



2ª edição

Descobrimdo a ESTATÍSTICA usando o SPSS

ANDY FIELD





F453d Field, Andy.

Descobrimo a estatística usando o SPSS [recurso eletrônico]
/ Andy Field ; tradução Lorí Viali. – 2. ed. – Dados eletrônicos. –
Porto Alegre : Artmed, 2009.

Editado também como livro impresso em 2009.
ISBN 978-85-363-2018-2

1. Estatística - Informática. I. Título.

CDU 311:004

Catálogo na publicação: Renata de Souza Borges – CRB-10/1922

ANDY FIELD

Senior Lecturer in Psychology at The University of Sussex, UK



2ª edição

Descobrimo a ESTATÍSTICA usando o SPSS

Tradução, consultoria e supervisão desta edição:

Lorí Viali

Professor Titular da Famat/PUCRS

Professor Adjunto do IM/UFRGS

Versão impressa
desta obra: 2009



2009

Obra originalmente publicada sob o título *Discovering Statistics with SPSS 2nd Edition*
ISBN 0-71619-4452-4

© English language edition published by Sage Publications of London, Thousand Oaks and New Delhi,
© Andy Field, 2005.

© Portuguese language translation by Artmed Editora 2009.

Capa: *Paola Manica*

Preparação de original: *Maria Francisca de Oliveira Vargas*

Leitura final: *Lara Frichenbruder Kengeriski e Amanda Munari*

Supervisão editorial: *Mônica Ballejo Canto*

Editoração eletrônica: *Techbooks*

Reservados todos os direitos de publicação, em língua portuguesa, à
ARTMED® EDITORA S.A.
Av. Jerônimo de Ornelas, 670 - Santana
90040-340 Porto Alegre RS
Fone (51) 3027-7000 Fax (51) 3027-7070

É proibida a duplicação ou reprodução deste volume, no todo ou em parte, sob quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição na Web e outros), sem permissão expressa da Editora.

SÃO PAULO
Av. Angélica, 1091 - Higienópolis
01227-100 São Paulo SP
Fone (11) 3665-1100 Fax (11) 3667-1333

SAC 0800 703-3444

IMPRESSO NO BRASIL
PRINTED IN BRAZIL

Como a primeira edição, este livro é dedicado ao meu irmão Paul e ao meu gato Fuzzy (que é o mesmo gato da última edição, cujo apelido é “Beana”, mas que agora é novamente chamado do nome que eu originalmente dei a ele!), porque um deles é uma fonte constante de inspiração intelectual e o outro me acorda de manhã sentando em mim e me lambendo o rosto até que eu lhe dê comida de gato – as manhãs serão muito melhores quando meu irmão superar seu amor por comida de gato no café da manhã.☺

AGRADECIMENTOS

Obrigado à SPSS pela permissão do uso de suas imagens. A SPSS pode ser contatada na Avenida North Michigan, 444, em Chicago, Illinois 60611 (Estados Unidos) ou no primeiro andar da St. Andrew's House, West Street, Working, GU21 1EB (UK). Para mais informações, acessem o site (<http://www.spss.com>). Obrigado, também, a Alex Buchner, por disponibilizar seu programa G*Power.

A primeira edição não teria acontecido se não fosse por Dan Wright, que além de ter fé incondicional em um recém-pós-graduado, também leu e revisou muitos capítulos. Esta nova edição recebeu ajuda de muitas pessoas que enviaram *e-mails* oferecendo dicas referentes à primeira edição: agradeço muito a todos que gastaram seu tempo para apontar erros, sugerir melhorias e dar retornos maravilhosos. Assim, meus sinceros agradecimentos vão para Peter de Heus, Tilly Houtmans, Don Hunt, Paul Tinsley, Keith Tolfrey, Jaap Dronkers e Nick Smith (para nomear apenas alguns!). A Sage também organizou várias revisões anônimas que nos deram ótimas sugestões, assim, sou grato a essas pessoas, quem quer que elas sejam (Tony Cassidy, que admitiu ser uma delas, recebe meu agradecimento particular!) Jeremy Miles evitou que eu

me tornasse um completo idiota (no livro – é uma pena que seus poderes não se estendam para o cotidiano) apontando muitos erros; ele também foi uma pessoa muito agradável de conhecer nestes últimos anos (menos quando ele diz que aquelas seções de rascunhos do meu livro são, e eu cito com suas próprias palavras, “besteira”!). Gareth Williams e Lynne Slocombe ofereceram inspiradas sugestões pedagógicas; entretanto, eu nunca vou perdô-los por terem sugerido um glossário de termos-chave. Laura também testou a primeira edição do livro como uma tutora na graduação e, mais tarde, na pós-graduação do meu curso; ela também tomou tempo do seu curso de pós-graduação para comentar capítulos desta nova edição e sugerir muitas melhorias. Agradecimentos especiais por esta edição vão para David Hitchin, o herói desconhecido da estatística na University of Sussex. Ele despendeu muito do seu tempo oferecendo comentários sobre a primeira edição, e sobre alguns rascunhos de capítulos desta nova edição, me ensinando inúmeras coisas sobre estatística por *e-mail*. Todas essas pessoas leram a primeira edição ou grande parte dela (senão toda), e apontaram erros, sugeriram aprimoramentos ou, simplesmente, sugeriram coisas

que eu antes desconhecia. Não sei o que isso sugere sobre seus estados mentais, mas sou eternamente agradecido por seu bom coração: que eles vivam muito e que seus conjuntos de dados sejam normais.

Com o risco de me prolongar tanto como a cerimônia do Oscar, também estou muito agradecido às seguintes pessoas por escreverem boas críticas à primeira edição, tanto na Amazon.com ou na Amazon.co.uk: George H. Marshall (Escócia), Andrew (Itália), Dr. Simon Marshall (University of Loughborough, Grã Bretanha), um leitor da Royal Holloway University of London, Dr. Keith Tolfrey (Manchester Metropolitan University, Grã Bretanha), James MacCabe (Institute of Psychiatry, Londres, Grã Bretanha), Geoff Bird (University College, Londres, Grã Bretanha), abare@telcel (Grã Bretanha), Jeremy Miles (embora levemente parcial, porque me conhece), Ken Kolosh (IL, EUA), Mehmet Yusuf Yahyagil (Turquia), Sandra Casillas (CA, EUA), um leitor de Missoula (MT, EUA), Paulo Ferreira Leite (Brasil), um leitor de Fair Lawn (NJ, EUA), Its1102 de Cresco (PA, EUA), Shaun Galloway (Hungria), smcol dos EUA, um leitor de Bowling Green (KY, EUA), Denis E. Hommrich (KY, EUA), Hsi Lung Wu (Taiwan), Iwan Wahyu (Singapura) e Mark Gray. Agradeço por vocês escreverem todas essas críticas. Se vocês algum dia forem a Brighton, lhes pago uma cerveja!

O pessoal da Sage é, sem dúvida, o mais beberrão que eu já conheci. Desde a primeira edição, tenho tido a felicidade de trabalhar com Michael Carmichael, que, apesar de suas falhas no campo de futebol (!), participou de memoráveis noites comigo. Em seu tempo livre, ele é também um ótimo editor e uma

pessoa maravilhosa que merece uma medalha por me aturar. Mas, para ser honesto, eu também mereço uma medalha por aguentar sua obsessão por glossários.

Se você tem um dos meus livros, deve saber que eu sempre escrevo ouvindo música. Devo a minha sanidade, por terem grandes músicas: Fugazi, Beck, Busta Rhymes, Abba, The Cardiacs, Mercury Ver, Ben e Jason, Plug, Roni Size, Supergrass, Massive Attack, Elvis Costello, The Smashing Pumpkins, Radiohead, Placebo, Money Mark, Love, Hefner, Nick Cave, DJ Shadow, Elliott Smith, Muse, Arvo Pärt, AC/DC e Quasi. Para essa segunda edição, a minha lista de músicas é um indicativo da angústia mental na qual afundei. Os tons encantadores das seguintes músicas me envolveram: Emperor, Cradle of Fifth, Frank Black and the Catholics, Blondie, Fugazi, Radiohead, Peter Gabriel, Genesis (a era de Peter Gabriel – tem um pouco da fase de rock progressivo por volta do Capítulo 5), Metallica, The White Stripes, Sevara Nazarkhan, Nusrat Fateh Ali Khan, Killing Joke, The Beyond, Jane's Addiction, Nevermore, The French and Hefner, Iron Maiden, The Mars Volta, Morrisey, Slipknot, Granddaddy, Mark Lanegan, PJ Harvey.

Escrever um livro exige horas solitárias (geralmente tarde da noite) de digitação. Sem a ajuda de meus amigos que me tiravam da minha sala pouco iluminada de tempos em tempos, seria mais abobado do que eu já sou. Em especial, sou grato a Graham e Benie, Martin, Doug, Paul, Darren, Helen Liddle e Mark por me lembrarem que existe vida além da estatística, e à Leonora, por achar que psicólogos que gostam de estatística dão bons maridos.☺

PREFÁCIO

Policia!l, prenda este homem,
ele fala em números
Ele zune como uma geladeira, ele
é como um rádio fora do ar.

(Radiohead, 1997)

INTRODUÇÃO

Os estudantes de ciências sociais sempre desprezaram a Estatística. Uma das razões é que muitos não entendem matemática, o que torna a compreensão de equações estatísticas complexas muito difícil. A maior vantagem em ter estudado estatística no início dos anos de 1990 (como eu estudei) comparado com os anos de 1960 foi o desenvolvimento de softwares para fazer todo o trabalho. A vantagem em aprender estatística hoje, comparado a dez anos atrás, é que esses pacotes agora são mais fáceis de usar graças ao Windows^{MR}/MacOS^{MR}. O SPSS é, em minha opinião, o melhor dos pacotes comerciais disponíveis, utilizado em muitas universidades. Então, o que me levou a escrever um livro sobre estatística e SPSS?

Já existem muitos bons livros que descrevem a teoria estatística. Howell (2002), Stevens (1992) e Wright (2002) escreveram livros

maravilhosos e claros, mas utilizam exemplos de computador somente como um adendo à teoria. Da mesma maneira, existem muitos bons livros sobre SPSS (Kinnear e Gray, 2000; Foster, 2001), mas eles se concentram em “fazer os testes”. Usar SPSS sem conhecimento algum de estatística pode ser perigoso (é somente uma ferramenta, não um recurso divino de sabedoria). Portanto, quero usar o SPSS como uma ferramenta para ensinar conceitos estatísticos. Agindo dessa maneira, espero que o leitor tenha um entendimento melhor, tanto da teoria quanto da prática.

Primeiro, quero responder o tipo de pergunta que eu fazia quando estudava estatística e usava o SPSS na faculdade (por exemplo, “Será que consigo entender como funciona esse teste estatístico sem saber muito de matemática?”, “O que este botão faz?”, “Mas que diabos significa esta saída?”). O SPSS tem um complexo conjunto de opções para cada teste, muitos dos quais são ensinados de modo superficial por livros e professores. Espero ser capaz de explicar o que essas opções realmente fazem e por que você deve usá-las. Quero ir além de fornecer receitas. Muitos livros dizem ao leitor o que fazer (“aperte esta tecla”, “faça isso”, “faça aquilo”, etc.), e isso cria uma impressão de que a estatística e o SPSS são infle-

xíveis. O SPSS tem muitas opções projetadas que lhe permitem modificar um teste já elaborado para seus fins específicos. Portanto, embora eu faça recomendações, espero fornecer aos leitores fundamentos teóricos suficientes para torná-los capazes de tomar suas próprias decisões sobre quais opções são mais apropriadas para as análises que eles querem fazer.

Meu segundo objetivo foi escrever um livro que pudesse ser lido por vários níveis (veja a próxima seção para mais detalhes). Existem capítulos para alunos do primeiro ano da faculdade (1, 2, 3, 4, 7 e 13), e capítulos para estudantes do segundo ano (5, 8, 9, 10, 11 e 12) e capítulos de tópicos mais avançados que alunos pós-graduados poderão usar (6, 14, 15 e 16). Todos esses capítulos devem ser acessíveis a qualquer um e espero conseguir isso indicando o nível adequado de cada seção (veja a próxima).

O QUE HÁ DE NOVO?

Eu era muito ruim em matemática. Aos 13 anos, era praticamente o último da minha turma. No entanto, 12 anos depois escrevi um livro-texto sobre estatística (e 17 anos depois terminei a segunda edição). A diferença entre o menino de 13 anos que não passou no seu exame e o de 15 anos que foi muito bem foi um bom professor: meu irmão, Paul. De fato, devo minha vida acadêmica à habilidade de Paul em fazer o que meus professores não fizeram: ensinar as matérias de uma maneira envolvente. Ainda hoje ele aparece, quando necessário, para me ensinar coisas (um curso intensivo de programação de computador, alguns Natais atrás, me vem à lembrança). De qualquer modo, a razão de ele ser um bom professor é que ele é capaz de tornar as coisas interessantes e relevantes para mim. Infelizmente, ele parece ter os genes de “bom professor” da família (e ele nem mesmo trabalha como professor, que desperdício!), mas eu tenho tentado usar a sua abordagem nas minhas aulas e livros. Algo que aprendi foi que as pessoas apreciam um toque de humanidade, assim, na primeira edição, tentei colocar mui-

to da minha personalidade (ou seria “falta de personalidade?”). Então, havia alguns exemplos despretensiosos (alguns disseram “indecentes”, mas prefiro “despretensiosos”) e uma grande dose de humor – o que pode ter sido uma coisa ruim porque eu sou um professor e, portanto, não tenho senso de humor!

Quando escrevi a primeira edição, meu objetivo era escrever um tipo de livro de estatística que eu gostaria de ler. Muito egoísta, eu sei, mas achei que se eu tivesse um livro de referência com exemplos que me divertissem, a vida se tornaria mais simples quando eu necessitasse de um livro para consulta. Não achei que alguém fosse comprar a obra (bem, fora minha mãe e meu pai) e eu antecipei desculpas para comentários do tipo “todo o capítulo X está completamente errado e você é um idiota”, ou “pense em quantas árvores morreram para este lixo ser escrito, você deveria se envergonhar”. Na verdade, nem a editora pensou que o livro iria vender (eles revelaram isso posteriormente, devo acrescentar). Eu realmente não esperava receber muitos *e-mails* extremamente gentis de pessoas que gostaram do livro. Até hoje ainda fico admirado que alguém leia o livro e gaste tempo para me escrever um *e-mail* gentil (e devo acrescentar, com risco de parecer sentimental, que saber que o livro ajuda as pessoas faz com que o sorriso na minha face nunca desapareça). Mas por que estou contando tudo isso? Bem, vendo que pelo menos algumas pessoas apreciaram o estilo da primeira edição, tomei isto como uma luz verde para incluir mais exemplos estúpidos, mais bobagens e mais mau gosto. Resumindo, muito mais sexo, drogas e *rock ‘n’ roll*. Para todos aqueles que odeiam isso, desculpa, mas isso me diverte.

Além de acrescentar mais bobagens, fui forçado, com relutância, a expandir o conteúdo acadêmico. Muitos dos acréscimos são resultados de alguém (geralmente muitas pessoas) que me enviaram *e-mails* perguntando como fazer algo. Assim, teoricamente, esta edição deveria responder as perguntas que me foram feitas nos últimos quatro anos! As mudanças gerais em cada capítulo são:

- **Exemplos:** Cada capítulo agora tem muitos exemplos no final para você trabalhar, incluindo as respostas.
- **Mais dados:** Por causa dos exemplos, você tem mais conjuntos de dados!
- **Relatando suas análises:** Existem seções sobre como relatar as análises na maioria dos capítulos. Essas seções foram baseadas no manual de estilo de publicação da APA (American Psychological Association, 2001).
- **Glossário:** Termos-chave estão agora identificados no texto e existe um glossário de todos eles.
- **Tamanho de efeito:** A maioria dos tópicos agora tem uma discussão sobre como calcular o tamanho de efeito.
- **Resumos:** Há muita informação no texto e meu estilo é divagador (prefiro “conversador”), assim, existem agora quadros com resumo dos pontos-chave (veja a próxima seção sobre como usar este livro).
- **Quadros:** Inúmeros quadros que ressaltam temas interessantes.
- **Sintaxe:** Existe agora discussão sobre o uso da sintaxe do SPSS ao longo do livro.
- **SPSS 13:** Embora essas fossem as menores mudanças, porque o SPSS 13 não é muito diferente da versão 9, atualizei tudo com a última versão do SPSS.

As mudanças específicas de cada capítulo são:

- **Capítulo 1 (Conceitos básicos):** Foi expandido para incluir mais fundamentos em teoria da estatística (por exemplo, intervalos de confiança, testes de significância, poder, tamanho de efeito).
- **Capítulo 2 (SPSS):** Estava condensado dentro do Capítulo 1, mas foi atualizado para lidar com o SPSS 13 e agora inclui discussão da assustadora janela de sintaxe.
- **Capítulo 3 (Explorando dados):** Tem alguma semelhança com o antigo Capítulo 2 da primeira edição. Ele agora aborda apresentação de dados, transformação de dados, verificação de modelos de dados e muito mais.
- **Capítulo 4 (Correlação):** Estava no Capítulo 3 e, na verdade, não mudou muito.
- **Capítulo 5 (Regressão):** Estava no Capítulo 4 e agora tem um material mais avançado, como uma seção em previsores categóricos, bem como uma discussão de autovalores. Também expandi as seções originais de forma mais técnica (por exemplo, uma discussão mais aprofundada do tamanho de efeito).
- **Capítulo 6 (Regressão Logística):** O antigo Capítulo 5 foi aumentado levemente de várias formas (a parte teórica principalmente), mas também reorganizei alguns conteúdos de modo que a teoria e as dicas do SPSS estão mais claramente separadas.
- **Capítulo 7 (Testes t):** Não foi alterado em relação ao antigo Capítulo 6 exceto algumas informações das quais não gostava que foram reescritas.
- **Capítulo 8 (MLG 1):** É basicamente o mesmo antigo Capítulo 7, exceto que agora cobre as versões de Welch e Brown-Forsythe da F quando a homogeneidade da variância não pode ser assumida.
- **Capítulo 9 (MLG 2):** Aborda a análise de covariância (ANCOVA), que antes dividia um capítulo com a ANOVA Fatorial, mas agora ganhou um capítulo próprio. Coloquei algum material extra sobre contrastes planejados e corriji alguns erros embaraçosos da primeira edição.☺
- **Capítulo 10 (MLG 3):** Cobre ANOVA Fatorial (que dividia um capítulo com a ANCOVA). Existem seções inteiramente novas sobre a teoria da ANOVA independente de dois fatores, como realizar contrastes (com sintaxe), como realizar análise de efeitos simples, interpretação de diagramas de interações e como podemos conceitualizar a ANOVA de dois fatores como um MLG (Modelo Linear Generalizado).
- **Capítulo 11 (MLG 4):** Era o Capítulo 9 sobre a ANOVA de medidas repetidas. Adicionei uma seção sobre a teoria da ANOVA de medidas repetidas e sobre a

condução de análises de efeitos simples na ANOVA de medidas repetidas com dois fatores.

- **Capítulo 12 (MLG 5):** Aborda a ANOVA mista (estava junto com ANOVA de medidas repetidas no antigo Capítulo 9). Existe um exemplo completamente novo e, com o novo capítulo, expandi um pouco as seções de interpretação.
- **Capítulo 13 (Estatística não-paramétrica):** É um capítulo completamente novo (embora pegue emprestadas algumas coisas do Capítulo 2). Acrescentei seções de teoria para os testes de Mann-Whitney e Wilcoxon; também acrescentei seções completamente novas que cobrem a ANOVA de Friedman, o teste de Jonckheere e o teste de Kruskal-Wallis.
- **Capítulo 14: (MANOVA):** Não mudou muito do Capítulo 10.
- **Capítulo 15 (Análise de fatores):** Reorganizei o material para que o que estava escondido na seção de interpretação ficasse agora claramente colocado na seção de teoria. É uma seção completamente nova sobre análise de confiabilidade.
- **Capítulo 16 (Dados categóricos):** É basicamente um capítulo novo (embora o exemplo do qui-quadrado seja antigo). Incluí uma seção de teoria para o teste qui-quadrado, razão de verossimilhança e correção de Yates, discussão sobre razão

de chances (*odds ratios*) e análise Loglinear. Nenhuma dessas seções estava na antiga edição!

ADEUS

Escrevi no início da primeira edição que “este livro é o resultado de dois anos (com uma ou duas semanas para escrever a minha tese de doutorado) tentando atingir esses objetivos. Ele não é perfeito, e eu gostaria de ter um retorno (bom ou ruim) das pessoas que realmente contam: vocês, os leitores”. Esse sentimento ainda se aplica, exceto que agora é o resultado de dois anos e meio de trabalho. Ao longo desses últimos quatro anos, fiquei muito ligado a este livro: ele começou com um trabalho de amor e ainda é. Com o inesperado sucesso da primeira edição, e, tendo feito muitas atualizações para a segunda edição (trezentas páginas comparadas com as cinquenta páginas que a editora queria que eu escrevesse!), estou muito preocupado que eu tenha mudado tudo! Portanto, embora eu não saiba se o livro terá uma terceira edição, mesmo assim, gostaria de ter o retorno das pessoas que realmente contam: vocês, os leitores.

Andy

E-mail: discoveringstatistics@sussex.ac.uk

Web: <http://www.sussex.ac.uk/Users/andyf/>

COMO USAR ESTE LIVRO

Quando a editora me pediu para escrever uma seção sobre “como usar este livro” foi tentador escrever “compre um frasco de creme antirrugas (o que você irá precisar para se defender dos efeitos do envelhecimento enquanto ler o livro), procure uma cadeira confortável, sente, dobre a capa, comece a ler e só pare quando chegar à última página”. Entretanto, acho que eles gostariam de algo mais útil.☺

QUE CONHECIMENTO ANTERIOR VOCÊ PRECISA TER?

Baseei-me na suposição de que você não sabe nada sobre estatística, mas tem algum conhecimento básico sobre computador (eu não vou dizer como ligá-lo, por exemplo) e matemática (embora eu tenha incluído uma revisão rápida sobre alguns conceitos básicos, assim, eu não suponho nenhum conhecimento prévio).

OS CAPÍTULOS FICAM MAIS DIFÍCEIS AO LONGO DO LIVRO?

De certa forma ficam (o Capítulo 14 sobre MANOVA é mais difícil do que o Capítulo 1), mas em outros casos, não (o Capítulo 13, so-

bre estatística não-paramétrica é menos complexo que o Capítulo 12, e o Capítulo 7, sobre o teste-t, é menos complexo do que Capítulo 6 sobre regressão logística).

Por que eu fiz isso? Ordenei os capítulos para ter coerência estatística (para mim, pelo menos). Muitos livros ensinam vários testes isoladamente e não apresentam a idéia global; isso, eu acho, cria um mistério desnecessário. A maioria dos testes neste livro é a mesma apresentada de uma maneira ligeiramente diferente. Assim, eu queria que o livro contasse essa história. Para fazer isso, tenho que explicar alguns conceitos, como o de regressão, bem no começo, porque quase tudo está baseado nesse conhecimento!

Entretanto, para ajudá-lo, codifiquei cada seção com um ícone, que representa a dificuldade do conteúdo. Eles não significam que você deva pular seções (veja o Alex Esperto na próxima seção), mas irão informar se a seção está no seu nível ou se irá exigir mais conhecimento de você. Eu baseei esses ícones no meu modo de ensinar, portanto, eles podem não se adequar a todos (especialmente porque sistemas variam em diferentes países!)

① Isto significa “nível um” e eu o equi-parei ao primeiro ano da faculdade no Reino

Unido. Todos deveriam ser capazes de entender essas seções.

② Este é o próximo nível e eu o equiparei ao segundo ano da faculdade no Reino Unido. São tópicos que eu leciono no meu segundo ano e qualquer um com pouco de conhecimento de estatística deveria ser capaz de entender. Entretanto, algumas dessas seções são desafiadoras mesmo para alunos do segundo ano. São seções intermediárias.

③ Este é o “nível 3” e representa tópicos difíceis. Espero que alunos do último ano do Reino Unido e recém-formados sejam capazes de lidar com essas seções.

④ Este é o nível mais alto e representa tópicos muito difíceis. Espero que esses tópicos sejam muito desafiadores para estudantes da graduação e recém-formados, mas os recém-formados, com um razoável conhecimento em métodos de pesquisa, não deverão ter muitas dificuldades.

POR QUE VOCÊ VÊ ROSTOS ESTÚPIDOS EM TODOS OS LUGARES?



Alex Esperto: Alex é um personagem muito importante porque ele aparece quando as coisas ficam difíceis. Ele é um pouco malandro, assim, quando a sua cara aparecer, você saberá que algo complicado irá ser explicado. Quando o conteúdo difícil termina, ele reaparece para informar que é seguro continuar. Mas isso não quer dizer que o resto do material do livro é fácil; ele apenas informa as partes do livro que você pode pular se tem coisas melhores a fazer com sua vida do que ler 800 páginas! Então, quando o Alex Esperto aparecer, você pode *pular a seção inteira* e ainda assim entender o que está acontecendo. O Alex também aparece no final de cada capítulo apresentando algumas tarefas para testar se você é tão esperto quanto ele. A propósito, qualquer semelhança entre Alex e o meu editor é pura coincidência!

Samanta Ferrinho: Samanta odeia estatística. Na verdade, ela acha que tudo isso é uma perda de tempo. Ela simplesmente quer

passar no exame e esquecer que teve de aprender distribuição normal. Assim, quando aparece, ela fornece um resumo dos pontos-chave que você precisa saber. A Samanta lhe dará a informação essencial para que você não fique procurando em centenas de páginas o que deseja saber.



Brian Cansado: Brian aparece fazendo perguntas e completamente confuso. Portanto, ele não é muito diferente do autor. À medida que o livro avança, ele se torna cada vez mais desanimado. Tire suas próprias conclusões!

Gato Curioso: Ele também aparece para fazer perguntas (porque ele é curioso). Na verdade, ele só está aqui é porque eu queria um gato no livro... e, de preferência, um que se parecesse com o meu. É claro que os especialistas em educação acham que ele precisa de um papel específico, assim, seu papel é parecer bonito e contar péssimas piadas sobre gatos.



PARA QUE SERVEM OS QUADROS?

Há quadros ao longo de todo o livro. Os quadros podem ser ignorados, se você quiser, porque são lembretes incluídos no texto. Eles explicam assuntos complicados ou ilustram pontos interessantes sobre estatística, com os quais você pode impressionar seus amigos.

O QUE ESTÁ NO SITE*?

O site contém muitas coisas interessantes:

- **Arquivo de dados:** Em geral, você irá usá-lo nos exemplos, porque ele contém todos os arquivos de dados dos exemplos

* N. de R. O conteúdo disponibilizado no site www.art-med.com.br está em inglês. Consulte também: www.sagepub.co.uk/field, que contém materiais adicionais para o estudante (conteúdo em inglês).

do livro (os dados para cada capítulo estão contidos numa pasta separada).

- **Respostas para o Alex Esperto:** Porque não quero matar mais árvores do que o necessário, há centenas de materiais adicionais no *site*. Os documentos mais importantes são as respostas às tarefas no final de cada capítulo. Para cada tarefa, o documento contém a saída do SPSS e uma breve explicação, assim, você pode checar suas respostas.
- **Material adicional:** Para alguns tópicos, escrevi longos documentos. São, principalmente, tópicos avançados ou tópicos para os quais uma explicação detalhada no corpo do livro seria uma distração. Em

caso de alguém querer saber mais, há descrições no *site*. Por exemplo, há arquivos sobre contrastes usando a sintaxe, a razão F de Welch, teste de Jonckheere e assim por diante. A maioria pode ignorar esses arquivos, mas eles estão no livro caso alguém esteja interessado.

- **Material do apêndice:** Algum material do apêndice (como os cálculos matemáticos) também foi colocado nos arquivos do *site* para economizar papel.
- **Programas:** Há uma cópia do programa G*Power no *site*.

Boa leitura!

SÍMBOLOS UTILIZADOS

Operadores matemáticos

Σ	Este símbolo (denominado sigma) significa “some tudo”. Assim, se você vir algo como Σx_i , isso quer dizer “some todos os valores que coletou”.	σ^2	A variância de um conjunto de dados populacionais
Π	Este símbolo significa “multiplique tudo”. Assim, se aparecer algo como Πx_i , isso apenas quer dizer que você deve “multiplicar todos os valores que coletou”.	σ	O desvio padrão de um conjunto de dados populacionais
\sqrt{x}	Isto quer dizer “tire a raiz de x ”.	σ_x	O erro padrão da média
		τ	Tau de Kendall (coeficiente de correlação não-paramétrico)
		ω^2	Ômega ao quadrado (medida do tamanho de efeito)

Símbolos gregos

α	É a probabilidade de cometer erro do Tipo I
β	É a probabilidade de cometer erro do Tipo II
β_i	Coefficiente da regressão padronizado
χ^2	Estatística teste qui-quadrado
χ^2_F	Estatística teste ANOVA de Friedman
ε	Normalmente significa “erro”
η^2	Eta ao quadrado (uma medida do tamanho de um efeito)
μ	A média de uma população de valores
ρ	O coeficiente de correlação populacional

Símbolos ingleses

b_i	O coeficiente de regressão (não-padronizado)
gl	Graus de liberdade ($df = \text{degrees of freedom}$)
e_i	O erro associado à i -ésima pessoa
F	Razão F (estatística teste utilizada na ANOVA)
H	Estatística teste de Kruskal-Wallis
k	O número de níveis de uma variável (isto é, o número de condições de um tratamento) ou o número de previsores num modelo de regressão
\ln	Logaritmo natural
MS	O erro médio ao quadrado (média ao quadrado). A variabilidade média nos dados

N, n, n_i	O tamanho da amostra. N normalmente representa o tamanho de uma amostra qualquer enquanto n representa o tamanho de uma particular amostra.
P	Probabilidade (o valor de uma probabilidade, o valor- p ou a significância de um teste são normalmente representados por p)
r	O coeficiente de correlação de Pearson
r_b, r_{pb}	Coefficiente de correlação bisserial e coeficiente de correlação ponto-bisserial, respectivamente
r_s	Coefficiente de correlação por postos de Spearman
R	O coeficiente de correlação múltiplo
R^2	O coeficiente de determinação (isto é, a proporção da variância dentro de alguns dados explicados pelo modelo)
s	O desvio padrão de uma amostra
s^2	A variância de uma amostra
SS	A soma dos quadrados ou a soma dos erros ao quadrado se for dado o nome completo
SS_A	A soma dos quadrados para a variável A
SS_M	A soma dos quadrados do modelo (isto é, a variabilidade explicada pelo modelo ajustado aos dados)
SS_R	A soma dos quadrados dos resíduos (isto é, a variabilidade que o modelo não pode explicar, o erro no modelo)
SS_T	A soma total dos quadrados (isto é, a total variabilidade dos dados)
t	A estatística teste de Student
T	A estatística para o teste de postos com sinais de Wilcoxon
U	A estatística para o teste de Mann-Whitney
W_s	A estatística para o teste da soma dos postos de Wilcoxon
\bar{X} ou \bar{x}	A média de uma amostra
z	Um ponto de dados expresso em unidades de desvio padrão

Revisão matemática

1. **Dois negativos fazem um positivo:** embora na vida dois erros não façam um

acerto, na matemática eles fazem! Quando multiplicamos um número negativo por outro negativo, o resultado é um número positivo. Por exemplo, $-2 \times -4 = 8$.

2. **Um número negativo multiplicado por um positivo produz um negativo:** se você multiplica um número positivo por um negativo, o resultado é outro número negativo. Por exemplo, $2 \times -4 = -8$ ou $-2 \times 6 = -12$

3. **PODMAS:** é um acrônimo para a ordem em que as operações matemáticas são executadas. Ele significa Parênteses, Ordem, Divisão, Multiplicação, Adição e Subtração, e essa é a ordem em que você deve executar operações em uma expressão. Muitas dessas operações são autoexplicativas (por exemplo, sempre calcule coisas dentro dos parênteses primeiro), exceto por ordens que de fato se referem a termos com potências como quadrados. Quatro ao quadrado, ou 4^2 , é lido como quatro elevado na ordem dois, por isso esses termos são denominados “ordem” em PODMAS (também, se os denominarmos de potência, vamos ter PPODMAS, o que será difícil de pronunciar). Vamos dar uma olhada em um exemplo de PODMAS: qual seria o resultado de $1 + 3 \times 5^2$? A resposta é 76 (não 100, como alguns de vocês podem ter pensado). Não existem parênteses, assim, o primeiro a fazer é lidar com o termo ordem: $5^2 = 25$, assim a expressão se torna $1 + 3 \times 25$. Não existe divisão, portanto, podemos passar para a multiplicação: 3×25 , que resulta 75. PODMAS nos diz que agora é hora de lidarmos com a adição: $1 + 75$, que fornece o resultado 76. Se eu tivesse escrito a expressão original como $(1 + 3) \times 5^2$, a resposta seria 100 porque nesse caso devemos lidar primeiro com os parênteses: $(1 + 3) = 4$, assim a expressão torna-se 4×5^2 . Agora devemos lidar com o termo ordem e a expressão torna-se: $4 \times 25 = 100$!

4. <http://www.easymaths.com> é um bom local para uma revisão de conceitos matemáticos básicos.

COMENTÁRIOS SOBRE A PRIMEIRA EDIÇÃO

DE ESTUDANTES:

O livro está muito bem escrito. Está tão claro que é como se eu visse a estatística, que eu estudei por tanto tempo, através de um par de óculos limpos.

Megan Gray, Universidade de Sussex

Sua combinação da parte técnica com o conceito é simplesmente maravilhosa. E o seu senso de humor, também!

Abigail Levy

Você tem a habilidade de explicar coisas complexas de uma maneira simples e compreensível.

Ruth Mann

Eu não seria capaz de terminar meus estudos tão facilmente se não fosse o seu livro.

Marleen Smits, Universidade de Maastricht

Seu livro me tornou um aluno muito mais feliz.

Lisa Oliver, Universidade da Columbia Britânica

Eu nunca vi conceitos tão bem explicados ou ilustrados tão claramente.

Joel Philip

Eu defendi minha dissertação em psicologia em março e seu livro SALVOU MINHA VIDA!

Noelle Leonard

Eu acho que você vai para história como a primeira pessoa a colocar humor em um livro de estatística!

Carol McSweeney, Universidade Brookes

Você é a mais pura expressão de um bom professor! Embora eu não o conheça ou tenha assistido às suas aulas, sua empatia e seu desejo genuíno de ajudar os estudantes se refletem no livro.

Johannah Sirkka

DE ACADÊMICOS

Ano passado descobri suas páginas na rede (e, desse modo, seu livro) e ele transformou completamente minhas aulas de estatís-

tica básica para alunos de pós-graduação de “ok” para “ótimas”. Como resultado, este ano (e num futuro próximo) usarei o seu livro como texto principal na minha classe.

Michael Marsiske, Universidade da Flórida

Eu recomendei seu livro para um dos meus cursos e estou ansioso para que meus alunos cheguem. Leciono estatística há 25 anos e usei muitos livros. Mas este é o melhor. Você me deu uma nova visão de coisas com as quais tive dificuldade por anos.

Henry Steel, Universidade de Stellenbosch, África do Sul

A maneira como você consegue trazer o leitor ao mundo da estatística avançada sem muito jargão técnico é admirável.

Michael A. Karchmer, Diretor do Instituto de Pesquisa Gallaudet, Universidade de Gallaudet

Muito obrigado por escrever um livro excelente e inspirador sobre um tópico que é quase impossível tornar interessante – você fez um milagre!

Dr. Keith Tolfrey, Professor Sênior, Departamento da Ciência do Esporte e do Exercício, Universidade Metropolitana de Manchester.

Desde que o descobri, recomendei seu livro para os meus colegas e acho que pelo menos metade do departamento tem seu livro na estante agora. Quando alguém não sabe o que fazer com seus dados, simplesmente dizemos: consulte o Andy!

Froukje Dijk, Universidade de Maastricht

SUMÁRIO

AGRADECIMENTOS vii
PREFÁCIO ix
COMO USAR ESTE LIVRO xiii
SÍMBOLOS UTILIZADOS xvi
COMENTÁRIOS SOBRE A PRIMEIRA EDIÇÃO xviii

1 TUDO O QUE VOCÊ SEMPRE QUIS SABER SOBRE ESTATÍSTICA (BEM, QUASE TUDO) 31

1.1 O que você vai aprender neste capítulo? ① 31

1.2 Construindo modelos estatísticos ① 31

1.3 Populações e amostras ① 33

1.4 Modelos estatísticos simples ① 33

1.4.1 A média, a soma dos quadrados, a variância e o desvio padrão ① 33

1.5 Distribuições de frequências ① 37

1.5.1 Propriedades das distribuições de frequências ① 37

1.5.2 O desvio padrão e a forma da distribuição ① 38

1.5.3 O que é a distribuição normal padrão? ① 39

1.6 A minha amostra é representativa da população? ① 42

1.6.1 O erro padrão ① 42

1.6.2 Intervalos de confiança ② 44

1.7 Modelos lineares ① 47

1.8 Como descobrir se o seu modelo representa o mundo real? ① 49

1.8.1 Estatísticas teste ① 53

1.8.2 Testes uni e bilaterais ① 54

1.8.3 Erros do tipo I e do tipo II ① 56

1.8.4 Tamanhos de efeito ② 56

1.8.5 Poder estatístico ② 57

1.9 Últimos conselhos 58

1.10 O que descobrimos sobre estatística? ① 59

1.11 Termos-chave que descobrimos 59

1.12 Questões de revisão do Alex Esperto 59

1.13 Leituras complementares 60

2 O AMBIENTE DO SPSS 61

2.1 O que você vai aprender neste capítulo? ① 61

2.2 Versões do SPSS ① 61

2.3 Iniciando ① 62

2.4 O editor de dados ① 62

2.4.1 Entrando com os dados no editor ① 68

2.4.2 Criando uma variável ① 69

2.4.3 O visualizador de variáveis (Variable View) ① 70

2.4.4 Criando variáveis codificadas ① 72

2.4.5 Tipos de variáveis ① 74

2.4.6 Valores que faltam (missing) ① 75

2.4.7	Mudando o formato da coluna ①	76	3.10	Tarefas do Alex Esperto	124
2.4.8	Teste rápido ①	77	3.11	Leituras complementares	124
2.5	O visualizador de saídas (Output Viewer) ①	77	4	CORRELAÇÃO	125
2.6	A janela de sintaxe ③	80	4.1	O que você vai aprender neste capítulo? ①	125
2.7	Salvando arquivos ①	81	4.2	Como medimos relacionamentos? ①	125
2.8	Recuperando um arquivo ①	82	4.2.1	Um desvio para o mundo da covariância ①	125
2.9	O que descobrimos sobre estatística? ①	83	4.2.2	Padronização e o coeficiente de correlação ①	128
2.10	Termos-chave que descobrimos	83	4.3	Entrada de dados para realizar uma análise de correlação no SPSS ①	129
2.11	Tarefa do Alex Esperto	83	4.4	Representando relacionamentos graficamente: o diagrama de dispersão ①	129
2.12	Leituras complementares	84	4.4.1	Diagrama de dispersão simples ①	130
3	EXPLORANDO DADOS	85	4.4.2	Diagrama de dispersão tridimensional ①	133
3.1	O que você vai aprender neste capítulo? ①	85	4.4.3	Diagrama de dispersão sobreposto ①	134
3.2	Dados paramétricos ①	85	4.4.4	Diagrama de dispersão matricial ①	136
3.2.1	Hipóteses “retirar” ①	85	4.5	Correlação bivariada ①	137
3.3	Apresentando dados graficamente ①	87	4.5.1	Coeficiente de correlação de Pearson ①	140
3.3.1	Passo 1: Apontando os erros óbvios utilizando histogramas ①	87	4.5.2	Um alerta sobre interpretação: Causalidade ①	142
3.3.2	Passo 2: Estatísticas descritivas e o diagrama de caixa e bigodes ①	92	4.5.3	Utilizando o R^2 para interpretação ①	143
3.3.3	Passo 3: Corrigindo problemas nos dados ②	99	4.5.4	Coeficiente de correlação de Spearman ①	144
3.3.4	Passo 4: Transformando dados utilizando o SPSS ②	101	4.5.5	Tau de Kendall (não-paramétrico) ①	145
3.4	Explorando grupos de dados ①	107	4.5.6	Correlações bisserial e bisserial por ponto ②	146
3.4.1	Executando a análise para todos os dados ①	107	4.6	Correlação parcial ②	148
3.4.2	Saída do SPSS para todos os dados ①	108	4.6.1	A teoria por trás da correlação parte e parcial ②	148
3.4.3	Executando a análise para grupos diferentes ①	110	4.6.2	Correlação parcial utilizando o SPSS ②	149
3.4.4	Saídas para grupos diferentes ①	110	4.6.3	Correlações semiparciais (ou por parte) ②	153
3.5	Testando se uma distribuição é normal ①	112	4.7	Como relatar coeficientes de correlação ①	153
3.5.1	Executando o teste de Kolmogorov-Smirnov no SPSS ①	113	4.8	O que descobrimos sobre estatística? ①	154
3.5.2	Saída do procedimento Explorar (Explore) ①	113	4.9	Termos-chave que descobrimos	155
3.6	Testando a homogeneidade da variância ①	116			
3.7	Representando médias graficamente ①	119			
3.8	O que descobrimos sobre estatística? ①	123			
3.9	Termos-chave que descobrimos	124			

4.10	Tarefas do Alex Esperto	155
4.11	Leituras complementares	155
5	REGRESSÃO	156
5.1	O que você vai aprender neste capítulo? ①	156
5.2	Uma introdução à regressão ①	156
5.2.1	Algumas informações importantes sobre retas ①	157
5.2.2	O método dos mínimos quadrados ①	158
5.2.3	Avaliando o ajuste: soma dos quadrados, R e R^2 ①	159
5.2.4	Interpretando previsões individuais ①	162
5.3	Executando regressão simples no SPSS ①	164
5.4	Interpretando a regressão simples ①	165
5.4.1	Ajuste global do modelo ①	165
5.4.2	Parâmetros do modelo ①	166
5.4.3	Utilizando o modelo ①	167
5.5	Regressão múltipla: o básico ②	168
5.5.1	Um exemplo de um modelo de regressão múltipla ②	168
5.5.2	Soma dos quadrados, R e R^2 ②	170
5.5.3	Métodos de regressão ②	170
5.5.3.1	Hierárquico (Entrada em blocos) ②	170
5.5.3.2	Entrada forçada ②	171
5.5.3.3	Métodos por passos (Stepwise) ②	171
5.5.3.4	Escolhendo um método ②	172
5.6	Quão acurado é o meu modelo de regressão? ②	172
5.6.1	Interpretando o modelo de regressão I: diagnósticos ②	172
5.6.1.1	Valores atípicos (outliers) e resíduos ②	172
5.6.1.2	Casos influentes ③	174
5.6.1.3	Um comentário final sobre estatísticas diagnósticas ②	178
5.6.2	Interpretando o modelo de regressão II: generalização ②	178
5.6.2.1	Avaliando hipóteses ②	178
5.6.2.2	Validação cruzada do modelo ③	180
5.6.2.3	Tamanho da amostra na regressão ③	181
5.6.2.4	Multicolinearidade ②	182
5.7	Como executar uma regressão múltipla no SPSS ②	184
5.7.1	Principais opções ②	184
5.7.2	Estatísticas ②	186
5.7.3	Diagramas da regressão ②	188
5.7.4	Salvando os diagnósticos da regressão ②	189
5.7.5	Opções adicionais ②	190
5.8	Interpretando a regressão múltipla ②	191
5.8.1	Descritivas ②	192
5.8.2	Resumo do modelo ②	193
5.8.3	Parâmetros do modelo ②	197
5.8.4	Variáveis excluídas ②	201
5.8.5	Avaliando as hipóteses de não-multicolinearidade ②	202
5.8.6	Diagnósticos por casos (Casewise) ②	204
5.8.7	Conferindo as hipóteses ②	208
5.9	Como relatar a regressão múltipla ②	212
5.10	Previsores categóricos e a regressão múltipla ③	212
5.10.1	Variáveis dummy ③	213
5.10.2	Saída do SPSS para variáveis auxiliares (dummy) ③	215
5.11	O que descobrimos sobre estatística? ①	219
5.12	Termos-chave que descobrimos	219
5.13	Tarefas do Alex Esperto	220
5.14	Leituras complementares	220
6	REGRESSÃO LOGÍSTICA	221
6.1	O que você vai aprender neste capítulo? ①	221
6.2	Pressupostos da regressão logística ①	221
6.3	Quais são os princípios por trás da regressão logística? ③	222
6.3.1	Avaliando o modelo: a estatística conhecidos de verossimilhança-log ③	223
6.3.2	Avaliando o modelo: R e R^2 ③	224
6.3.3	Avaliando a contribuição dos previsores: a estatística de Wald ②	225
6.3.4	Exp b ③	226
6.3.5	Métodos de regressão logística ②	227
6.3.5.1	O método da entrada forçada	227
6.3.5.2	Métodos passo a passo (stepwise)	227

6.3.5.3	Como selecionamos um método? ②	228	7.3.2.1	Passo 1: Calcular a média para cada participante ②	275
6.4	Executando a análise: um exemplo de pesquisa ②	228	7.3.2.2	Passo 2: Calcular a média geral ②	275
6.4.1	A análise principal ②	229	7.3.2.3	Passo 3: Calcular o fator de ajustamento ②	276
6.4.2	Métodos de regressão ②	230	7.3.2.4	Passo 4: Criar os valores ajustados para cada variável ②	276
6.4.3	Previsores categóricos ②	230	7.3.2.5	Traçar o diagrama de barras de erros ②	278
6.4.4	Obtendo resíduos ②	231	7.4	Testando diferenças entre médias: o teste t ①	278
6.4.5	Opções adicionais ②	232	7.4.1	Fundamentação do teste t ①	279
6.5	Interpretando a regressão logística ②	233	7.4.2	Suposições do teste t ①	280
6.5.1	O modelo inicial ②	233	7.5	O teste t dependente ①	280
6.5.2	Passo 1: Entendendo falsas crenças ③	235	7.5.1	Distribuições amostrais e o erro padrão ①	281
6.5.3	Listando probabilidades previstas ②	242	7.5.2	A equação do teste t dependente explicada ①	281
6.5.4	Interpretando os resíduos ②	243	7.5.3	O teste t dependente utilizando o SPSS ①	283
6.5.5	Calculando o tamanho de efeito ②	247	7.5.4	Saídas do teste t dependente ①	284
6.6	Como relatar a regressão logística ②	247	7.5.5	Calculando o tamanho de efeito ②	286
6.7	Outro exemplo ②	247	7.5.6	Relatando o teste t dependente ①	286
6.7.1	Executando a análise: regressão por entrada em blocos ②	248	7.6	O teste t independente ①	287
6.7.2	Interpretando saídas ③	250	7.6.1	A equação do teste t independente explicada ①	287
6.8	Testes para a multicolinearidade ③	256	7.6.2	O teste t independente utilizando o SPSS ①	289
6.9	O que pode dar errado ④	260	7.6.3	Saídas do teste t independente ①	290
6.9.1	Informação incompleta dos previsores ④	260	7.6.4	Calculando o tamanho de efeito ①	292
6.9.2	Separação completa ④	261	7.6.5	Relatando o teste t independente ①	292
6.10	O que descobrimos sobre Estatística? ①	262	7.7	Entre grupos ou medidas repetidas? ①	293
6.11	Termos-chave que descobrimos	262	7.8	O teste t como um modelo linear generalizado ②	294
6.12	Tarefas do Alex Esperto	262	7.9	O que fazer se os dados não forem normalmente distribuídos? ②	295
6.13	Leituras complementares	264	7.10	O que descobrimos sobre Estatística? ①	295
7	COMPARANDO DUAS MÉDIAS	265	7.11	Termos-chave que descobrimos	296
7.1	O que você vai aprender neste capítulo? ①	265	7.12	Tarefas do Alex Esperto	296
7.2	Revisão da pesquisa experimental ①	266	7.13	Leituras complementares	297
7.2.1	Os dois métodos de coletar dados ①	266			
7.2.2	Dois tipos de variação ①	266			
7.2.3	Aleatorização ①	268			
7.3	Atribuindo dados e apresentando médias com diagramas de barras de erros	269			
7.3.1	Diagramas de barras de erros para delineamentos entre grupos ①	271			
7.3.2	Diagramas de barras de erros para delineamentos de medidas repetidas ②	274			

8 COMPARANDO VÁRIAS MÉDIAS: ANOVA (MLG 1) 298

- 8.1 O que você vai aprender neste capítulo? ① 298
- 8.2 A teoria por trás da ANOVA ② 298
 - 8.2.1 Taxas de erros infladas ② 298
 - 8.2.2 ANOVA como regressão ② 299
 - 8.2.3 A lógica da razão F ② 303
 - 8.2.4 Soma dos quadrados Total (SS_T) ② 305
 - 8.2.5 Soma dos Quadrados do Modelo (SS_M) ② 306
 - 8.2.6 Soma dos Quadrados dos Resíduos (SS_R) ② 307
 - 8.2.7 Médias ao quadrado ② 308
 - 8.2.8 A razão F ② 308
 - 8.2.9 Suposições da ANOVA ② 309
 - 8.2.10 Contrastes planejados ② 310
 - 8.2.10.1 Escolhendo que contrastes realizar ② 310
 - 8.2.10.2 Definindo contrastes ponderados ③ 314
 - 8.2.10.3 Comparações não-ortogonais ② 319
 - 8.2.10.4 Contrastes padrões ② 320
 - 8.2.10.5 Contrastes polinomiais: análise da tendência ② 321
 - 8.2.11 Procedimentos *post hoc* ② 322
 - 8.2.11.1 Procedimentos *post hoc* e erros do Tipo I (α) e Tipo II (β) ② 323
 - 8.2.11.2 Procedimentos *post hoc* e violações das hipóteses do teste ② 323
 - 8.2.11.3 Resumo dos Procedimentos *post hoc* ② 324
- 8.3 Realizando a ANOVA de um fator (One-Way) no SPSS ② 324
 - 8.3.1 Comparações planejadas utilizando o SPSS ② 325
 - 8.3.2 Testes *post hoc* no SPSS ② 327
 - 8.3.3 Opções ② 328
- 8.4 Saídas da ANOVA de um fator (One Way) ② 328
 - 8.4.1 Saídas para a análise principal ② 328
 - 8.4.2 Saídas para as comparações planejadas ② 333
 - 8.4.3 Saídas para os testes *post hoc* ② 334
- 8.5 Calculando o tamanho de efeito ② 338

- 8.6 Relatando resultados da ANOVA independente de um fator ② 339
- 8.7 Violações das hipóteses da ANOVA independente de um fator ② 339
- 8.8 O que descobrimos sobre Estatística? ① 340
- 8.9 Termos-chave que descobrimos 340
- 8.10 Tarefas do Alex Esperto 340
- 8.11 Leituras complementares 341

9 ANÁLISE DE COVARIÂNCIA, ANCOVA (MLG 2) 343

- 9.1 O que você vai aprender neste capítulo? ② 343
- 9.2 O que é ANCOVA ② 343
- 9.3 Executando a ANCOVA no SPSS ② 344
 - 9.3.1 Atribuindo dados ① 344
 - 9.3.2 Análise principal ② 345
 - 9.3.3 Contrastes e outras opções ② 345
- 9.4 Interpretando as saídas da ANCOVA ② 347
 - 9.4.1 Análise principal ② 347
 - 9.4.2 Contrastes ② 351
 - 9.4.3 Interpretando as covariáveis ② 352
- 9.5 ANCOVA realizada como uma Regressão Múltipla ② 353
- 9.6 Suposições adicionais da ANCOVA ③ 357
 - 9.6.1 Homogeneidade dos parâmetros da regressão ③ 357
 - 9.6.2 Testando a homogeneidade dos parâmetros da regressão no SPSS ③ 359
- 9.7 Calculando o tamanho de efeito ② 361
- 9.8 Relatando resultados ② 361
- 9.9 O que descobrimos sobre estatística? ② 362
- 9.10 Termos-chave que descobrimos 362
- 9.11 Tarefas do Alex esperto 362
- 9.12 Leituras complementares 363

10 ANOVA FATORIAL (MLG 3) 364

- 10.1 O que você vai aprender neste capítulo? ② 364
- 10.2 Teoria da ANOVA fatorial (entre grupos) ② 364
 - 10.2.1 Delineamentos fatoriais ② 364
 - 10.2.2 Um exemplo com duas variáveis independentes ② 365
 - 10.2.3 A Soma dos Quadrados (SS_T) ② 366

10.2.4	<i>A Soma dos Quadrados do Modelo</i> (SS_M) ②	367	11.2.4	<i>Qual é o efeito de violar a hipótese de esfericidade</i> ③	397
10.2.4.1	O efeito principal do Gênero (SS_A) ②	367	11.2.5	<i>O que fazer se a hipótese de esfericidade for violada?</i> ②	398
10.2.4.2	O efeito principal do Álcool (SS_B) ②	368	11.3	Teoria da ANOVA de Medidas Repetidas de Um Fator ②	399
10.2.4.3	O efeito interação ($SS_{A:B}$) ②	368	11.3.1	<i>A Soma dos Quadrados Total</i> (SS_T) ②	401
10.2.5	<i>A Soma dos Quadrados dos Resíduos</i> (SS_R) ②	368	11.3.2	<i>A variação dentre participantes</i> (SS_W) ②	401
10.2.6	<i>As razões F</i> ②	369	11.3.3	<i>A Soma dos Quadrados do Modelo</i> (SS_M) ②	402
10.3	ANOVA fatorial utilizando o SPSS ②	370	11.3.4	<i>A Soma dos Quadrados dos Resíduos</i> (SS_R) ②	403
10.3.1	<i>Entrando com dados e utilizando</i>	370	11.3.5	<i>A média dos quadrados</i> ②	403
10.3.2	<i>Modelos customizados</i> ②	371	11.3.6	<i>A razão F</i> ②	403
10.3.3	<i>Ilustrando interações</i> ②	372	11.4	ANOVA de Medidas Repetidas de um Fator utilizando o SPSS ②	404
10.3.4	<i>Contrastes</i> ②	373	11.4.1	<i>A análise principal</i> ②	404
10.3.5	<i>Testes post hoc</i> ②	373	11.4.2	<i>Definindo contrastes para medidas repetidas</i> ②	405
10.3.6	<i>Opções</i> ②	374	11.4.3	<i>Testes post hoc e opções adicionais</i> ③	406
10.4	Saídas da ANOVA fatorial ②	375	11.5	Saídas para a ANOVA de Medidas Repetidas de Um Fator ②	408
10.4.1	<i>Saídas para as análises preliminares</i> ②	375	11.5.1	<i>Descritivas e outros diagnósticos</i> ①	408
10.4.2	<i>O teste de Levene</i> ②	375	11.5.2	<i>Interpretando e corrigindo a esfericidade</i> ②	409
10.4.3	<i>A tabela principal da ANOVA</i> ②	376	11.5.3	<i>A ANOVA principal</i> ②	410
10.4.4	<i>Contrastes</i> ②	378	11.5.4	<i>Contrastes</i> ②	413
10.4.5	<i>Análise post hoc</i> ②	380	11.5.5	<i>Testes post hoc</i> ②	414
10.5	Interpretando gráficos de interações ②	383	11.6	Tamanho de efeito para ANOVA de Medidas Repetidas ③	415
10.6	Calculando o tamanho de efeito ③	387	11.7	Relatando a ANOVA de Medidas Repetidas com Um Fator ②	417
10.7	Relatando os resultados da ANOVA de dois fatores (two way) ②	388	11.8	Medidas Repetidas com várias Variáveis Independentes ②	417
10.8	ANOVA de dois fatores como Regressão ③	389	11.8.1	<i>A análise principal</i> ②	418
10.9	O que descobrimos sobre Estatística? ②	392	11.8.2	<i>Contrastes</i> ②	421
10.10	Termos-chave que descobrimos	393	11.8.3	<i>Ilustrando interações</i> ②	422
10.11	Tarefas do Alex Esperto	393	11.8.4	<i>Outras opções</i> ②	422
10.12	Leituras complementares	394	11.9	Saídas para a ANOVA Fatorial de Medidas Repetidas ②	426
11	DELINEAMENTOS DE MEDIDAS REPETIDAS (MLG 4)	395	11.9.1	<i>Descritivas e análise principal</i> ②	426
11.1	O que você vai aprender neste capítulo? ②	395	11.9.2	<i>O efeito da bebida</i> ②	427
11.2	Introdução aos Delineamentos de Medidas Repetidas ②	395	11.9.3	<i>O efeito da imagem</i> ②	430
11.2.1	<i>A suposição de esfericidade</i> ②	396	11.9.4	<i>O efeito interação (bebida × imagem)</i> ②	431
11.2.2	<i>Como a esfericidade é medida?</i> ②	396			
11.2.3	<i>Interpretando as violações da hipótese de esfericidade</i> ②	397			

11.9.5	<i>Contrastes para variáveis de medidas repetidas</i> ②	433	12.4.5.1	Interação 1 carisma × gênero: alta × algum carisma, homem × mulher ②	457
11.9.5.1	Cerveja <i>versus</i> água, percepção positiva <i>versus</i> neutra ②	434	12.4.5.2	Interação 2 carisma × gênero: sem carisma <i>versus</i> algum carisma, homem <i>versus</i> mulher ②	457
11.9.5.2	Cerveja <i>versus</i> Água, Percepção Negativa <i>versus</i> Neutra ②	435	12.4.6	<i>A interação entre atraente e carismático</i> ②	458
11.9.5.3	Vinho <i>versus</i> água, percepção positiva <i>versus</i> neutra ②	435	12.4.6.1	Interação 1 aparência × carisma: atraente <i>versus</i> mediano, carisma alto <i>versus</i> algum carisma ②	459
11.9.5.4	Vinho <i>versus</i> água, percepção negativa <i>versus</i> neutra ②	436	12.4.6.2	Interação 2 aparência × carisma: atraente <i>versus</i> mediano, sem carisma <i>versus</i> algum carisma ②	460
11.9.5.5	Limitações desses contrastes ②	437	12.4.6.3	Interação 3 aparência × carisma: feio <i>versus</i> mediano, carisma alto <i>versus</i> algum carisma ②	461
11.10	Tamanho de efeito para a ANOVA Fatorial de Medidas Repetidas ③	437	12.4.6.4	Interação 4 aparência × carisma: feio <i>versus</i> mediano, sem carisma <i>versus</i> algum carisma ②	461
11.11	Relatando os resultados da ANOVA Fatorial de Medidas Repetidas ②	438	12.4.7	<i>A interação entre aparência, carisma e gênero</i> ③	462
11.12	O que descobrimos sobre Estatística? ②	439	12.4.7.1	Interação 1 aparência × carisma × gênero: atraente <i>versus</i> mediano, carisma alto <i>versus</i> algum carisma, homem <i>versus</i> mulher ③	464
11.13	Termos-chave que descobrimos	439	12.4.7.2	Interação 2 aparência × carisma × gênero: atraente <i>versus</i> mediano, sem carisma <i>versus</i> algum carisma, homem <i>versus</i> mulher ③	464
11.14	Tarefas do Alex Esperto	440	12.4.7.3	Interação 3 aparência × carisma × gênero: feio <i>versus</i> mediano, carisma alto <i>versus</i> algum carisma, homem <i>versus</i> mulher ③	465
11.15	Leituras complementares	440	12.4.7.4	Interação 4 aparência × carisma × gênero: atraente <i>versus</i> mediano, sem carisma <i>versus</i> algum carisma, homem <i>versus</i> mulher ③	466
12	ANOVA COM DELINEAMENTOS MISTOS (MLG 5)	441	12.4.8	<i>Conclusões</i> ③	467
12.1	O que você vai aprender neste capítulo? ②	441	12.5	Calculando o tamanho de efeito ③	468
12.2	O que homens e mulheres procuram em um companheiro? ②	441	12.6	Relatando os resultados da ANOVA Mista ②	469
12.3	ANOVA mista com o SPSS ②	442			
12.3.1	<i>A análise principal</i> ②	442			
12.3.2	<i>Outras opções</i> ②	445			
12.4	Saídas para a ANOVA mista fatorial: Análise principal ③	446			
12.4.1	<i>O efeito de gênero</i> ②	450			
12.4.2	<i>O efeito da aparência</i> ②	451			
12.4.3	<i>O efeito do carisma</i> ②	454			
12.4.4	<i>A interação entre gênero e aparência</i> ②	454			
12.4.4.1	Interação 1 aparência × gênero: atrativo × mediano, homem × mulher ②	456			
12.4.4.2	Interação 2 aparência × gênero: feio × mediano, homem × mulher ②	456			
12.4.5	<i>A interação entre gênero e carisma</i> ②	456			

- 12.7 O que descobrimos sobre Estatística? ② 471
- 12.8 Termos-chave que descobrimos 472
- 12.9 Tarefas do Alex Esperto 472
- 12.10 Leituras complementares 473
- 13 TESTES NÃO-PARAMÉTRICOS 474**
- 13.1 O que você vai aprender neste capítulo? ① 474
- 13.2 Comparando duas condições independentes: o teste da soma dos postos de Wilcoxon e o Teste de Mann-Whitney ① 475
- 13.2.1 Teoria ② 475
- 13.2.2 Atribuindo dados e análise condicional ① 478
- 13.2.3 Executando a análise ① 478
- 13.2.4 Saídas para o teste de Mann-Whitney ① 481
- 13.2.5 Calculando o tamanho de efeito ② 482
- 13.2.6 Escrevendo os resultados ① 483
- 13.3 Comparando duas condições relacionadas: o teste dos Sinais de Wilcoxon ① 484
- 13.3.1 Teoria do teste dos postos com sinais de Wilcoxon ② 484
- 13.3.2 Executando a análise ① 486
- 13.3.3 Saídas para o grupo do ecstasy ① 488
- 13.3.4 Saídas para o grupo do álcool ① 489
- 13.3.5 Calculando o tamanho de efeito ② 490
- 13.3.6 Escrevendo e interpretando os resultados ① 490
- 13.4 Diferenças entre vários grupos independentes: o teste de Kruskal-Wallis ① 490
- 13.4.1 Teoria do teste de Kruskal-Wallis ② 492
- 13.4.2 Atribuindo dados e fazendo análise condicional ① 493
- 13.4.3 Executando o teste de Kruskal-Wallis no SPSS ① 493
- 13.4.4 Saídas para o teste de Kruskal-Wallis ① 495
- 13.4.5 Análise post hoc para o teste de Kruskal-Wallis ② 496
- 13.4.6 Testes para a tendência: o teste de Jonckheere-Terpstra ② 500
- 13.4.7 Calculando o tamanho de efeito ② 502
- 13.4.8 Escrevendo e interpretando os resultados ① 502
- 13.5 Diferenças entre vários grupos relacionados: a ANOVA de Friedman ① 503
- 13.5.1 Teoria da ANOVA de Friedman ② 504
- 13.5.2 Atribuindo dados e fazendo análise condicional ① 505
- 13.5.3 Executando a ANOVA de Friedman no SPSS ① 505
- 13.5.4 Saídas para a ANOVA de Friedman ① 506
- 13.5.5 Análise post hoc para a ANOVA de Friedman ② 508
- 13.5.6 Calculando o tamanho de efeito ② 510
- 13.5.7 Escrevendo e interpretando os resultados ① 510
- 13.6 O que descobrimos sobre Estatística? ① 511
- 13.7 Termos-chave que descobrimos 512
- 13.8 Tarefas do Alex Esperto 512
- 13.9 Leituras complementares 513
- 14 ANÁLISE DE VARIÂNCIA MULTIVARIADA (MANOVA) 514**
- 14.1 O que você vai aprender neste capítulo? ② 514
- 14.2 Introdução: Semelhanças e Diferenças com a ANOVA ② 514
- 14.2.1 Advertência ② 515
- 14.2.2 Controvérsias ② 515
- 14.2.3 O exemplo do capítulo ② 516
- 14.3 Teoria da MANOVA ③ 517
- 14.3.1 Introdução às matrizes ③ 518
- 14.3.2 Algumas matrizes importantes e suas funções ③ 518
- 14.3.3 Calculando a MANOVA manualmente: um exemplo ③ 519
- 14.3.3.1 ANOVA univariada para VD 1 (Ações) ② 519
- 14.3.3.2 ANOVA univariada para VD 2 (pensamentos) ② 520
- 14.3.3.3 O relacionamento entre VDs: produtos cruzados ② 521
- 14.3.3.4 A matriz SSCP TOTAL (T) ③ 523
- 14.3.3.5 A matriz SSCP Residual (E) ③ 524
- 14.3.3.6 A matriz SSCP do modelo (H) ③ 525

14.3.4	<i>Princípios do teste estatístico MANOVA</i> ④	525
14.3.4.1	A função discriminante ④	526
14.3.4.2	Traço de Pillai-Bartlett (V) ④	528
14.3.4.3	T^2 de Hotelling ④	529
14.3.4.4	Lambda de Wilks (Λ) ④	529
14.3.4.5	Maior raiz de Roy ④	529
14.4	Hipóteses da MANOVA ③	530
14.4.1	Verificando as hipóteses ③	530
14.4.2	Escolhendo o teste estatístico ③	531
14.4.3	Análise adicional ③	531
14.5	MANOVA no SPSS ②	532
14.5.1	A análise principal ②	532
14.5.2	Comparações múltiplas na MANOVA ②	533
14.5.3	Opções adicionais ③	533
14.6	Saídas da MANOVA ③	534
14.6.1	Análise preliminar e checagem de hipóteses ③	534
14.6.2	Teste estatístico MANOVA ③	536
14.6.3	Teste estatístico Univariado ②	536
14.6.4	Matrizes SSCP ③	538
14.6.5	Contrastes ③	540
14.7	MANOVA adicional com Análise Discriminante ③	542
14.8	Saídas da Análise Discriminante ④	545
14.9	Algumas considerações finais ④	548
14.9.1	A interpretação final ④	548
14.9.2	ANOVA univariada ou Análise discriminante? ③	549
14.10	O que descobrimos sobre Estatística? ②	549
14.11	Termos-chave que descobrimos	551
14.12	Tarefas do Alex Esperto	551
14.13	Leituras complementares	552
15	ANÁLISE DE FATORES EXPLORATÓRIA	553
15.1	O que você vai aprender neste capítulo? ②	553
15.2	Fatores ②	553
15.2.1	Representação gráfica dos fatores ②	555
15.2.2	Representação matemática dos fatores ②	556
15.2.3	Escores dos fatores ②	558
15.2.3.1	O método da regressão ④	559
15.2.3.2	Outros métodos ②	560
15.2.3.3	Usos dos escores dos fatores ②	560
15.3	Descobrimos fatores ②	561
15.3.1	Escolhendo um método ②	561
15.3.2	Comunalidades ②	562
15.3.3	Análise de fatores versus Análise de Componentes Principais ②	562
15.3.4	A teoria por trás da Análise de Componentes Principais ③	563
15.3.5	Extração de fatores: autovalores e o diagrama de inclinação ②	564
15.3.6	Melhorando a interpretação: a rotação de fatores ③	566
15.3.6.1	Escolhendo um método de rotação dos fatores ③	568
15.3.6.2	Importância da carga dos fatores ②	568
15.4	Exemplo de pesquisa ②	569
15.4.1	Primeiras considerações ②	570
15.4.1.1	Tamanho da amostra ②	570
15.4.1.2	Vasculhando os dados ②	571
15.5	Executando a análise ②	572
15.5.1	Extração de fatores no SPSS ②	573
15.5.2	Rotação ②	575
15.5.3	Escores ②	576
15.5.4	Opções ②	576
15.6	Interpretando saídas do SPSS ②	577
15.6.1	Análise preliminar ②	577
15.6.2	Extração de fatores ②	581
15.6.3	Rotação de fatores ②	587
15.6.3.1	Rotação ortogonal (Varimax) ②	587
15.6.3.2	Rotação oblíqua ②	589
15.6.4	Escores do fator ②	592
15.6.5	Resumo ②	593
15.7	Análise de confiabilidade ②	593
15.7.1	Medidas de confiabilidade ③	593
15.7.2	Interpretando o α de Cronbach (advertências) ②	594
15.7.3	Análise de confiabilidade no SPSS ②	597
15.7.4	Interpretando a saída ②	598
15.8	O que descobrimos sobre Estatística? ②	602
15.9	Termos-chave que descobrimos	603
15.10	Tarefas do Alex Esperto	604
15.11	Leituras complementares	604

16 DADOS CATEGÓRICOS 606

- 16.1 O que você vai aprender neste capítulo? ① 606
- 16.2 Teoria da Análise de Dados Categóricos ① 606
 - 16.2.1 *Teste qui-quadrado de Pearson* ① 607
 - 16.2.2 *A razão de verossimilhança* ② 608
 - 16.2.3 *Correção de Yates* ② 609
- 16.3 Hipóteses do teste qui-quadrado ① 609
- 16.4 Realizando o teste qui-quadrado no SPSS ① 610
 - 16.4.1 *Entrando com os dados: escores brutos* ① 610
 - 16.4.2 *Entrando com os dados: ponderando casos* ① 610
 - 16.4.3 *Executando a análise* ① 611
 - 16.4.4 *Saída para o teste qui-quadrado* ① 613
 - 16.4.5 *Calculando o tamanho de efeito* ② 616
 - 16.4.6 *Relatando os resultados do qui-quadrado* ① 617
- 16.5 Diversas variáveis categóricas: Análise Loglinear ③ 618
 - 16.5.1 *Qui-quadrado como regressão* ④ 618
 - 16.5.2 *Análise Loglinear* ③ 623
- 16.6 Hipóteses na Análise Loglinear ② 625
- 16.7 Análise Loglinear utilizando o SPSS ② 625
 - 16.7.1 *Considerações iniciais* ② 625
 - 16.7.2 *A Análise Loglinear* ② 626

- 16.8 Saídas da Análise Loglinear ③ 629
- 16.9 Continuação da Análise Loglinear ② 635
- 16.10 Tamanhos de efeito na Análise Loglinear ③ 636
- 16.11 Relatando os resultados da Análise Loglinear ③ 637
- 16.12 O que descobrimos sobre Estatística? ① 637
- 16.13 Termos-chave que descobrimos 638
- 16.14 Tarefas do Alex Esperto 638
- 16.15 Leituras complementares 639

17 EPÍLOGO 640**GLOSSÁRIO** 641**APÊNDICE** 663

- A.1 Tabela da distribuição normal padrão 663
- A.2 Valores críticos da distribuição t 667
- A.3 Valores críticos da distribuição F 668
- A.4 Valores críticos da distribuição qui-quadrado 671
- A.5 O teste F de Welch 671
- A.6 Calculando efeitos simples 671
- A.7 Teste da tendência de Jonckheere 671
- A.8 Capítulo 14 671
- A.9 Cálculos dos coeficientes dos escores dos fatores 671
- Referências 673
- Índice 679

TUDO O QUE VOCÊ SEMPRE QUIS SABER SOBRE ESTATÍSTICA (BEM, QUASE TUDO)

1.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

Sei que muitas pessoas usarão este livro para consultar somente os capítulos que descobrem os testes estatísticos que elas descobriram de uma hora para outra que precisam utilizar. No entanto, para que esses capítulos façam sentido, talvez seja útil saber algumas coisas sobre estatística. Este capítulo é uma tentativa de fornecer um breve panorama de alguns conceitos estatísticos importantes, como o de utilizar modelos estatísticos para responder questões científicas.

1.2 CONSTRUINDO MODELOS ESTATÍSTICOS ①

Nas ciências sociais, normalmente estamos interessados em descobrir algo sobre um fenômeno que acreditamos que realmente exista (um fenômeno do mundo real). Esses fenômenos reais podem ser qualquer coisa desde o comportamento da taxa de juros no mercado econômico até comportamento de universitários numa festa de final de semestre. Qualquer que seja o fenômeno que desejamos explicar, procuramos explicá-lo coletando dados do mundo real e então utilizando esses da-

dos para tirar conclusões sobre o que está sendo estudado. Como estatísticos, nosso trabalho é pegar os dados disponíveis e utilizá-los de uma forma apropriada e isso, frequentemente, envolve construir modelos estatísticos do fenômeno de interesse.

A razão para construirmos modelos estatísticos dos dados do mundo real é melhor explicada por analogia. Imagine que um engenheiro quer construir uma ponte sobre um rio. Esse engenheiro seria bem bobo se construísse qualquer ponte com materiais ultrapassados porque ela teria grandes chances de cair. Em vez disso, ele coleta dados do mundo real: olha para as pontes do mundo real e observa de que material elas são feitas, que estruturas foram utilizadas (ele pode inclusive coletar dados sobre os defeitos que elas apresentam!). Ele, então, usa essa informação para construir um modelo. O engenheiro constrói um modelo, em escala, da ponte do mundo real porque é impraticável, além de caro, construir a ponte real. O modelo será diferente da realidade em muitos aspectos – para começar, ele será menor – mas o engenheiro irá tentar construir o modelo que melhor se adapte ao seu interesse com base nos dados disponíveis. Uma vez que o modelo foi construído, ele pode ser usado para prever coisas sobre o mundo real: por



exemplo, o engenheiro poderá testar se a ponte irá suportar fortes ventos colocando o modelo num túnel de vento. Obviamente, é necessário que o modelo seja uma representação precisa do mundo real. Cientistas sociais fazem a mes-

ma coisa que os engenheiros: constroem modelos de processos do mundo real na tentativa de prever como esses processos operam sob certas circunstâncias. Não temos acesso direto a esses processos, assim, coletamos dados que representam o processo e usamos esses dados para construir modelos estatísticos (reduzimos o processo a um modelo estatístico). Depois, usamos esse modelo estatístico para fazer previsões sobre o fenômeno do mundo real. Assim como o engenheiro, queremos que nosso modelo seja o mais preciso possível, para que possamos garantir que as previsões que fizermos serão também precisas. Entretanto, diferente dos engenheiros, não temos acesso a situações do mundo real e, desse modo, podemos somente *inferir* coisas sobre

processos psicológicos, sociais ou econômicos baseados nos modelos que construímos. Se quisermos que nossas deduções sejam precisas, o modelo estatístico que construímos deve representar os dados que coletamos (os *dados observados*) sempre que possível. O grau com que o modelo estatístico representa os dados coletados é conhecido como *aderência* do modelo e esse é um termo que você irá ver frequentemente.

A Figura 1.1 ilustra os tipos de modelos que um engenheiro pode construir para representar a ponte do mundo real que ele quer criar. O primeiro modelo (a) é uma excelente representação de uma situação do mundo real e é conhecida como uma *boa aderência* (isto é, existem algumas diferenças, mas o modelo é uma boa réplica da realidade). Se esse modelo for utilizado para fazer previsões sobre o mundo real, o engenheiro pode confiar que essas previsões serão precisas, porque o modelo representa bem a realidade. Assim, se o modelo desabar com um vento forte, haverá uma boa chance de que a ponte real desabará também. O segundo modelo (b) tem algumas semelhanças com o mundo real: o modelo inclui algumas características básicas de estrutura, mas

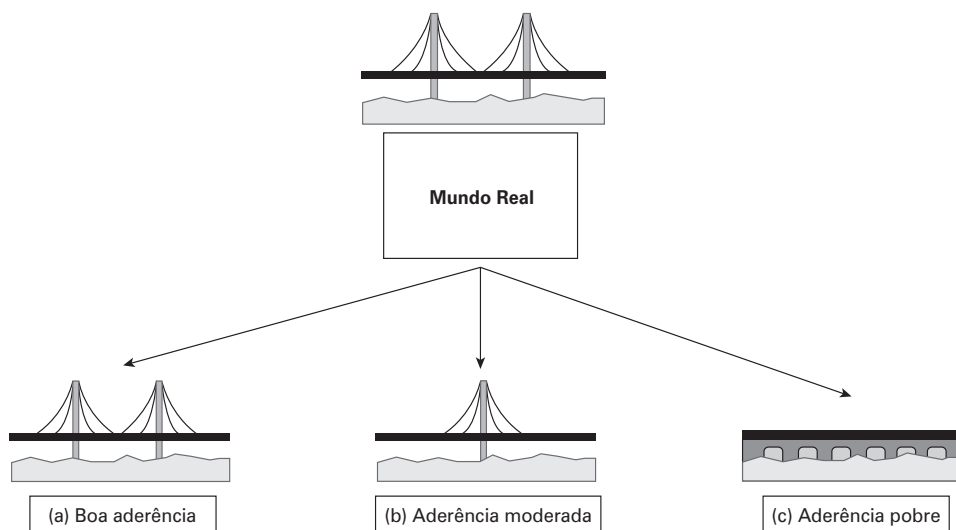


Figura 1.1 Ajustando modelos ao mundo real (veja o texto para detalhes).

há grandes diferenças da ponte do mundo real (isto é, a ausência de uma das torres de suporte). Chamamos isso de *aderência moderada* (isto é, há algumas diferenças entre o modelo e os dados, mas há também grandes similaridades). Se o engenheiro usar esse modelo para fazer previsões sobre o mundo real, essas previsões poderão ser imprecisas e possivelmente catastróficas (por exemplo, se a ponte desabar com ventos fortes isso será devido a ausência da segunda torre de suporte). Portanto, o uso desse modelo resulta em previsões nas quais podemos confiar parcialmente, mas não totalmente. O modelo final (c) é completamente diferente da situação do mundo real. Esse modelo não possui semelhanças estruturais com a ponte real e pode ser chamado de modelo de pouca aderência (na verdade, pode ser mais precisamente descrito como uma aderência abismal!). Assim, qualquer previsão baseada nesse modelo provavelmente será imprecisa. Estendendo essa analogia para as ciências sociais, podemos dizer que é importante quando um modelo estatístico adere bem a um conjunto de dados. **Se o nosso modelo não aderir bem aos dados observados, nossas previsões serão igualmente pobres.**

1.3 POPULAÇÕES E AMOSTRAS ①

Como pesquisadores, estamos interessados em encontrar resultados que se apliquem a toda uma **população** de pessoas ou coisas. Por exemplo, psicólogos querem descobrir processos que ocorrem em todos humanos, biólogos podem estar interessados em processos que ocorrem em todas as células, economistas querem construir modelos que se aplicam a todos os salários, e assim por diante. A população pode ser geral (todos os seres humanos) ou muito pequena (todos os gatos machos de pelo castanho-avermelhado chamados Bob), mas em ambos os casos os cientistas raramente, senão nunca, terão acesso a cada membro de uma população. Psicólogos não podem coletar dados de cada ser humano e ecologistas não podem observar cada gato macho de pelo castanho-avermelhado

chamado Bob. Portanto, coletamos dados de um pequeno subconjunto de uma população (chamado de *amostra*) e usamos essas informações para inferir coisas sobre toda a população. O engenheiro construtor de pontes não pode fazer um modelo de tamanho real da ponte que ele quer construir, assim, ele constrói um modelo em escala menor e testa esse modelo sob várias condições. Dos resultados obtidos do modelo em escala menor, o engenheiro infere coisas sobre como a ponte de tamanho real irá responder. O modelo em escala menor poderá responder diferentemente da versão em escala real, mas, quanto maior o modelo, maior a probabilidade de ele se comportar como a ponte de tamanho real. Essa metáfora pode ser estendida a cientistas sociais. **Nunca temos acesso à população inteira (a ponte de tamanho real), assim, coletamos pequenas amostras (a ponte em escala menor) e usamos o comportamento dentro da amostra para inferir coisas sobre o comportamento da população. Quanto maior a amostra, maior a probabilidade de ela refletir a população inteira. Se pegarmos muitas amostras aleatórias da população, cada uma dessas amostras fornecerá resultados ligeiramente diferentes. Entretanto, em média, resultados de grandes amostras deverão ser bastante similares.**

1.4 MODELOS ESTATÍSTICOS SIMPLES ①

1.4.1 A média, a soma dos quadrados, a variância e o desvio padrão ①

Um dos modelos mais simples usados em estatística é a média. Alguns de vocês podem estranhar pensar na média como um modelo, mas, na verdade, é porque ela representa um resumo dos dados. A **média é um valor hipotético que pode ser calculado para qualquer conjunto de dados; ela não precisa ser um valor realmente observado no conjunto de dados.** Por exemplo, se pegarmos cinco professores de estatística e contarmos o número de amigos que eles têm, poderemos encontrar os seguintes dados: 1, 2, 3, 3 e 4. Se pegarmos

a média do número de amigos, isso poderá ser calculado adicionando os valores obtidos e dividindo pelo número de valores contados: $(1 + 2 + 3 + 3 + 4)/5 = 2,6$. Sabemos que é impossível ter 2,6 amigos (a não ser que você corte alguém com uma serra e seja amigo do seu braço), assim, a média é um valor *hipotético*. Portanto, a média é um modelo criado para resumir nossos dados. Agora, podemos determinar se esse modelo é preciso verificando quão diferente os nossos dados reais são do modelo que criamos. Uma maneira de fazer isso é olhar a diferença entre os dados que observamos e o modelo ajustado. A Figura 1.2 mostra o número de amigos que cada professor de estatística tem e, também, o número médio que anteriormente calculamos. A linha representando a média é o nosso modelo e os círculos são os dados observados. O diagrama tem, também, uma série de linhas verticais que conectam cada valor observado ao valor da média. Essas linhas representam os **desvios** entre os dados observados e o nosso modelo e podem ser pensadas como o erro do modelo. Podemos calcular a magnitude desses desvios simplesmente subtraindo a média (\bar{x}) de cada

um dos valores observados (x_i)¹. Por exemplo, o professor 1 tem somente 1 amigo, portanto, a diferença é: $x_i - \bar{x} = 1 - 2,6 = -1,6$. Você deve notar que o desvio é um número negativo e ele representa o fato de que nosso modelo superestimou a popularidade desse professor: ele prevê que o professor terá 2,6 amigos quando, na verdade, ele tem somente 1 amigo (sorte para ele!). Agora, como podemos usar esses desvios para estimar a precisão do modelo? Uma possibilidade é somar os desvios (isso fornecerá uma estimativa do erro total). Se fizermos isso, acharemos:

Erro total = soma dos desvios

$$= \sum (x_i - \bar{x}) = (-1,6) + (-0,6) + (0,4) + (0,4) + (1,4) = 0$$

O resultado diz que não há um erro total entre nosso modelo e os dados observados,

¹ O x_i simplesmente se refere ao escore observado para a i -ésima pessoa (assim, pode ser trocado por um número que representa um determinado indivíduo). Para esses dados: o professor 1, $x_i = x_1 = 1$; o professor 3, $x_i = x_3 = 3$; o professor 5, $x_i = x_5 = 4$.

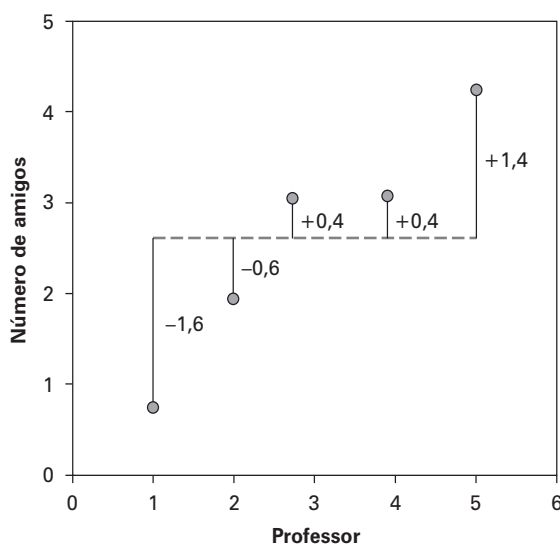


Figura 1.2 Gráfico mostrando as diferenças entre o número observado de amigos que cada professor de estatística tem e o número médio de amigos (linha tracejada).

assim, a média é uma representação perfeita dos dados. Mas isso obviamente não é verdade: houve erros, alguns deles foram positivos e alguns negativos e eles simplesmente cancelaram uns aos outros. Devemos evitar o problema do erro direcionado (isto é, positivo ou negativo) e uma maneira matemática de fazer isso é elevar cada erro ao quadrado², ou seja, multiplicar cada erro por ele mesmo. Assim, em vez de calcularmos a soma dos erros, calculamos a soma dos quadrados dos erros. Neste exemplo:

Soma dos erros

$$\begin{aligned} \text{ao quadrado (SS)} &= \sum (x_i - \bar{x}) (x_i - \bar{x}) \\ &= (-1,6)^2 + (-0,6)^2 \\ &\quad + (0,4)^2 + (0,4)^2 + (1,4)^2 \\ &= 2,56 + 0,36 + 0,16 \\ &\quad + 0,16 + 1,96 = 5,20 \end{aligned}$$

A soma dos erros ao quadrado (SS) é uma boa medida da acurácia do nosso modelo. Contudo, é óbvio que a soma dos erros ao quadrado depende do total de dados que foram coletados – quanto mais dados, maior o SS. Para solucionar esse problema, calculamos a média dos erros dividindo o SS pelo número de observações (N). Se estivermos somente interessados na média do erro para a amostra, podemos dividir apenas por N. Entretanto, geralmente estamos interessados em usar o erro na amostra para estimar o erro na população e, assim, dividimos o SS pelo número de observações menos 1 (a explicação está no Quadro 8.2). Essa medida é conhecida como **variância** e você a usará muito.

$$\begin{aligned} \text{Variância} = s^2 &= \frac{\text{SS}}{N - 1} \\ &= \frac{\sum (x_i - \bar{x})^2}{N - 1} = \frac{5,20}{4} = 1,30 \end{aligned}$$

A variância é, portanto, a média do erro entre a média e as observações feitas (e é a medida de como o modelo corresponde aos dados reais). Existe um problema da variância como

medida: ela é expressa em unidades quadradas (porque colocamos cada erro ao quadrado no cálculo). Em nosso exemplo, teremos que dizer que a média do erro em nossos dados (a variância) foi 1,3 amigos ao quadrado. Se já faz pouco sentido falar de 1,3 amigos, faz menos sentido ainda falar de amigos ao quadrado! Por essa razão, geralmente tiramos a raiz quadrada da variância (o que garante que o erro médio será expresso na mesma unidade da variável). Essa medida é conhecida como **desvio padrão** e é simplesmente a raiz quadrada da variância. Neste exemplo o desvio padrão é:

$$\begin{aligned} \text{Desvio padrão} = s &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}} \\ &= \sqrt{1,30} \\ &= 1,14 \end{aligned}$$

O desvio padrão é, portanto, uma medida de quão bem a média representa os dados. Pequenos desvios padrões (relativos ao valor da própria média) indicam que pontos de dados estão próximos da média. Um desvio padrão grande (relativo à média) indica que os pontos de dados estão distantes da média (isto é, a média não é uma representação precisa dos dados). Um desvio padrão de 0 significaria que todos os escores são os mesmos. A Figura 1.3 mostra um índice geral (em uma escala de cinco pontos) dos dois professores depois de cada uma das cinco aulas. Ambos os professores tiveram uma média de 2,6 em uma escala de 5 pontos. Entretanto, o primeiro professor tem um desvio padrão de 0,55 (relativamente pequeno se comparado à média). Deve ficar claro no gráfico que os índices para esse professor estavam consistentemente próximos do valor da média. Houve uma pequena flutuação, mas em geral suas aulas não variaram em popularidade. Como tal, a média é uma representação precisa dos seus índices. A média teve uma boa aderência aos dados. O segundo professor, entretanto, teve um desvio padrão de 1,82 (relativamente alto comparado à média). Os índices para esse professor estão claramente mais espalhados em torno da média; isto é, para

² Quando você multiplica um número negativo por ele mesmo ele se torna positivo.

algumas aulas ele recebeu índices muito altos e para outras seus índices foram horríveis. Portanto, **a média não é uma representação tão precisa da sua performance porque houve muita variação na popularidade das suas aulas. A média mostrou uma aderência pobre aos dados, nesse caso. Esta ilustração deverá deixar claro por que o desvio padrão é uma medida de quão bem a média representa os dados.**

A discussão das médias, somas dos quadrados e da variância pode parecer um desvio do ponto inicial, mas, na verdade, a média é provavelmente um dos modelos estatísticos mais simples que pode ser ajustado aos dados. O que eu quero dizer com isso? Bem, tudo em estatística se resume a uma equação:

$$\text{Saída}_i = (\text{Modelo})_i + \text{erro}_i \quad (1.1)$$

Isso apenas significa que os dados que observamos podem ser previstos pelo modelo que escolhemos para ajustar os dados mais um erro. Quando eu digo que a média é um modelo estatístico simples, o que quero dizer é que podemos trocar a palavra “modelo” pela palavra “média”, na equação (1.1). Se voltarmos ao nosso exemplo envolvendo o número de amigos que os professores de estatística têm e olharmos para o primeiro professor, por

exemplo, observamos que ele tinha um amigo e a média de todos os professores foi 2,6. Assim, a equação se torna:

$$\begin{aligned} \text{Saída}_{\text{professor1}} &= \bar{X} + \varepsilon_{\text{professor1}} \\ 1 &= 2,6 + \varepsilon_{\text{professor1}} \end{aligned}$$

Disso podemos entender que o erro é $1 - 2,6$, ou $-1,6$. Se substituirmos esse valor dentro da equação, teremos $1 = 2,6 - 1,6$ ou $1 = 1$. Embora isso pareça óbvio, vale ter essa equação em mente ao longo do livro, porque você irá descobrir que a maioria das coisas se resume a uma ideia simples!

Da mesma forma, **a variância e o desvio padrão ilustram outro conceito fundamental: como a aderência de um modelo pode ser medida. Se estivermos interessados em avaliar quão bem o modelo se adequou aos dados (nesse caso nosso modelo é a média), então geralmente olhamos para os desvios do modelo, avaliamos a soma dos erros ao quadrado e, de forma geral, podemos escrever isso como:**

$$\text{Desvio} = \sum (\text{observado} - \text{modelo})^2 \quad (1.2)$$

Ou seja, avaliamos modelos comparando os dados que observamos ao modelo utilizado e elevamos as diferenças observadas ao qua-

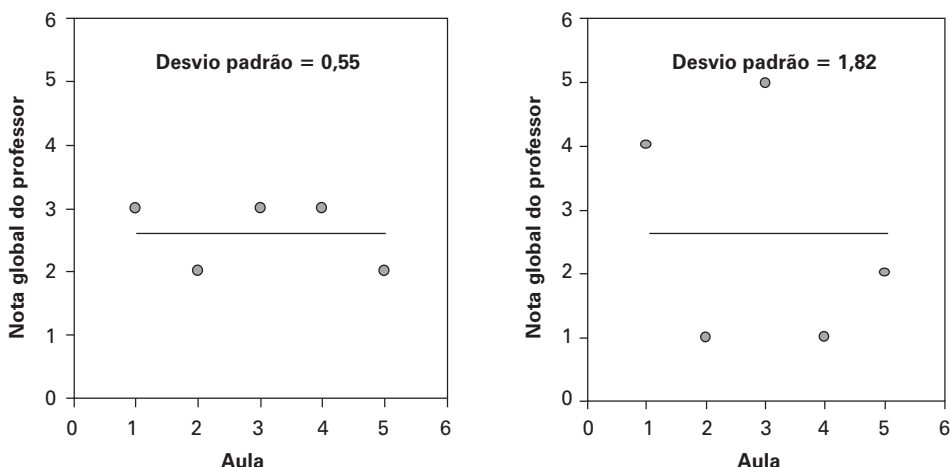


Figura 1.3 Gráfico ilustrando dados que apresentam a mesma medida, mas desvios padrão diferentes.

drado. Essa ideia fundamental se repetirá ao longo de todo o livro.

1.5 DISTRIBUIÇÕES DE FREQUÊNCIAS ①

O que é uma distribuição de frequências e quando ela é normal?



Depois de coletar alguns dados, algo muito útil a fazer é um gráfico de quantas vezes cada escore ocorre. Isso é conhecido como **distribuição de frequências ou histograma**, que é simplesmente um gráfico com os valores observados

no eixo horizontal, com barras mostrando quantas vezes cada valor ocorreu no conjunto de dados. A distribuição de frequências pode ser útil para avaliar as propriedades de um conjunto de valores. Olhando para o escore que apresenta a barra mais alta, podemos imediatamente ver a **moda**, que é simplesmente o escore que ocorre mais frequentemente num conjunto de dados.

Distribuições de frequências ocorrem de muitas formas e tamanhos diferentes. É muito importante, portanto, ter algumas descrições gerais para os tipos mais comuns de distribuições. Em um mundo ideal, nossos dados estariam distribuídos simetricamente em volta do centro de todos os escores. Assim, se traçássemos uma linha vertical pelo centro da distribuição, ela deveria ser a mesma em ambos os lados. Isso é conhecido como **distribuição normal** e é caracterizado por uma curva em forma de sino, a qual você já deve conhecer. Essa forma basicamente sugere que a maioria dos escores está em torno do centro da distribuição (assim, as barras maiores no histograma estão em volta do valor central). Também, à medida que nos distanciamos do centro, as barras ficam menores, sugerindo que à medida que os escores começam a se desviar do centro, sua frequência diminui. Afastando-nos ainda mais do centro a frequência dos nossos escores se torna muito baixa (as barras são muito pequenas). Um exemplo de distribuição normal é mostrado na Figura 1.4.

1.5.1 Propriedades das distribuições de frequências ①

Uma distribuição pode se desviar de uma normal de duas maneiras principais: (1) falta de simetria (chamado de **assimetria**) e (2) achatamento (chamado de **curtose**). Distribuições assimétricas não são pares e, em vez disso, os escores mais frequentes (a parte mais alta do gráfico) estão concentrados em um dos lados da escala. Uma distribuição assimétrica pode ser positivamente assimétrica (a maioria dos escores está concentrada à esquerda da escala) ou negativamente assimétrica (a maioria dos pontos está concentrada à direita da escala). A Figura 1.5 mostra exemplos dessas distribuições.

As distribuições também variam no seu achatamento ou curtose. A curtose, apesar de soar como um tipo de doença exótica refere-se ao grau que os escores estão concentrados na cauda da distribuição. Essa característica é avaliada por um grau de achatamento (achatada ou pontiaguda) da distribuição. Uma distribuição platocúrtica é aquela que têm muitos escores nas caudas (chamada de distribuição com cauda pesada) sendo, portanto, bem achatada. Em contraste, as distribuições leptocúrticas são relativamente finas e parecem bem pontudas. Você pode lembrar disso assim: “distribuições leptocúrticas se elevam no ar enquanto que as distribuições platocúrticas ficam achatadas como um platô (Figura 1.6), ou para os surrealistas entre vocês podem lembrar-se delas como “as distribuições platocúrticas são ‘boas de bico’ e as distribuições leptocúrticas ‘compulsivamente roubam coisas’.” Idealisticamente queremos que nossos dados sejam distribuídos de forma normal (isto é, nem muito assimétricos, nem muito pontiagudos ou achatados!).

Em uma distribuição normal os valores da assimetria e da curtose são 0* (isto é, a dis-

* N de R. T. De fato, o coeficiente de curtose fornece um valor igual a 3 para a curva normal. Centrado o valor (isto é, subtraindo 3), temos o coeficiente de curtose centrado que é igual a zero para a curva normal. É esse valor ao qual o autor está se referindo.

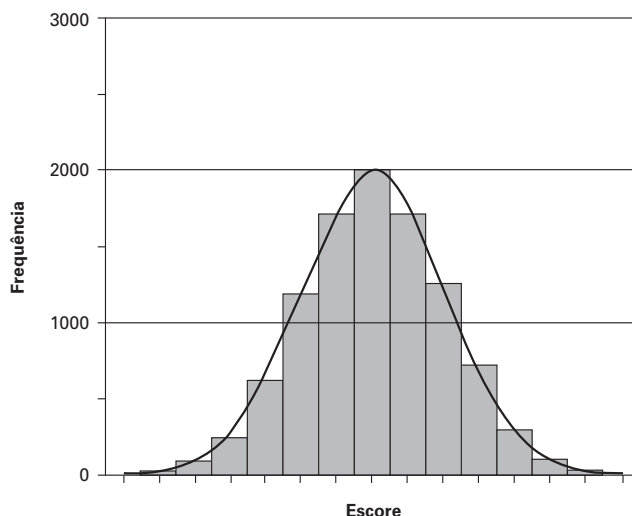


Figura 1.4 Uma distribuição "normal" (a curva mostra a forma idealizada).

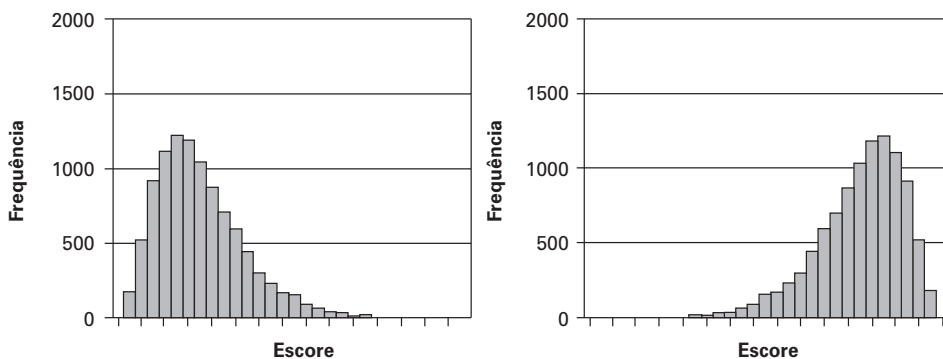


Figura 1.5 Uma distribuição positivamente (esquerda) e negativamente (direita) assimétrica.

tribuição não é muito pontiaguda nem muito achatada e é perfeitamente simétrica). Se a distribuição tem valores de assimetria ou curtose acima ou abaixo de 0, isso indica um desvio da normal (veja o Capítulo 3).

1.5.2 O desvio padrão e a forma da distribuição ①

Assim como fornecem uma ideia da precisão da média como um modelo do nosso conjunto de dados, a variância e o desvio pa-

drão também nos informam sobre a forma da distribuição dos escores. Se a média representa bem os dados, a maioria dos escores irá se concentrar perto da média e o desvio padrão resultante é pequeno quando comparado à média. Quando a média é uma representação ruim dos dados, os escores se espalham mais ao redor da média (lembre-se da Figura 1.3) e o desvio padrão é maior. A Figura 1.7 mostra duas distribuições que tem a mesma média (50), mas desvios padrão diferentes. Uma tem um desvio padrão maior em relação à média

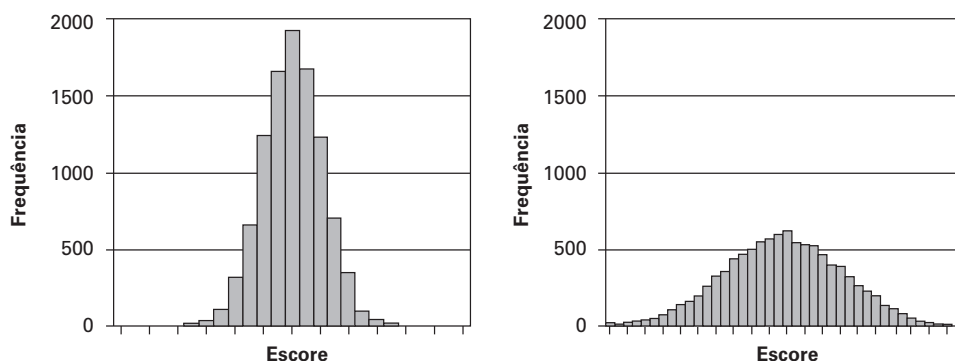


Figura 1.6 Uma distribuição leptocúrtica (esquerda) e platicúrtica (direita).

(DP = 25) e isso resulta em uma distribuição achatada que está mais espalhada, enquanto a outra tem um desvio padrão menor em relação à média (DP = 15), resultando em uma distribuição mais pontiaguda em que os escores próximos da média são mais frequentes, mas escores distantes da média tornam-se muito infreqüentes. A **mensagem principal é que à medida que o desvio padrão fica maior, a distribuição tende a ficar mais achatada.**

1.5.3 O que é a distribuição normal padrão? ①

Outra maneira de olhar as distribuições de frequência é em termos de probabilidade: eles fornecem uma ideia da probabilidade de

um dado escore acontecer. Beachy Head é um rochedo grande e ventoso na costa de Sussex (não muito longe de onde eu moro) e tem uma reputação de atrair suicidas, que parecem gostar de se jogar dele (e após meses e meses reescrevendo este livro, meus pensamentos levam-me, cada vez mais, na direção daquele rochedo tranquilo). A Figura 1.8 mostra a distribuição de frequências de alguns dados anuais inventados sobre o número de suicídios em Beachy Head, cometidos por pessoas de diferentes idades (embora eu tenha inventado esses dados, eles são baseados, grosseiramente, em estatísticas gerais de suicídio como os de Williams, 2001). Houve 171 suicídios no total e você pode ver que a população de suicidas tinha, mais frequentemente, idade entre 25

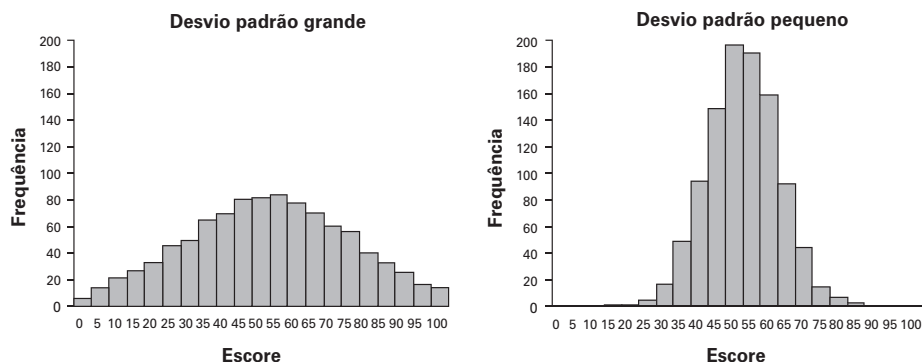


Figura 1.7 Duas distribuições com a mesma média, mas com desvios padrão grande e pequeno.

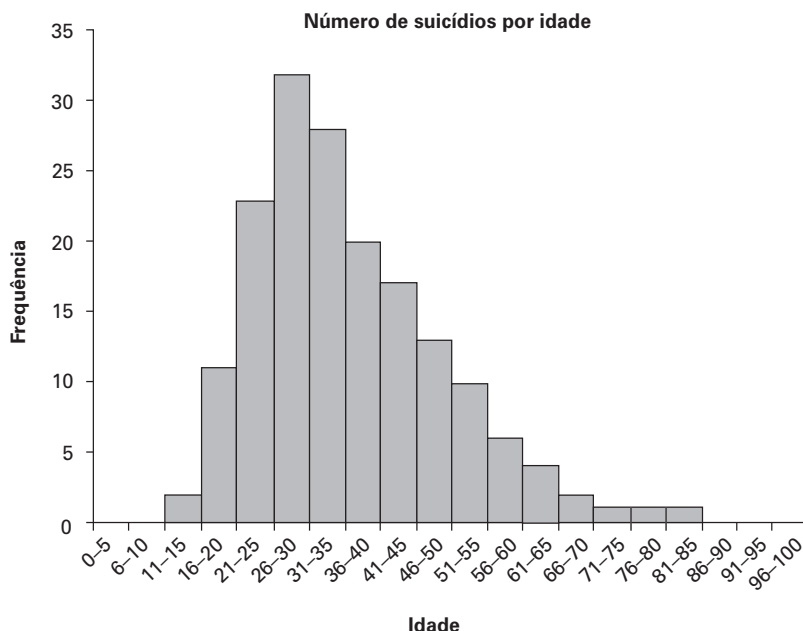


Figura 1.8 Distribuição de frequências mostrando o número de suicídios anuais por faixa etária.

e 30 anos (a barra mais alta), o que é bastante preocupante para mim (embora isso signifique que se eu conseguir passar por este ano a vontade irá diminuir no ano que vem!). O gráfico também nos diz que, por exemplo, poucas pessoas acima de 70 anos cometeram suicídio em Beachy Head.

Afirmo anteriormente que podíamos pensar nas distribuições de frequência em termos de distribuição de probabilidade. Para explicar isso, imagine que alguém pergunte a você o seguinte: “qual é a probabilidade de uma pessoa de 70 anos cometer suicídio em Beachy Head?”. Qual seria a sua resposta? As chances são, se você olhar a distribuição de frequências, de sua resposta ser “pouco provável”, pois há apenas duas pessoas do grupo de 171 suicidas com idade em torno de 70 anos. E se alguém perguntar “qual é a probabilidade de uma pessoa de 30 anos cometer suicídio?”. De novo, olhando o gráfico, você poderá dizer “é bastante provável”, pois 32 dos 171 suicidas eram pessoas com idade em torno de 30 anos (isso é mais do que uma em cada cinco pessoas que cometeram suicídio).

Assim, baseado na frequência de diferentes escores deve começar a ficar claro que podemos usar essa informação para estimar a probabilidade que um determinado escore irá ocorrer. Portanto, poderíamos perguntar, com base em nossos dados, qual é a probabilidade de uma vítima de suicídio ter entre 16 e 20 anos? O valor da probabilidade pode variar de 0 (não há chance alguma de o evento acontecer) a 1 (o evento definitivamente irá acontecer). Assim, por exemplo, quando eu falo com meus editores digo que há a probabilidade de 1 que irei completar a revisão deste livro por volta de junho de 2003. Entretanto, quando falo com outras pessoas, posso mais realisticamente dizer que há a probabilidade de 0,1 de eu terminar as revisões a tempo (ou seja, 10% de chance* ou 1 chance em 10 que eu completarei o livro no tempo previsto). Na verdade, a probabilidade de eu terminar no tempo

* N. de T.: Como você já deve ter notado, o autor utiliza o termo “chance” como sinônimo de probabilidade, o que nem todos o fazem e nem sempre é o caso.

previsto é 0 (nenhuma chance mesmo!) porque nunca consigo terminar na data prevista pelos editores! Se as probabilidades não fazem sentido para você, apenas ignore o ponto decimal e pense nelas como percentagens (isto é, a probabilidade de 0,1 de algo acontecer é o mesmo que 10% de chance de que algo aconteça).

Falei vagamente sobre como distribuições de frequências podem ser usadas para obter uma ideia aproximada da probabilidade de um escore ocorrer. Entretanto, é necessário ser preciso. Para qualquer distribuição de escores poderíamos, teoricamente, calcular a probabilidade de obter um escore de certo tamanho – seria muito entediante e complexo fazê-lo, mas poderíamos. Para poupar nossa sanidade, os estatísticos identificaram muitas distribuições. Para cada uma eles elaboraram uma fórmula matemática que especifica versões idealizadas dessas distribuições (elas estão especificadas em termos de uma curva). Essas distribuições idealizadas são conhecidas como distribuições de probabilidade e a partir delas é possível calcular a probabilidade de conseguir escores baseados nas frequências com que um escore particular ocorre. Uma dessas distribuições é a distribuição normal, já mencionada. Os estatísticos calcularam a probabilidade de certos escores ocorrerem numa distribuição normal com a média 0 e o desvio padrão de 1. Portanto, se tivermos qualquer conjunto de dados que possua a forma de uma distribuição normal e se a média e o desvio padrão são 0 e 1, respectivamente, podemos usar a tabela de probabilidades para a distribuição normal a fim de ver qual é a probabilidade de um determinado escore ocorrer nos dados (fiz uma tabela no Apêndice deste livro)*.

O problema óbvio é que nem todos os dados que coletamos terão uma média 0 e o desvio padrão 1! Por exemplo, podemos ter um conjunto que tem uma média de 567 e um desvio padrão de 52,98! Felizmente, qualquer

conjunto de dados pode ser convertido em um conjunto que tenha a média 0 e o desvio padrão 1. Primeiro, para centrar os dados em zero, pegamos cada escore e subtraímos dele a média de todos os escores. Depois, dividimos o escore resultante pelo desvio padrão para assegurar que os resultados terão um desvio padrão de 1. Os escores resultantes são conhecidos como escores- z e, na forma de equação, a conversão que acabei de descrever é:

$$z = \frac{X - \bar{X}}{s}$$

A tabela das probabilidades que foi calculada para a distribuição normal padrão é mostrada no Apêndice. Por que essa tabela é importante? Bem, se olharmos os dados sobre suicídio, podemos responder a pergunta “qual é a probabilidade de alguém que se jogou de Beachy Head ter 70 anos ou mais?”. Primeiro, convertamos 70 em um escore- z . Digamos que a média dos escores de suicídio foi 36 e o desvio padrão 13. O valor 70 irá tornar-se $(70 - 36)/13 = 2,62$. Depois, olhamos esse valor na coluna chamada “porções menores” (isto é, a área acima do valor 2,62). Devemos encontrar que a probabilidade é 0,0044, ou seja, somente 0,44% de chance que uma vítima de suicídio tenha 70 anos ou mais. Olhando a coluna chamada de “porções maiores”, também podemos ver a probabilidade de que a vítima de suicídio tenha 70 anos ou menos! Essa probabilidade é 0,9956, ou seja, há 99,56% de chance de que uma vítima de suicídio tenha menos de 70 anos!

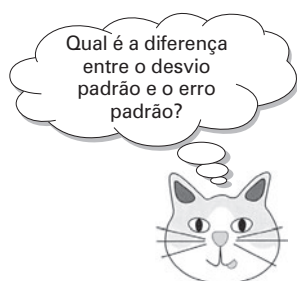
Espero que você possa ver a partir desses exemplos que a distribuição normal e os escores- z nos permitem ir, um passo além dos nossos dados, no sentido que de um conjunto de escores podemos calcular a probabilidade que um determinado escore irá acontecer. Assim, podemos ver se os escores de certo tamanho podem ou não ocorrer em uma distribuição de um tipo particular. Você verá quão útil isso é em seu devido tempo, mas vale mencionar, neste estágio, que certos escores- z são particularmente importantes. Isso porque seus valores são pontos de corte destacando percentagens importantes da distribuição. O primeiro

* N. de T.: O autor está falando de uma forma bastante informal, pois, de fato, a probabilidade de um valor específico ocorrer em uma distribuição normal é zero. A probabilidade só é diferente de zero em um intervalo de valores. Note que ele, ao calcular, utiliza intervalos (70 ou mais e 70 ou menos).

valor importante de z é 1,96, pois ele separa os 2,5% do topo da distribuição e o seu oposto ($-1,96$) destaca os 2,5% da cauda inferior da distribuição. Assim, tomados juntos, esses valores destacam 5% dos escores ou, colocando de outra maneira, 95% dos escores- z estão entre $-1,96$ e $+1,96$. Os outros dois importantes pontos de referência são os valores $\pm 2,58$ e $\pm 3,29$, que destacam 1% e 0,1% dos escores, respectivamente. Ou seja, 99% dos escores- z estão entre $-2,58$ e $2,58$ e 99,9% deles estão entre $-3,29$ e $3,29$. Lembre-se desses valores porque eles irão aparecer com frequência!

1.6 A MINHA AMOSTRA É REPRESENTATIVA DA POPULAÇÃO? ①

1.6.1 O erro padrão ①



Vimos que o desvio padrão e as distribuições das frequências nos dizem algo sobre quão bem a média representa aqueles dados, mas mencionei anteriormente que em geral coletamos dados de

amostras em vez de toda uma população. Também falei que se você retirar várias amostras de uma população, elas irão diferir um pouco uma da outra. Portanto, é **também importante saber quão bem uma amostra em particular representa a população. Aqui é onde usamos o erro padrão. Muitos estudantes não entendem a diferença entre o desvio padrão e o erro padrão (em geral porque essa diferença nunca é explicada claramente). Contudo, o erro padrão é um conceito importante que deve ser entendido, portanto, farei o possível para explicá-lo.**

Já aprendemos que cientistas sociais usam amostras como uma forma de estimar o comportamento de uma população. Imagine que estamos interessados nos índices de todos os professores (assim, professores em geral é a população). Podemos pegar uma amostra dessa população. Quando alguém toma uma

amostra da população, está pegando uma de muitas amostras possíveis. **Se pegarmos muitas amostras de uma mesma população, cada amostra terá a sua própria média e em várias dessas amostras as médias serão diferentes (nem toda amostra terá a mesma média).**

A Figura 1.9 ilustra o processo de selecionar amostras de uma população. Imagine que podemos ter os índices de todos os professores do planeta e que o índice médio é 3 (essa é a média da população). É claro que não podemos coletar os índices de todos os professores, portanto, usamos uma amostra. Para cada uma dessas amostras podemos calcular uma medida de posição ou média da amostra. Vamos imaginar que selecionamos nove amostras diferentes (como no diagrama). Você pode ver que algumas das amostras têm a mesma média que a população e algumas têm médias diferentes; a primeira amostra de professores teve média igual a 3, mas a segunda teve média igual a 2. Isso ilustra a variação amostral, isto é, as médias irão variar porque elas contêm elementos diferentes da população; a amostra que por acaso incluir alguns bons professores irá ter uma média maior do que uma amostra que, por acaso, incluir alguns péssimos professores! Podemos representar as médias amostrais como uma distribuição de frequências ou histograma³, exatamente como fiz no diagrama.

Essa distribuição mostra que três amostras tinham a média 3, médias de 2 e 4 ocorreram em duas amostras cada e médias de 1 e 5 ocorreram apenas em uma amostra cada. O resultado final é uma bela distribuição simétrica conhecida como distribuição amostral. Uma distribuição amostral é simplesmente uma distribuição de frequências das médias de todas as amostras de uma mesma população. Teoricamente, selecionaríamos centenas ou milhares de amostras para construir uma distribuição amostral, mas estou usando apenas nove para simplificar o diagrama! A distribuição amostral apresenta o comportamento das amostras da população e você notará que ela está centrada

³ Ele é um gráfico de cada média amostral representada contra o número de amostras que apresentam essa média – ver o capítulo 2 para detalhes.

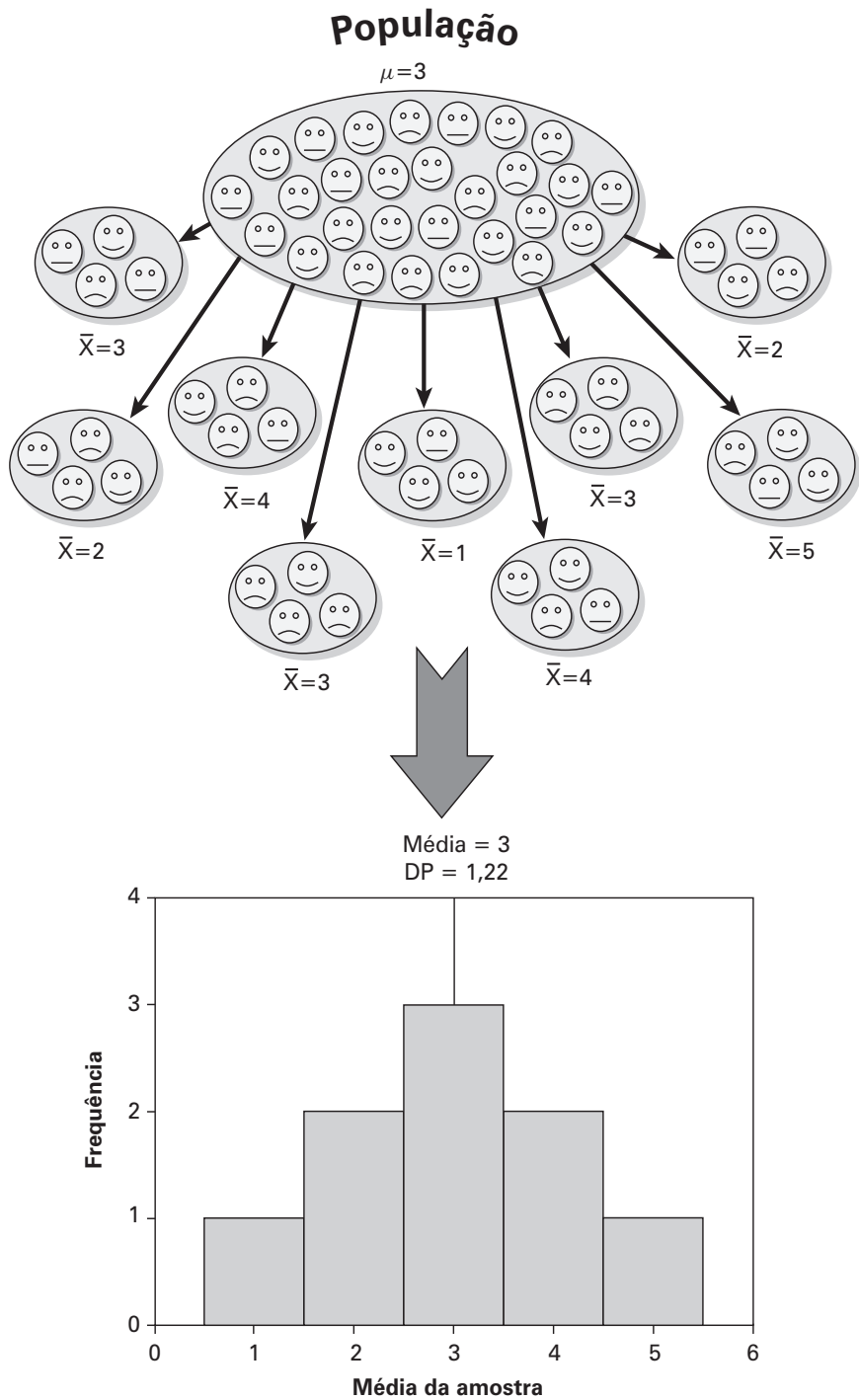


Figura 1.9 Figura ilustrando o erro padrão (veja o texto para detalhes).

no mesmo valor que a média da população (isto é, 3). Isso significa que, se calcularmos a média de todas as médias das amostras, teremos o mesmo valor da média da população. Agora, se a média das médias das amostras é igual à média da população e se conhecermos com precisão aquela média, saberemos algo sobre quão provável é que uma amostra qualquer represente a população. Então, como determinamos a precisão da média da população?

Lembre a discussão sobre o desvio padrão. Usamos o desvio padrão como uma medida de quão representativa a média é dos dados observados. Pequenos desvios padrões representam um cenário no qual a maioria dos dados está próxima da média e um desvio padrão grande representa uma situação na qual os dados estão bem mais espalhados em torno da média. Se você calcular o desvio padrão entre as médias das amostras, isso fornecerá uma medida de quanta variabilidade existe entre as médias de diferentes amostras. O desvio padrão entre as médias das amostras é denominado erro padrão da média (EP). Portanto, o erro padrão poderia ser calculado fazendo a diferença entre cada média da amostra e a média geral (média das médias de todas as amostras ou média da população), elevando ao quadrado essas diferenças, somando-as, e, então, dividindo-as pelo número de amostras.

É claro que, na realidade, não podemos selecionar centenas de amostras e, assim, nos baseamos numa aproximação do erro padrão (felizmente, muitos estatísticos espertos calcularam maneiras em que o erro padrão pode ser

estimado a partir do desvio padrão da amostra). Felizmente, não precisamos entender por que essa aproximação funciona. Podemos apenas confiar que essas pessoas são inteligentes e sabem do que estão falando. O erro padrão pode ser calculado dividindo o desvio padrão da amostra (s) pela raiz quadrada do tamanho da amostra (N):⁴

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{N}}$$

1.6.2 Intervalos de confiança ②

Lembre que normalmente estamos interessados em utilizar a média da amostra como uma estimativa do valor da média verdadeira (isto é, da média da população). Acabamos de ver que amostras diferentes fornecerão valores diferentes da média e que podemos usar o erro padrão para ter uma ideia da extensão da diferença da média da amostra. Uma abordagem diferente para determinar a precisão da média da amostra como uma estimativa da média da população é calcular os limites entre os quais acreditamos que o valor da média verdadeira estará. Tais limites são chamados de intervalos de confiança. A ideia básica por trás dos intervalos de confiança é construir uma gama de valores dentro dos quais achamos que o valor da população estará.

⁴ De fato, ele deve ser o desvio padrão da população (σ), que é dividido pela raiz quadrada do tamanho da amostra; no entanto, para amostras grandes essa é uma aproximação bastante boa.

Dica da Samanta Ferrinho



Resumindo, o erro padrão é o desvio padrão das médias das amostras. Como tal, ele é uma medida de quão representativa a amostra poderá ser da população. Um erro padrão grande (comparado à média da amostra) informa que existe muita variabilidade entre as médias das diferentes amostras e, dessa forma, a amostra que temos pode não ser representativa da população. Um erro padrão pequeno indica que muitas médias amostrais são similares (estão próximas) à média da população e, assim, a nossa amostra será provavelmente uma boa representação da população.

Vamos imaginar um exemplo: Domjan, Blesbois e Williams (1998) examinaram a liberação de esperma de uma codorna japonesa. A ideia básica é que se a codorna pode copular com a fêmea em um certo contexto (em uma câmara experimental), então esse contexto irá servir como uma deixa para a cópula e isso irá afetar a liberação do sêmen (embora durante a fase de testes a pobre codorna foi induzida a copular com um pedaço de pano com uma cabeça embalsamada de uma fêmea fincada no topo – doentio, não?). De qualquer modo, se olharmos a quantidade média de espermatozoides liberado na câmara experimental, existe uma média verdadeira (a média na população); vamos imaginar que são 15 milhões de espermatozoides. Agora, na nossa amostra real, poderemos constatar que a quantidade média de espermatozoide liberado foi de 17 milhões. Como não conhecemos a média verdadeira, realmente não sabemos se o valor da nossa amostra de 17 milhões é uma estimativa boa ou ruim desse valor. O que podemos fazer é usar um intervalo de estimativas: usamos nosso valor amostral como ponto do meio, mas determinamos um limite inferior e um superior. Assim, podemos dizer que o valor real da média de espermatozoide liberado é algo entre 12 e 22 milhões (note que os 17 milhões estão exatamente entre esses valores). É claro, nesse caso, o valor real (15 milhões) está entre esses limites. Mas e se determinarmos limites menores, por exemplo, se acharmos que os valores estão entre 16 e 18 milhões (novamente, note que os 17 milhões estão no meio)? Nesse caso, o intervalo não contém o valor real da média. Vamos imaginar que você está particularmente interessado no exemplo do espermatozoide da codorna japonesa e repetiu o experimento 50 vezes usando diferentes amostras. Cada vez que você realizou o experimento, construiu um intervalo em volta da média da amostra como acabei de descrever. A Figura 1.10 mostra esse cenário: os círculos representam a média para cada amostra com as linhas saindo de fora deles representando os intervalos para essas médias. O valor real da média (a média da população) é de 15 milhões e é representada pela linha vertical. O primeiro aspecto a ser notado

é que a maioria das médias das amostras é diferente da média real (isso é por causa da variação da amostragem, como foi descrito na seção anterior). Em segundo lugar, embora a maioria dos intervalos contenha a média real (eles atravessam a linha vertical significando que 15 milhões de espermatozoides estão em algum lugar entre os limites mais altos e mais baixos), alguns não contêm.

Até agora, evitei o assunto de como devemos calcular os intervalos. É crucial construir os intervalos de confiança de uma maneira que eles nos informem algo útil. Portanto, nós os calculamos de uma forma que eles tenham certas propriedades: especificamente, que eles nos indiquem a probabilidade de conterem o valor real do que estamos tentando estimar (nesse caso, a média).

Tipicamente, se prestarmos atenção aos intervalos de confiança de 95% e, algumas vezes, aos intervalos de confiança de 99%, veremos que eles têm interpretações semelhantes: são limites construídos para que em certa percentagem das vezes (seja 95% ou 99%) o valor real da média da população esteja dentro desses limites. Assim, quando você tiver um intervalo de confiança de 95% para uma média, pense nele assim: se selecionarmos 100 amostras, calcularmos a média e, depois, determinarmos o intervalo de confiança para aquela média (parecido com a Figura 1.10), 95% dos intervalos de confiança conterão o valor real da média da população.

Para calcular o intervalo de confiança, precisamos saber os limites nos quais 95% das médias estarão. Como calculamos esses limites? Lembre-se da seção 1.5.3, onde eu disse que 1,96 era um valor de z importante (um escore de uma distribuição normal com uma média 0 e um desvio padrão 1) porque 95% dos escores- z estão entre $-1,96$ e $1,96$. Isso significa que se a nossa média da amostra têm média 0 e erro padrão 1, os limites de nosso intervalo de confiança serão $-1,96$ e $+1,96$. Você deve também lembrar que podemos converter escores em escores- z usando esta equação:

$$z = \frac{X - \bar{X}}{s}$$

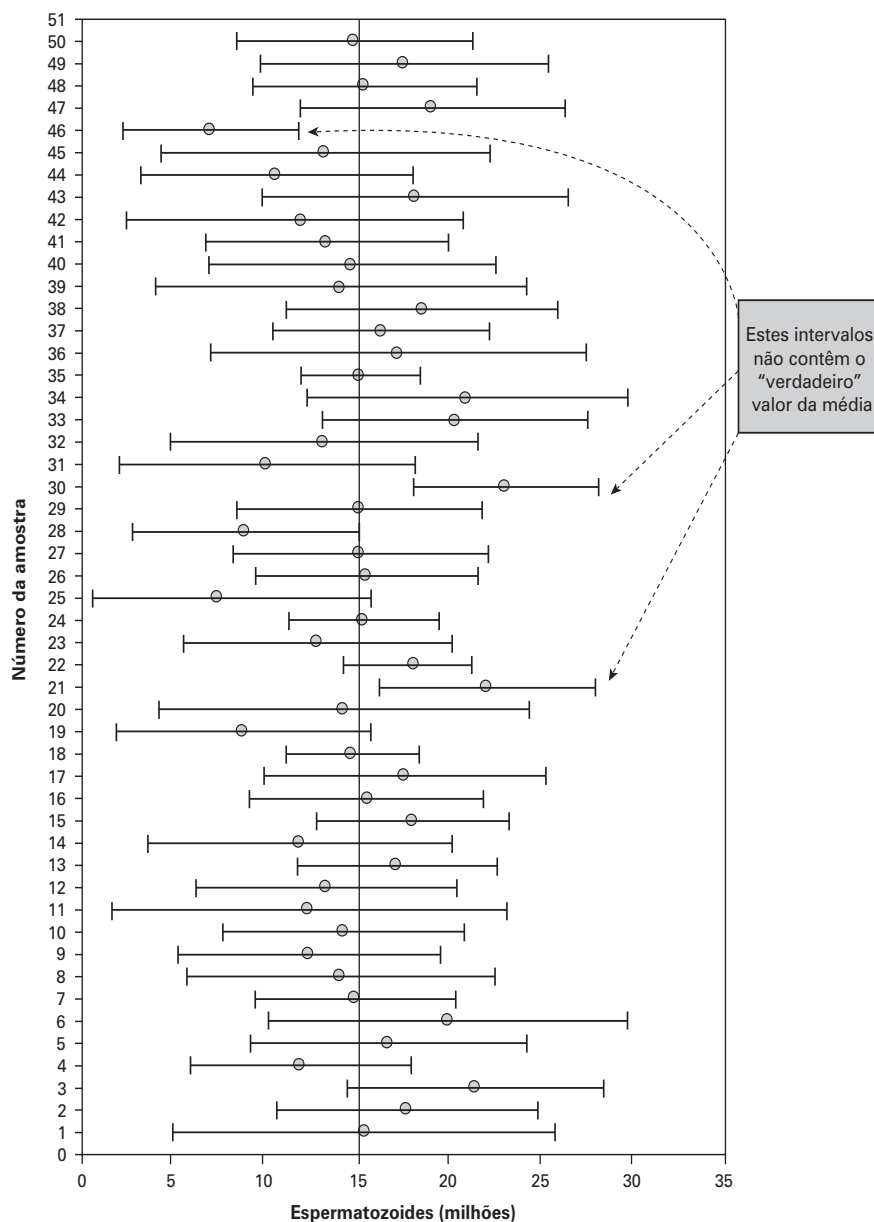


Figura 1.10 Figura mostrando intervalos de confiança da contagem de espermatozoides da cordona japonesa (eixo horizontal) para 50 diferentes amostras (eixo vertical).

Se soubermos que nossos limites serão $-1,96$ e $1,96$ em escores- z , quais são os escores correspondentes em valores dos nossos da-

dos? Para encontrar isso, podemos recolocar z na equação (porque existem dois valores, nós temos duas equações):

$$1,96 = \frac{X - \bar{X}}{s} \quad -1,96 = \frac{X - \bar{X}}{s}$$

O que precisamos saber é o valor de X nessas equações e, para descobrir isso, nós simplesmente as reorganizamos:

$$\begin{aligned} 1,96 \times s &= X - \bar{X} & -1,96 \times s &= X - \bar{X} \\ 1,96 \times s + \bar{X} &= X & -1,96 \times s + \bar{X} &= X \end{aligned}$$

Portanto, o intervalo de confiança pode ser facilmente calculado uma vez que o desvio padrão (s nas equações acima) e a média (\bar{X} nas equações) são conhecidos. Na verdade, usamos o erro padrão e não o desvio padrão porque estamos interessados na variabilidade das médias das amostras e não na variabilidade das observações dentro da amostra. O limite mais baixo do intervalo de confiança é, portanto, a média menos 1,96 vezes o erro padrão e o limite superior é a média mais 1,96 erros padrão.

$$\begin{aligned} \text{Limite inferior do intervalo de confiança} \\ &= \bar{X} - (1,96 \times \text{EP}) \end{aligned}$$

$$\begin{aligned} \text{Limite superior do intervalo de confiança} \\ &= \bar{X} + (1,96 \times \text{EP}) \end{aligned}$$

Assim, a média está sempre no centro do intervalo de confiança. Se a média representa com precisão a média real, o intervalo de confiança deverá ser pequeno. Porque 95% dos intervalos de confiança contem a média real, podemos assumir que esse intervalo de confiança contém a média real: portanto, se o intervalo é pequeno, a média da amostra deve estar perto da média real. Já se o intervalo de confiança é muito grande, então a média da amostra deverá ser diferente da média real, indicando que ela é uma péssima representante da população. Você verá que os intervalos de confiança aparecerão frequentemente ao longo do livro.

1.7 MODELOS LINEARES ①

A média é um exemplo de um modelo estatístico, mas você pode perguntar que outros tipos de modelos estatísticos podem ser construídos. Bem, verdade seja dita, há somente

um modelo geralmente usado, conhecido como modelo linear. Para alguns cientistas sociais, pode não ser óbvio que minha afirmação anterior está correta, mas um estatístico concordaria comigo prontamente. A razão

para isso é que há uma variedade de nomes diferentes dados aos procedimentos estatísticos que são baseados no modelo linear. Um exemplo clássico é que a análise de variância (ANOVA) e a regressão são sistemas idênticos (Cohen, 1968), mas eles têm nomes diferentes e são usados em contextos diferentes (devido a uma divisão nas filosofias metodológicas – ver Cronbach, 1957).

A palavra linear literalmente significa “relativo a uma linha”, mas em termos estatísticos, a linha referida é uma linha reta. Um modelo linear é, portanto, um modelo baseado sobre uma linha reta; isso significa que geralmente estamos tentando resumir nossos dados observados em termos de uma linha reta. Por exemplo, no capítulo descrevendo regressão, ficará claro que duas variáveis podem ser negativamente relacionadas (isso significa que à medida os valores de uma variável aumentam, valores de outra variável diminuem). Em tais circunstâncias, o relacionamento pode ser resumido por uma linha reta.

Suponha que nós mensuramos quantos capítulos deste livro uma pessoa leu e, então, medimos seu enriquecimento espiritual; podemos representar esses dados hipotéticos na forma de um diagrama de dispersão no qual cada ponto representa um escore individual em ambas variáveis. A Figura 1.11 mostra tal gráfico e também mostra o mesmo gráfico, mas com uma linha que resume o padrão desses dados. Uma terceira versão do diagrama de dispersão é também incluída, mas tem uma curva para resumir o padrão geral dos dados. A Figura 1.11 ilustra como podemos adequar diferentes tipos de modelos aos mesmos dados. Nesse caso podemos usar uma linha reta para



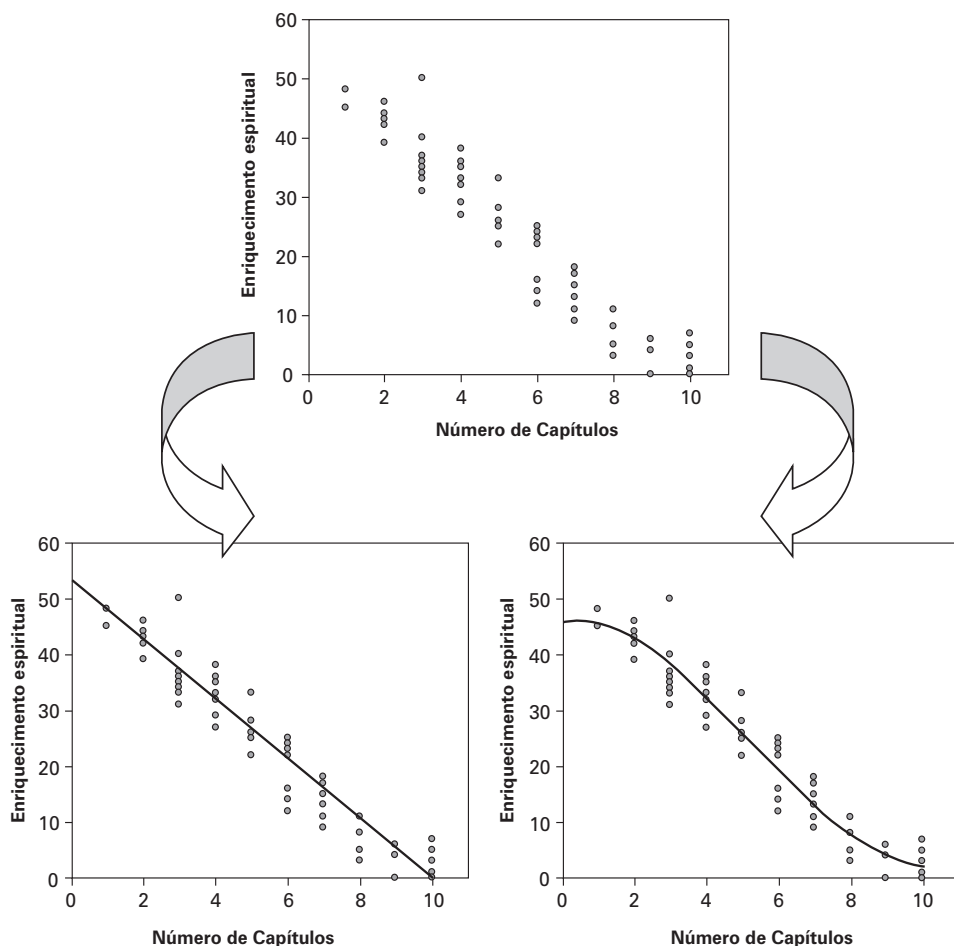


Figura 1.11 Diagramas de dispersão de alguns dados sem um modelo, com um modelo linear e um com um modelo não-linear ajustados.

representar nossos dados e ela irá mostrar que quanto mais capítulos uma pessoa ler, menor será o seu enriquecimento espiritual. Entretanto, também podemos usar uma linha curva para resumir os dados e isso mostrará que quando a maioria ou todos os capítulos foram lidos, o enriquecimento espiritual aumentará levemente (provavelmente porque uma vez que o livro for lido, tudo repentinamente faz sentido – imagine!). Nenhum dos dois tipos de modelos é necessariamente correto, mas existem casos em que um modelo representa melhor os dados

do que outro. Assim, quando usamos modelos estatísticos é importante avaliar quão bem um determinado modelo adere aos dados.

A maioria da estatística usada nas ciências sociais é baseada em modelos lineares, o que significa que tentamos adequar modelos de linhas retas aos dados coletados. Isso é interessante porque a maioria dos estudos científicos publicados são aqueles com resultados estatísticos significantes. Dado que a maioria dos cientistas sociais aprende a usar somente técnicas baseadas no modelo linear, os resulta-

dos publicados serão aqueles que utilizarem os modelos lineares com sucesso. Dados que se ajustam a padrões não-lineares provavelmente serão erroneamente ignorados (porque o modelo errado teria sido aplicado aos dados, levando a resultados não-significantes). Por essa razão, bons pesquisadores primeiro representam seus dados graficamente; diagramas dos dados nos dizem muito sobre quais modelos podem ser aplicados aos dados. É possível, portanto, que algumas áreas da ciência estejam sendo tendenciosas; **se você coletar dados que parecerem não-lineares, por que não corrigir a tendência e investigar técnicas estatísticas diferentes (o que é fácil para eu dizer quando não coloco tais técnicas no meu livro!)?**

1.8 COMO DESCOBRIR SE O SEU MODELO REPRESENTA O MUNDO REAL? ①

Até agora, vimos que usamos amostras para estimar o que está acontecendo em uma população grande a qual não temos acesso. Vimos, também, que é importante estabelecer se um modelo é bom ou ruim para os dados e se

ele é representativo da população. Expliquei tudo isso usando a média como um exemplo, mas o que acontece quando nosso modelo é mais complicado? Cientistas estão geralmente interessados em cenários mais complexos, como “existe uma relação entre o monte de bobagens que as pessoas falam e o monte de gelatina de vodka que elas comeram?” ou “a quantidade média de chocolate que eu como quando estou escrevendo o livro é maior do que quando não estou escrevendo?”. Nesses casos, adequamos modelos que são mais complexos: na verdade, estamos detectando efeitos na nossa população e quantificando esses efeitos. **Esta seção explica algumas maneiras de quantificar efeitos e decidir se eles são significativos. Essencialmente, esse é um processo de quatro estágios (veja o Quadro 1.1):**

1. Crie uma hipótese (ou hipóteses) – isso será, geralmente, uma previsão de que algum tipo de efeito existe na população.
2. Colete alguns dados úteis.
3. Ajuste um modelo estatístico aos dados – esse modelo irá testar suas previsões originais.

Quadro 1.1

Fraude em pesquisa ①

O processo de quatro estágios descrito neste capítulo funcionará somente se você seguir os passos sem falcatuas. Suponha que eu quero apostar em quem vencerá a Copa do Mundo de Rúgbi. Como sou inglês, apostarei que a Inglaterra vencerá o torneio. Para fazer isso eu:

1. Faço minhas apostas, escolhendo meu time (Inglaterra) e as chances disponíveis na lotérica (atualmente 6/4).
2. Vejo qual time vence o torneio.
3. Coleta meus ganhos (se a Inglaterra vencer).

Para todos ficarem felizes, esse processo precisa ser justo: as lotéricas determinam suas chances de modo que não paguem muito (o que as torna felizes), mas às vezes paguem algo (para manter os clientes felizes). A lotérica pode oferecer qualquer chance antes de o torneio terminar, mas não pode mudá-las depois que o torneio acabar (ou o último jogo iniciou). Da mesma forma, posso escolher qualquer time antes do torneio, mas não posso mudar de ideia no meio do torneio ou no final do jogo!

(Continua)

Quadro 1.1 (Continuação)

A situação na pesquisa é semelhante: podemos escolher qualquer hipótese (time de rúgbi) que queremos antes de os dados serem coletados, mas não podemos mudar de ideia no meio da coleta dos dados (ou após a sua coleta). Da mesma forma, temos que decidir nosso nível de probabilidade (ou chances de apostas) antes de coletarmos os dados. Se fizermos isso, o processo de quatro estágios funciona. Entretanto, pesquisadores às vezes fraudam. Eles não escrevem suas hipóteses antes de conduzirem seus experimentos e, algumas vezes, eles as trocam quando os dados são coletados (como eu, trocando meu time depois que a Copa do Mundo terminou) ou, pior ainda, decidem sobre elas depois de coletar os dados! Com exceção de alguns procedimentos complicados chamados de testes *post hoc*, isso é fraude. Do mesmo modo, pesquisadores são culpados se escolherem o nível de significância depois de os dados terem sido coletados e analisados, assim como a lotérica que muda as chances depois do torneio.

Cada vez que você trocar sua hipótese ou os detalhes da sua análise, aumenta a chance de encontrar um resultado significativo, mas você também aumenta a probabilidade de publicar resultados que outro pesquisador não conseguirá reproduzir (o que é muito constrangedor). Se, entretanto, você seguir cuidadosamente as regras e fizer seu teste com uma significância de 5%, pelo menos saberá que no máximo um resultado em cada 20 poderá sair errado.

(Agradecimentos a David Hitchin por esse quadro – peça desculpas por torná-lo em um exemplo de rúgbi!).

4. Avalie esse modelo para ver se ele suporta suas previsões iniciais.

Cientistas, em geral, estão interessados em testar hipóteses; isto é, testar as questões científicas que eles criam. Dentro dessas questões há, usualmente, a previsão que um pesquisador fez. Essa previsão é chamada de hipótese experimental (ela é a previsão de que sua manipulação experimental terá algum efeito ou que de certas variáveis irão se relacionar entre si). A possibilidade contrária – que sua previsão está errada e que o efeito previsto não existe – é chamada de hipótese nula. Dois exemplos são:

1. *Hambúrgueres engordam*: a hipótese experimental é que quanto mais hambúrgueres você comer, mais parecido com uma baleia você vai ficar; a hipótese nula é que as pessoas irão ficar gordas independentemente da quantidade de hambúrgueres que comerem.
2. *Queijo dá pesadelos*: a hipótese experimental é que aqueles que comem queijo antes de dormir têm mais pesadelos do que quem não come; a hipótese nula seria que

quem come queijo antes de dormir tem mais ou menos o mesmo número de pesadelos do que aqueles que não. (No caso de você estar interessado, eu frequentemente como queijo antes de dormir e não tenho pesadelos – exceto quando estou escrevendo livros de estatística e sonho que sou atormentado por um número mau.)

Uma grande parte deste livro aborda a estatística inferencial, que informa se a hipótese experimental pode ser verdadeira – ela nos ajuda a confirmar ou rejeitar nossas previsões. *Grosso modo*, aplicamos nosso modelo estatístico aos nossos dados e observamos quão bem ele se encaixa nos mesmos (em termos de quanta variância ele explica). Se ele se ajusta bem aos dados (isto é, explica muito da variância dos escores), assumimos que nossa previsão inicial é verdadeira: aceitamos a hipótese experimental. É claro, nunca podemos ter absoluta certeza de que qualquer hipótese está correta e, portanto, trabalhamos com probabilidades. Mais precisamente, calculamos a probabilidade de que os resultados que obtivemos ocorreram por acaso – à medida que essa probabilidade diminui, confirmamos que a hi-

pótese experimental é correta e que a hipótese nula pode ser rejeitada.

Para ilustrar essa ideia, Fisher⁵(1925) descreve um experimento projetado para testar a afirmação de uma mulher de que ela conseguia determinar, provando uma xícara de chá, se o leite ou o chá foi colocado primeiro na xícara. Fisher pensou que poderia dar à mulher algumas xícaras de chá em que o leite fosse colocado primeiro e outras em que o leite fosse colocado por último e ver se ela conseguiria identificá-las corretamente. A mulher saberia que existiria um número igual de xícaras com o leite colocado em primeiro lugar e por último, mas não saberia a ordem em que foi colocado. Se pegarmos a situação mais simples na qual há somente duas xícaras, a mulher tem 50% de chance de adivinhar corretamente. Se ela realmente adivinhasse corretamente, não poderíamos ter certeza de que ela realmente sabe a diferença entre as xícaras nas quais o leite foi colocado primeiro daquelas em que o leite foi colocado por último, porque a maioria de nós poderia se sair bem nessa tarefa apenas adivinhando. Mas e se complicássemos as coisas colocando seis xícaras? Existem 20 maneiras nas quais essas xícaras podem ser arranjadas e a mulher somente adivinharia a ordem correta uma vez em 20 (ou 5% das vezes). Se ela acertasse a ordem correta, estaríamos provavelmente muito confiantes de que ela poderia, realmente, perceber a diferença (e reverenciá-la por seu apurado paladar). Se você quer saber mais sobre Fisher e suas palhaçadas de degustação de chá, veja Field e Hole (2003), mas, para os nossos propósitos, o ponto é que somente quando havia uma pequena probabilidade de a mulher completar a tarefa do chá somente por sorte concluiríamos que ela tinha uma habilidade genuína para detectar se o leite fora colocado na xícara antes ou depois do chá.

Não é coincidência eu ter escolhido o exemplo das seis xícaras acima (onde a provadora de chá tinha 5% de chance de acertar a

tarefa por adivinhação), porque Fisher sugeriu que somente quando estamos 95% certos de que um resultado é genuíno (isto é, não resultante do acaso) devemos aceitá-lo como verdadeiro.

⁶ Ou seja, se há somente 5% de probabilidade de algo acontecer por acaso, podemos aceitar que é uma descoberta verdadeira – dizemos que é uma descoberta *estatisticamente significativa* (veja o Quadro 1.2 para entender melhor o que estatisticamente significativo quer dizer). Esse critério de 95% de confiança forma a base da moderna estatística, mas existe pouca justificativa para esse valor além do que Fisher disse; no entanto, como ele foi um sujeito muito inteligente, confiaremos no seu julgamento (Figura 1.12). Todavia, algumas vezes quando olho para a minha estante cheia de resultados experimentais que tiveram uma significância de somente 93% de probabilidade, imagino quão diferente minha carreira seria se Fisher tivesse acordado aquele dia com um humor tipo 90%.

Periódicos de pesquisas têm predisposição a publicar resultados positivos (nos quais a hi-

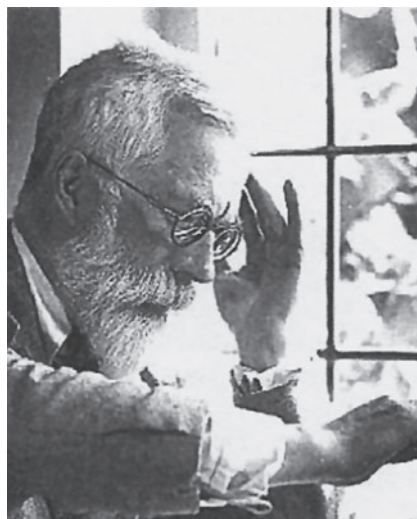


Figura 1.12 Ronald Fisher contemplando as consequências de estabelecer $p = 0,05$.

⁵ Para saber mais sobre Fisher, visite o site <http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafrader.htm>.

⁶ É claro, na realidade, ela pode não ser verdadeira – estamos apenas preparados para acreditar que ela seja!

Quadro 1.2

O que podemos ou não concluir de um teste de significância estatística ②

1. **A importância de um efeito:** Já vimos que a ideia básica por trás do teste de hipóteses envolve a geração de uma hipótese experimental e de uma hipótese nula, ajustar um modelo estatístico e avaliar aquele modelo com uma estatística teste. Se a probabilidade de obter o valor da nossa estatística teste por acaso for menor do que 0,05, então, geralmente aceitamos a hipótese experimental como verdadeira: há um efeito na população. Normalmente dizemos “existe um efeito *significativo* de...”. Contudo, não seja enganado pela palavra “significativo”, porque mesmo que a probabilidade do nosso efeito ter ocorrido por acaso seja pequena (menor do que 0,05), isso não quer dizer que o efeito é importante. Efeitos muito pequenos não importantes podem se tornar estatisticamente significativos somente porque um grande número de pessoas foi usado no experimento (veja Field e Hole, p. 60).
2. **Resultados não-significativos:** Uma vez que você calculou sua estatística teste, você calcula a probabilidade de que ela tenha ocorrido por acaso; se essa probabilidade é maior do que 0,05 você rejeita sua hipótese experimental. Contudo, isso *não* significa que a hipótese nula é verdadeira. Lembre que a hipótese nula é que não existe efeito na população. Tudo o que um resultado não-significativo nos diz é que o efeito não é grande o suficiente para não ser outra coisa do que um resultado causal – ele não nos diz que o efeito é 0. Como Cohen (1990) destaca, um resultado não-significativo nunca deveria ser interpretado (apesar de sempre ser) como “nenhuma diferença entre médias” ou “nenhuma relação entre variáveis”. Cohen também destaca que hipótese nula *nunca* é verdadeira porque sabemos pela distribuição de amostras (veja a Seção 1.6) que amostras ao acaso terão médias ligeiramente diferentes e, mesmo que essas diferenças sejam muito pequenas (isto é, uma média pode ser 10 e a outra 10.000001), elas são diferentes. Na verdade, mesmo uma pequena diferença seria considerada estatisticamente significante se uma amostra grande o suficiente fosse utilizada. Assim, testes de significância nunca nos dizem que a hipótese nula é verdadeira, porque ela nunca é!
3. **Resultados significativos:** OK, podemos não aceitar a hipótese nula como sendo verdadeira, mas podemos pelo menos concluir que ela é falsa quando nossos resultados são significativos, certo? Errado! Uma estatística teste significativa é baseada em uma argumentação probabilística, com grandes limites do que podemos concluir. Novamente, Cohen (1994), que era um ótimo escritor de estatística, destaca que uma argumentação formal se baseia em uma afirmação inicial de um fato seguida por uma afirmação sobre o atual estado dos assuntos e uma conclusão inferida. Este silogismo ilustra o que eu quero dizer:

- Se um homem não tem braços ele não pode tocar guitarra.
 - Este homem toca guitarra.
 - Portanto, este homem tem braços.

O silogismo inicia com uma afirmação do fato que permite alcançar a conclusão final, porque você pode negar que o homem não tem braços (o anterior), negando que ele não pode tocar guitarra (o seguinte). Uma versão parecida de hipótese nula é:

- Se a hipótese nula é correta, esta estatística teste não pode acontecer.
 - Esta estatística teste aconteceu.
 - Portanto, a hipótese nula é falsa.

Isso é ótimo, mas a hipótese nula não é representada dessa maneira porque ela é baseada em probabilidades. Ela deveria ser colocada da seguinte forma:

- Se a hipótese nula está correta, esta estatística teste é pouco provável.
 - Esta estatística teste aconteceu.
 - Portanto, a hipótese nula é pouco provável.

(*Continua*)

Quadro 1.2 *(Continuação)*

Se voltarmos ao exemplo da guitarra, podemos ter uma afirmação similar:

- Se um homem toca guitarra, provavelmente ele não toca na banda Fugazi (isso é verdade porque há milhares de pessoas que tocam guitarra, mas somente duas que tocam guitarra na banda Fugazi!).
 - Guy Picciotto toca na Fugazi.
 - Portanto, Guy Picciotto provavelmente não toca guitarra.

Isso deve parecer, espero, completamente ridículo – a conclusão está errada porque Guy Picciotto toca guitarra. Isso ilustra um erro comum em testes de hipóteses. De fato, testes de significância nos permitem dizer pouco sobre a hipótese nula.

pótese experimental é apoiada) e, se Fisher tivesse acordado com um humor em 90% aquela manhã, eu teria tido muito mais experimentos “com sucesso” e provavelmente seria vice-reitor da minha universidade agora (hum, talvez tenha sido melhor para minha universidade que ele tenha acordado com um humor de 95%!).

1.8.1 Estatísticas teste ①

Então, como sabemos que o nosso modelo é uma boa representação do que está acontecendo no mundo real? Quando selecionamos dados, esses dados irão variar (vimos como medir essa variância na Seção 1.4.1) Há dois tipos de variância (veja também o Capítulo 7):

- **Variação sistemática:** Essa variação se deve a um efeito genuíno (seja um efeito de um pesquisador fazendo algo a todos os participantes de uma amostra, mas não em outras, ou a variação natural entre conjuntos de variáveis). Você pode pensar nisso como uma variação que pode ser explicada pelo modelo que ajustamos aos dados.
- **Variação não-sistemática:** É uma variação que não se deve ao efeito em que estamos interessados (assim, pode ser devido a diferenças naturais entre pessoas em diferentes amostras, como diferenças de inteligência e motivação). Você pode pensar nisso como uma variação que não pode ser explicada pelo modelo que ajustamos aos dados.

Se estivermos tentando estabelecer se um modelo é uma representação razoável do que

está acontecendo na população, geralmente calculamos uma estatística teste. Uma estatística teste é uma estatística que tem propriedades conhecidas; nós sabemos a frequência com que diferentes valores dessa estatística ocorrem. Sabendo isso, podemos calcular a probabilidade de obter um determinado valor. Uma analogia usada por Field e Hole (2003) é a idade em que as pessoas morrem. Dados passados nos dizem a distribuição da idade de morte. Por exemplo, sabemos que em média os homens morrem por volta dos 75 anos e essa distribuição é bem pesada no topo; isto é, a maioria das pessoas morre com idade acima dos 50 anos e é muito incomum morrer por volta dos 20. Assim, frequências da idade do falecimento em mais velhos são muito altas, mas mais baixas em mais jovens. Desses dados seria possível calcular a probabilidade de alguém morrer com certa idade. Se aleatoriamente pegarmos alguém e perguntarmos sua idade, e for 53, poderíamos dizer qual a probabilidade de ele morrer antes do seu próximo aniversário (mas provavelmente levaríamos um soco!). Também, digamos que encontramos um homem com 110 anos; poderíamos calcular qual é a probabilidade de ele ter vivido tanto (seria uma probabilidade muito pequena porque a maioria das pessoas morre antes de alcançar essa idade). A maneira como usamos estatísticas teste é muito similar: sabemos suas distribuições e isso nos permite, uma vez calculada a estatística teste, descobrir a probabilidade de achar um valor tão grande como o que temos. Assim, se calculássemos a estatística teste e seu

valor fosse 110, poderíamos, então, calcular a probabilidade de obter um valor tão grande.

Assim, como calculamos essas estatísticas testes? Isso depende de qual estatística você está usando (e como você verá mais adiante neste livro, existem muitos: t , F e χ^2 , para listar apenas três). Entretanto, a maioria dessas estatísticas representa essencialmente a mesma coisa:

$$\text{Estatística teste} = \frac{\text{Variância explicada pelo modelo}}{\text{Variância não-explicada pelo modelo}}$$

A forma exata dessa equação muda de teste para teste, mas essencialmente estamos comparando a quantidade da variância explicada pelo modelo que adequamos aos dados à variância que não pode ser explicada pelo modelo (veja os Capítulos 5 e 7 para uma explicação mais detalhada). A razão por que essa proporção é tão útil é intuitiva: **se nosso modelo é bom, esperamos que o número de variâncias que ele pode explicar seja maior do que o número que ele não pode explicar. Nesse caso, a estatística teste será sempre maior do que 1 (mas não necessariamente significativa).**

Dado que sabemos quão frequentemente diferentes valores dos testes estatísticos ocorrem por acaso, uma vez calculada uma estatística teste particular, podemos calcular a probabilidade de conseguir esse valor. Como vimos na seção anterior, esses valores de probabilidade nos dizem quão provável é que nosso modelo, ou efeito, seja genuíno e não somente resultado do acaso. Quanto mais variação nosso modelo explicar (comparado com a variância que ele não pode explicar), maior será o teste estatístico e menor a possibilidade de ele ocorrer por acaso (como nosso homem de 110 anos). Portanto, quanto maior a estatística teste, menor a probabilidade de eles ocorrerem. Quando essa probabilidade cai para abaixo de 0,05 (Critério de Fisher), aceitamos isso como uma confiança suficiente para assumir que a estatística teste é assim grande porque nosso modelo explica um montante suficiente de variações para refletir o que realmente está acontecendo no mundo real (a população). Em

outras palavras, aceitamos nossa hipótese experimental e rejeitamos nossa hipótese nula – contudo, o Quadro 1.2 explica alguns mal-entendidos comuns sobre este processo.

1.8.2 Testes uni e bilaterais ①

Quando usamos testes estatísticos ou temos uma previsão específica do que irá acontecer, como “quanto mais alguém ler este livro maior a sua vontade de matar o autor”, ou realmente não sabemos o que irá acontecer, como “ler mais deste livro poderá aumentar ou diminuir o desejo do leitor de matar o autor”. O primeiro exemplo é direcionado: explicitamente dissemos que as pessoas iriam querer me matar à medida que lessem o livro. Se testarmos essa hipótese estatisticamente, o teste é chamado de **teste unilateral**. A segunda hipótese não é direcionada: afirmamos que o desejo de me matar irá mudar por causa da leitura do livro, mas não dissemos se esse desejo irá aumentar ou diminuir. Se testarmos essa hipótese estatisticamente, esse teste será **bilateral**.

Suponha que queremos descobrir se ler este livro aumenta ou diminui o desejo de me matar. Poderíamos fazer isso (experimentalmente) pegando dois grupos, um que leu o livro e outro que não leu, ou (correlacionalmente) medindo o montante do livro que foi lido e o desejo correspondente de me matar. Se não tivermos uma hipótese direcional, há três possibilidades. (1) As pessoas que leram este livro querem me matar mais do que as que não o leram, assim, a diferença (a média dos que leram o livro menos a média dos que não leram) é positiva. Em correlação, quanto mais você ler o livro, mais você irá querer me matar – uma relação positiva. (2) As pessoas que leram este livro querem me matar menos do que as que não leram, assim, a diferença (a média das pessoas que estão lendo o livro menos a das que não leram) é negativa. Em correlação, quanto mais você ler o livro, menor o seu desejo de me matar – uma relação negativa. (3) Não há diferença no desejo de me matar entre leitores e não-leitores – a média dos leitores menos a média dos não-leitores é exatamente zero. Em correlação, não há relação entre ler este livro e a vontade de me

matar. Essa última opção é uma hipótese nula. A direção da estatística teste (isto é, se é positiva ou negativa) depende se a diferença é positiva ou negativa. Assumindo que há uma diferença positiva ou um relacionamento (ler este livro faz com que você queira me matar), então, para detectar essa diferença temos que levar em conta o fato de que a média para leitores é maior do que para não-leitores (e assim obtemos uma estatística teste positiva). Entretanto, se fizemos uma previsão incorreta e ler este livro faz com que leitores queiram me matar menos, a estatística teste será realmente negativa.

Quais as consequências disso? Bem, se no nível 0,05 precisássemos ter uma estatística teste maior do que, por exemplo, 10, e o que realmente temos é -12 , então rejeitaríamos a hipótese mesmo que uma diferença exista. Para evitar isso, podemos olhar os dois lados (ou caudas) da distribuição de possíveis estatísticas teste. Isso significa que iremos pegar ambas as estatísticas teste, negativas ou positivas. Entretanto, fazer isso tem um preço, pois, para manter nosso critério de probabilidade em 0,05, devemos dividir essa probabilidade entre as duas caudas: assim, temos 0,025 no lado positivo da distribuição e 0,025 no lado negativo. A Figura 1.13 mostra essa situação – as áreas

escurecidas são as áreas acima do necessário para o teste estatístico ao nível 0,025 de significância. Combine as probabilidades em ambos os lados e você terá 0,05 – nosso critério de valor. Agora, se fizemos uma previsão, colocamos todos nossos ovos em uma cesta e olhamos somente para o final da distribuição (ou para o lado positivo ou o negativo, dependendo na direção da previsão que fizemos). Consequentemente, podemos apenas procurar pelo valor de uma estatística teste que ocorreria por acaso com a uma probabilidade de 0,05. Na Figura 1.13, as linhas diagonais mostram a área acima da estatística teste positiva necessária ao nível 0,05 de significância. O ponto importante é que se fizemos uma previsão específica, precisamos de uma estatística teste menor para encontrar um resultado significativo (porque estamos procurando somente em uma cauda), mas se nossa previsão está na direção errada, não conseguiremos detectar o efeito que realmente existe! Nesse contexto, é importante lembrar o que foi dito no Quadro 1.1: você não pode apostar ou mudar sua aposta depois que o torneio terminar – se você não fez a previsão da direção antes de coletar os dados, será tarde demais para prever a direção e reivindicar as vantagens do teste unilateral.

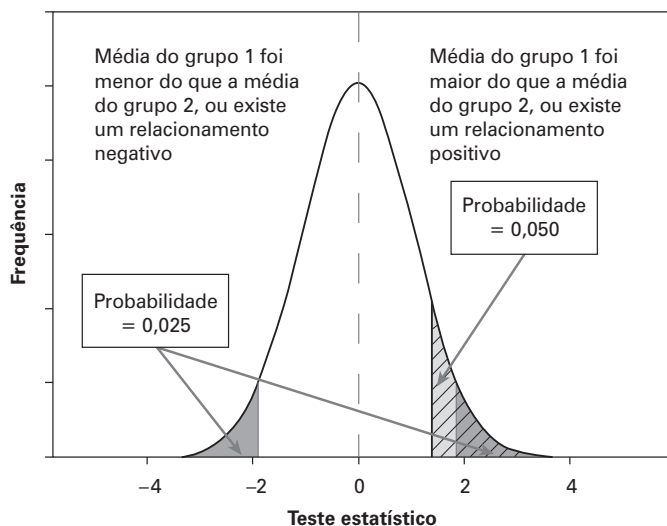


Figura 1.13 Diagrama para mostrar a diferença entre testes uni e bilaterais.

1.8.3 Erros do tipo I e do tipo II ①

Usamos testes estatísticos para nos informar sobre o estado real do mundo (a certo grau de confiança). Especificamente, estamos tentando ver se existe um efeito na nossa população. Temos duas possibilidades no mundo real: existe, na realidade, um efeito na população, ou não existe, na realidade, efeito na população. Não temos como saber quais dessas possibilidades são verdadeiras: entretanto, podemos olhar as estatísticas teste e suas probabilidades associadas para descobrir qual dos dois é mais provável. Obviamente, é importante que sejamos o mais preciso possível, por isso Fisher afirmou que devemos ser muito conservadores e somente acreditar que o resultado é verdadeiro quando estivermos 95% confiantes de que ele é – ou quando houver somente 5% de chance de que os resultados possam ocorrer por acaso. Existem dois erros que podemos cometer: erro do Tipo I e erro do Tipo II. O **erro do Tipo I** ocorre quando acreditamos que há um efeito verdadeiro na nossa população e, de fato, não há. Se usarmos o critério de Fisher, a probabilidade de erro é 0,05 (ou 5%) quando não existe efeito na população – esse valor é conhecido como nível de significância e sua probabilidade é representada por α . Assumindo que não há efeito na população, e se reproduzirmos nossa coleção de dados 100 vezes, poderíamos esperar que em cinco ocasiões obtivéssemos uma estatística teste grande o suficiente para nos fazer pensar que há um efeito genuíno na população mesmo que ele não exista. O oposto é o erro do Tipo II, que ocorre quando acreditamos que não exista um efeito na população, mas na realidade, ele existe. Isso irá ocorrer quando obtivermos um valor da estatística teste pequeno (talvez porque exista muita variação natural entre nossas amostras). Em um mundo ideal, queremos que a probabilidade desse erro seja bem pequena (se existe um efeito na população então é importante que possamos detectá-lo). Cohen (1992) sugere que a probabilidade máxima aceitável para um erro do Tipo II seria 0,2 (ou 20%) – isso é denominado de nível- β . Isso significa que, se tomássemos 100 amostras de dados da população na qual um efeito existe, falharíamos em detectar esse efei-

to em 20 dessas amostras (assim, perderíamos 1 em 5 dos efeitos verdadeiros).

Existe, obviamente, uma troca entre esses dois erros: se diminuirmos a probabilidade de aceitar um efeito como verdadeiro (isto é, tornar α menor), aumentamos a probabilidade de que rejeitaremos um efeito que realmente existe (porque somos muito rígidos sobre o nível no qual aceitaremos um efeito como verdadeiro). A relação exata entre o erro do Tipo I e o erro do Tipo II não é direta porque eles estão baseados em diferentes suposições: para fazer um erro do Tipo I não deve haver efeito na população, enquanto para fazer um **erro do Tipo II** vale o contrário (deve haver um efeito que perdemos). Assim, embora saibamos que a probabilidade de cometer um erro do Tipo I diminui à medida que a probabilidade de cometer um erro do Tipo II aumenta, a natureza exata do relacionamento é geralmente adivinhada pelo pesquisador com base em sua experiência e conhecimento (Howell, 2002, p. 104-107, fornece uma ótima explicação sobre a troca entre erros).

1.8.4 Tamanhos de efeito ②

A estrutura para testar se os efeitos são verdadeiros que acabamos de apresentar apresenta alguns problemas – a maioria deles foi explicada brevemente no Quadro 1.2. O primeiro problema

que encontramos foi saber quão importante um efeito é: apenas porque uma estatística teste é significativa não quer dizer que o efeito que ela mede é significativo ou importante. A solução é medir o tamanho de efeito que estamos testando de uma maneira padronizada. Quando medimos o tamanho de um efeito (seja uma manipulação experimental ou a força do relacionamento entre variáveis), isso é conhecido como tamanho de efeito. O tamanho de efeito é simplesmente uma medida de magnitude padronizada do efeito observado. O fato de que a medida é padronizada apenas significa que



podemos comparar os tamanhos dos efeitos por meio de diferentes estudos que mediram diferentes variáveis ou usaram medidas de escala distintas (assim, um tamanho de efeito baseado em velocidade de milésimos de segundos poderia ser comparado a um tamanho de efeito baseado em batimentos cardíacos). Muitas medidas do tamanho de efeito foram propostas, as mais comuns são o d de Cohen e o coeficiente de correlação r de Pearson (veja Field, 2001 e Capítulo 4). Muitos de vocês estão familiarizados com o coeficiente de correlação como uma medida da força do relacionamento entre duas variáveis (veja o Capítulo 4); contudo, ela é também uma medida muito versátil da força de um efeito experimental. É um pouco difícil aceitar que o humilde coeficiente de correlação também pode ser usado dessa maneira; entretanto, isso é somente porque os estudantes aprendem sobre ele dentro de um contexto de pesquisa não-experimental. Não quero falar sobre isso agora, mas se você ler os Capítulos 4, 7 e 8 entenderá melhor (eu espero!) o que quero dizer. Pessoalmente, prefiro o coeficiente r de correlação de Pearson como uma medida do tamanho de efeito porque ele está limitado ao intervalo entre 0 (sem efeito) e 1 (um efeito perfeito).⁷

Tamanhos de efeito são úteis porque eles dão uma medida objetiva da importância de um efeito. Assim, não importa qual é o efeito que você está procurando, quais as variáveis que foram medidas: sabemos que um coeficiente de correlação 0 significa que não existe efeito e um valor 1 significa que existe um efeito perfeito. Cohen (1988, 1992) estipulou o que é um efeito pequeno ou grande:

- $r = 0,10$ (efeito pequeno): nesse caso, o efeito explica 1% da variância total.
- $r = 0,30$ (efeito médio): o efeito é responsável por 9% da variância total.

- $r = 0,50$ (efeito grande): o efeito é responsável por 25% da variância total.

Podemos usar essas diretrizes para acessar a importância do nosso efeito (independentemente da significância da estatística teste). Entretanto, r não é medido numa escala linear, assim, um efeito com $r = 0,6$ não é duas vezes maior do que um com $r = 0,3$! A utilidade da estimativa do tamanho de efeito é tamanha que a *American Psychological Association* (APA) recomenda que todos os psicólogos informem esses efeitos nos resultados de qualquer trabalho publicado. Portanto, é um hábito que vale a pena ter.

Um ponto final a mencionar é que quando calculamos os tamanhos de efeito, nós os calculamos para uma determinada amostra. Quando olhamos as médias em uma amostra, vemos que as usamos para fazer inferências sobre a média de uma população inteira (que é o valor que realmente estamos interessados). O mesmo é verdadeiro sobre os tamanhos de efeito: o tamanho de efeito na população é o valor no qual estamos interessados, mas porque não temos acesso a esse valor, usamos o tamanho de efeito na amostra para estimar o provável tamanho de efeito na população (veja Field, 2001).

1.8.5 Poder estatístico ②

Vimos que os tamanhos de efeito são uma maneira inestimável de expressar a importância de uma descoberta na pesquisa. O tamanho de efeito em uma população é intrinsecamente ligado a três outras propriedades estatísticas: (1) o tamanho da amostra no qual o tamanho de efeito da amostra é baseado; (2) o nível da probabilidade no qual aceitaremos que um efeito é estatisticamente significativo (nível α); (3) a habilidade de um teste detectar um efeito daquele tamanho (conhecido como o poder estatístico). Assim, se conhecemos três dessas propriedades, podemos calcular a que falta. Irá depender, também, se o teste é uni ou bilateral (veja a Seção 1.8.2). Tipicamente, em psicologia, usamos o nível α em 0,05, assim, já conhecemos esse valor. O poder de um teste é a probabilidade que um determinado teste irá

⁷ O coeficiente de correlação pode ter valores negativos (mas não abaixo de -1), o que é útil quando estamos medindo o relacionamento entre duas variáveis porque o sinal r nos diz sobre a direção do relacionamento, mas em pesquisas experimentais o sinal r meramente reflete a maneira como os pesquisadores codificam seus grupos (veja o Capítulo 4).

encontrar um efeito assumindo que um já exista na população. Se você refletir, irá lembrar que já vimos a probabilidade de falhar em detectar um efeito que realmente existe (β , a probabilidade de um erro Tipo II). Segue que a probabilidade de detectar um efeito, se um existe, deve ser o oposto da probabilidade de não detectar aquele efeito (isto é, $1 - \beta$). Também já mencionei que Cohen (1988, 1992) sugere que deveríamos esperar uma probabilidade 0,2 na falha em detectar um efeito verdadeiro e, assim, o nível correspondente do poder que ele recomendou foi de $1,0 - 0,2 = 0,8$. Nosso objetivo deve ser alcançar um poder de 0,8 ou 80% de probabilidade de detectar um efeito se ele genuinamente existe. O tamanho de efeito na população pode ser estimado do tamanho de efeito em uma amostra e o tamanho da amostra é determinado pelo pesquisador de qualquer modo, assim, o valor é fácil de calcular. Agora, existem duas coisas úteis que podemos fazer sabendo que essas quatro variáveis estão relacionadas:

1. **Calcular o poder de um teste:** Se conduzirmos nosso experimento, já teremos selecionado um valor de α , poderemos estimar o tamanho de efeito baseado na nossa amostra e saberemos quantos participantes usamos. Portanto, podemos usar esses valores para calcular β , o poder de nosso teste. Se esse valor for 0,8 ou mais, podemos garantir que alcançamos poder suficiente para detectar qualquer efeito que poderá ter existido, mas se o valor resultante for menor, deveremos refazer o experimento usando mais participantes para aumentar o poder.
2. **Calcular o tamanho da amostra necessário para alcançar um nível dado de poder:** Dado que sabemos o valor de α e β , podemos utilizar pesquisas anteriores para estimar o tamanho de efeito que esperamos detectar em um experimento. Mesmo que ninguém tenha previamente feito o experimento exato que você pretende fazer, ainda assim podemos estimar a provável tamanho de efeito baseados em experimentos similares. Podemos usar esse tamanho de efeito estimado para cal-

cular quantos participantes necessitamos para detectar esse efeito (baseado nos valores de α e β que escolhemos).

O último uso é o mais comum: para determinar quantos participantes devem ser usados para alcançar o desejado nível de poder. Os cálculos exatos são muito volumosos, mas, felizmente, existem agora programas de computadores disponíveis que os farão para você (um exemplo é o G* Power que é grátis e pode ser encontrado no *site* www.artmed.com.br, outro é o nQuery Adviser – veja Field (1998b) para uma avaliação – mas esse tem que ser comprado!) Também, Cohen (1988) fornece tabelas extensas para calcular o número de participantes para um determinado nível de poder (e vice-versa). Baseado em Cohen (1992), podemos usar as seguintes diretrizes: se tomarmos o nível padrão α de 0,05 e requisitarmos o poder recomendado de 0,8, então precisamos de 783 participantes para detectar um tamanho de efeito pequeno ($r = 0,1$), 85 participantes para detectar um tamanho de efeito médio ($r = 0,3$) e 28 participantes para detectar um tamanho de efeito grande ($r = 0,5$).

1.9 ÚLTIMOS CONSELHOS

Quero terminar este capítulo com alguns conselhos gerais. Procedimentos estatísticos são uma forma de processar números e, portanto, se você colocar baboseiras em uma análise mesmo assim obterá conclusões que tem significado estatístico, mas que provavelmente não terão sentido empírico (prático). Existe uma tentação de ver a estatística como uma forma milagrosa de determinar a verdade, mas a estatística é somente um recurso. Meu professor de estatística costumava dizer que se “você entra com lixo, você vai obter lixo”. Eu nunca havia entendido o que ele estava querendo dizer até que as pessoas começaram a me apresentar conjuntos de dados contendo milhões de variáveis aleatórias aparentemente conectadas que eles queriam que eu as ajudasse a interpretar. Na análise estatística não existe um substituto para o pensamento empírico! Tenha isso em mente ao

longo deste livro. Um último conselho: nunca acredite em algo que um psicólogo diga sobre estatística. 😊

1.10 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Ok, este foi o seu curso rápido de iniciação à teoria da estatística! Espero que seu cérebro ainda esteja relativamente intacto. O ponto principal que quero que você entenda é que quando executa a sua pesquisa, você estará querendo ver se algum efeito verdadeiro existe na população (o efeito procurado depende dos seus interesses de pesquisa e suas previsões específicas). Você não será capaz de coletar dados sobre toda a população (a menos que queira gastar toda a sua vida e provavelmente várias outras coletando dados), portanto, deverá utilizar amostras. Usando os dados da amostra, você ajustará um modelo estatístico para testar suas previsões, ou seja, detectar o efeito de interesse. A estatística resume-se a uma ideia simples: dados observados podem ser previstos a partir de algum tipo de modelo mais um erro associado a tal modelo. O modelo é usado (e normalmente o erro associado a ele) para calcular a estatística teste. Se tal modelo pode explicar grande parte da variação nos dados coletados (a probabilidade de obter essa estatística teste é menor do que 5%), você vai inferir que o efeito de interesse realmente existe na população. Se a probabilidade de obter tal estatística teste é mais do que 5%, você poderá concluir que o efeito é muito pequeno para ser detectado. Em vez de confiar na significância, você pode quantificar o efeito na amostra de uma forma padrão como um *tamanho de efeito* e isso poderá ser útil para determinar a importância de tal efeito. Agora, para o SPSS!

1.11 TERMOS-CHAVE QUE DESCOBRIMOS

- Amostra
- Assimetria
- Curtose
- Desvio padrão

- Desvios
- Distribuição amostral
- Distribuição de probabilidade
- Distribuição normal
- Erro do Tipo I
- Erro do Tipo II
- Erro padrão
- Escore-z
- Estatística teste
- Hipótese nula
- Hipóteses experimentais
- Distribuição de frequências
- Histograma
- Intervalo de confiança
- Leptocúrtica
- Média
- Moda
- Modelo linear
- Nível α
- Nível β
- Platicúrtica
- Poder estatístico
- População
- Soma dos erros ao quadrado (SS)
- Tamanho de efeito
- Teste bilateral
- Teste unilateral
- Variação não-sistemática
- Variação sistemática
- Variância

1.12 QUESTÕES DE REVISÃO DO ALEX ESPERTO



Alex Esperto sabe tudo sobre estatística e SPSS. Ele adora fazer perguntas sobre estatística para os outros a fim de mostrar o quanto ele sabe. Que tal irritá-lo e responder corretamente todas as questões?

1. Por que utilizamos amostras? ①
2. O que é a média e como podemos saber se ela é representativa dos nossos dados? ①
3. Qual é a diferença entre desvio padrão e erro padrão? ①
4. O que é uma estatística teste e o que ela nos informa? ①

5. O que são os erros do Tipo I e do Tipo II ①
6. O que é o tamanho de efeito e como ele é mensurado? ②
7. O que é poder estatístico? ②

Algumas respostas breves podem ser encontradas no arquivo **Answers (Chapter 1).pdf** disponível no *site* www.artmed.com.br.

1.13 LEITURAS COMPLEMENTARES

COHEN, J. Things I have learned (so far).

American Psychologist. v. 45, n. 12, 1990. p. 1304-1312.

COHEN, J. The earth is round ($p < .05$). *American Psychologist*. v. 49, 1994. p. 997-1003. Alguns artigos maravilhosos do melhor escritor moderno de estatística.

FIELD, A. P., Hole, G. J. *How to design and report experiments*. Londres: Sage, 2003. Sou suspeito para falar, mas acredito que esse livro traça um bom panorama da teoria estatística básica.

WRIGHT, D. B. *First steps in statistics*. London: Sage. 2002. Capítulos 1, 4 e 5 são introduções bem claras sobre amostragem, intervalos de confiança e outras ideias estatísticas importantes.

O AMBIENTE DO SPSS

2.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

Espero que o Capítulo 1 não o tenha feito desistir do livro. Se não, o Capítulo 2 lida com outras informações que você precisará saber: o ambiente do SPSS. Existem vários textos excelentes que fornecem uma introdução ao ambiente de uma forma geral dentro do qual o SPSS opera. Os melhores incluem Kinnear e Gray (2000) e Foster (2001). Esses textos valem a pena ser lidos se você não possui familiaridade com computadores e o SPSS. No entanto, eu tenho consideração pela pouca grana de muitos estudantes e assim para tornar este texto útil para aqueles que não têm experiência com o SPSS, esta seção fornece um guia para o ambiente do *software* – mas se você precisa ir mais a fundo então recomendo que veja os textos citados anteriormente e os manuais do SPSS (você pode também achar os arquivos de ajuda (*help*) úteis). Este capítulo trata de janela base do SPSS (o editor de dados – *data editor* e visualizador – *viewer*) e também apresenta como criar variáveis, entrar com dados e ajustar propriedades das variáveis. Terminaremos vendo como carregar arquivos e salvá-los!

2.2 VERSÕES DO SPSS ①

Este livro é baseado principalmente na versão 13.0 do SPSS (pelo menos em termos de diagramas); no entanto, não se deixe impressionar muito pelos números das versões porque a SPSS tem o hábito de lançar “novas” versões regularmente. Embora isso os faça ganhar muito dinheiro e criar um bom mercado para pessoas que escrevem livros sobre o recurso, existem na realidade poucas diferenças entre essas novas versões e a maioria de nós nem irá percebê-las. Ocasionalmente eles apresentam uma grande mudança (a versão 7.0 mostrou uma mudança dramática em relação à versão 6.0 e a versão 10.0 alterou a forma de entrada das variáveis), mas a maior parte do tempo você pode se virar bem com um livro que não cobre explicitamente a versão que você está utilizando (a primeira edição desse livro foi baseado na versão 9, mas pode ser utilizado sem problemas com as versões 7, 8, 10, 11 e 12). Assim esse livro revisado, embora lide com a versão 13.0, poderá ser utilizado com versões anteriores (7.0, 7.5 e 8.0 – existe algumas poucas diferenças entre as versões 7.0, 8.0 e 9.0, mas qualquer diferença óbvia está destacada quando for relevante). Eu também suspeito que ele possa ser utilizado com



as versões 14, 15 e 16 quando elas aparecerem (embora a SPSS possa decidir mudar tudo só para me contrariar!). Existem várias diferenças em termos da entrada de dados com as versões anteriores a 10 assim se estás utilizando a versão 9 ou anterior em inclui um arquivo denominado de **Field2000(Chapter1).pdf** no *site* www.artmed.com.br. Esse arquivo é uma cópia do Capítulo 1 da primeira edição do livro e irá mostrar como entrar com dados nessas versões mais antigas.

2.3 INICIANDO ①

O SPSS utiliza principalmente duas janelas: o editor de dados – **data editor** (que é o local onde se entra com os dados e se executa as funções estatísticas) e o visualizador – **viewer** (que é onde os resultados de qualquer análise irão ser apresentados).¹ Existem várias janelas adicionais que poderão ser ativadas como o editor de sintaxe – **syntax editor** (veja a Seção 2.6), que permite que se entre com comandos manualmente (em vez de se utilizar os menus baseados em janelas). Para a maioria dos níveis de conhecimento a janela de sintaxe é redundante, pois você pode executar a maioria das análises simplesmente com o mouse. No entanto, existem muitas funções adicionais que podem ser acessadas utilizando-se a sintaxe e loucos que gostam de estatística podem encontrar muitos usos para ela. Apenas para provar que eu sou o mais louco de todos, existem umas poucas seções onde eu irei forçá-lo a usá-la. ☺

¹ Em versões do SPSS anteriores a versão 7.0, os gráficos aparecem em uma janela separada conhecida como carrossel gráfico – *chart carousel*; no entanto, as versões 7.0 e posteriores incluem os gráficos na janela de saída, que é denominada de navegador de saída – *output navigator* (versão 7.0) e o visualizador de saída – *output viewer* (versão 8.0 e posteriores).

Uma vez que o SPSS tenha sido ativado, uma janela inicial aparecerá (veja a Figura 2.1), que permitirá a seleção de várias opções.² Se você já tem um arquivo de dados guardado em algum disco que você gostaria de abrir então selecione *Open an existing data source* (Abrir uma fonte de dados existente) clicando no de forma que ele apareça assim: ; essa é a opção por omissão (*default*). No espaço embaixo dessa opção existe uma lista dos arquivos abertos recentemente que você poderá selecionar com o mouse. Para abrir o arquivo selecionado clique em . Se você quiser abrir um arquivo de dados que não está na lista selecione *More files...* (Mais arquivos) com o mouse e clique em . Isso irá abrir a janela de exploração padrão que permitirá a navegação pelos arquivos do computador para tentar encontrar o arquivo desejado (veja a Seção 2.8). Pode, também, ocorrer que você queira abrir algo que não seja um arquivo de dados, por exemplo, um documento *viewer* (visualização) contendo resultados da sua última análise. Você poderá fazer isso selecionando *Open another type of file* (Abrir Outro Tipo de Arquivo) clicando no de forma que ele apareça assim: , e escolhendo um arquivo da lista ou então selecionando *More files...* (Mais arquivos...) e navegando pelo computador. Se você está iniciando uma nova análise (como estamos fazendo aqui) então queremos digitar nossos dados em um novo editor de dados. Para isso precisamos selecionar *Type in data* (Digite os dados) clicando novamente no e então no . Isso irá abrir uma janela de edição de dados (*data editor*) em branco.

2.4 O EDITOR DE DADOS ①

A janela principal do SPSS inclui um editor para a entrada dos dados. Essa janela é onde a maioria da ação acontece. No topo dessa janela existe uma barra de menus semelhante a outras que você já deve ter visto em

² De fato, essa janela não aparece em versões anteriores a 10.0. Em vez disso, um editor de dados em branco é carregado e você poderá abrir os arquivos utilizando os menus (veja a seção 2.8 ou o arquivo **Field2000(Chapter1).pdf** no *site* www.artmed.com.br).

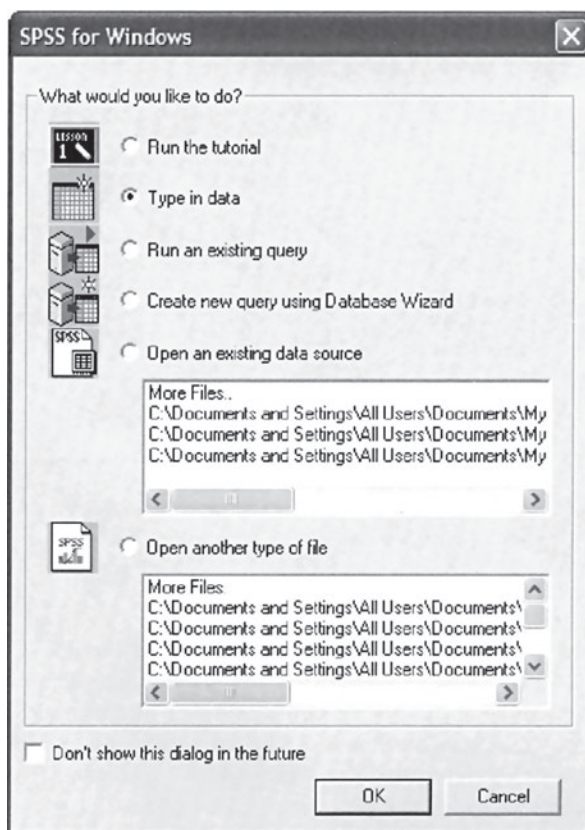


Figura 2.1 A janela inicial do SPSS.

outros programas (tal como o Word). A figura 2.2 mostra essa barra de menu e o editor de dados. Existem vários menus no topo da tela (por exemplo, *File* (Arquivo), *Edit* (Editar), etc.) que podem ser ativados pela utilização do mouse para mover o cursor sobre o menu desejado e então pressionar o botão esquerdo uma vez (pressionar esse botão é normalmente conhecido como ‘clique’). Quando você clica em um menu, uma janela se abre mostrando uma lista de opções que podem ser ativadas movendo o cursor para a opção desejada e então clicando-a. Frequentemente selecionar uma opção do menu faz que uma nova janela apareça, essas janelas são denominadas de “caixas de diálogo” (*dialog boxes*). Quando me referir a opções selecionáveis de um menu eu irei representar a ação utilizando

negrito com uma seta indicando o caminho do mouse (assim, cada seta representa colocar o cursor sobre a palavra e clicar com o botão esquerdo do mouse). Por exemplo, se eu quiser dizer que você deve selecionar a opção *Save As...* (Salvar Como) do menu *File* (Arquivo), escreverei isso como **File⇒Save As...** (**Arquivo⇒Salvar Como**).

O editor de dados tem dois painéis (*views*): o de **dados** e o de **variáveis**. O editor de dados é para entrar obviamente dados e o editor de variáveis permite que definamos várias características das variáveis do editor de dados.³ Na

³ Antes do SPSS versão 10, a entrada de dados era feita utilizando um único painel (veja Field, 2000, Capítulo 1 que está disponível *site* www.artmed.com.br como **Field2000(Chapter1).pdf**).

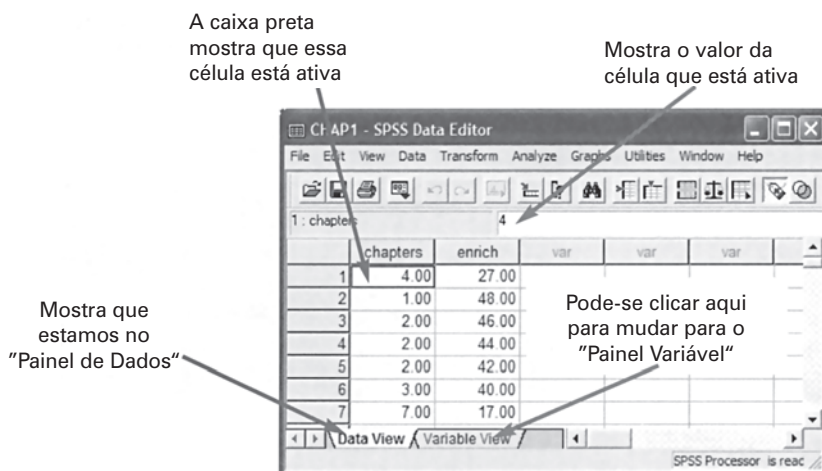


Figura 2.2 O editor de dados do SPSS (versão 10 em diante).

parte de baixo do editor de dados, você pode observar que existem duas orelhas (*tabs*) denominadas como *Data View* (Painel de Dados) e *Variable View* (Painel de Variáveis) (◀▶ Data View Variable View) e tudo o que temos que fazer para passar de um para outro é clicar com o mouse. Se estivermos no editor de dados então essa orelha terá um fundo branco e o painel das variáveis será cinza e se estivermos no painel das variáveis então o oposto irá ocorrer. Primeiro, vamos dar uma olhada em algumas características do editor de dados, essas são características que não se alteram quando nós mudamos de um painel para outro. Antes, vamos verificar os menus.

Em vários *softwares* dentro dos menus serão encontradas algumas letras sublinhadas, isso representa um “atalho de teclado” para facilitar o acesso a tal função. É possível selecionar muitas funções sem a utilização do mouse, e um usuário experiente pode achar esses atalhos mais rápidos que posicionar o mouse sobre o local adequado na tela. As letras sublinhadas nos menus indicam que a opção pode ser ativada pressionando simultaneamente *Alt* no teclado seguido da letra sublinhada. Assim para acessar a opção *Save As...* (Salvar Como), utilizando apenas o teclado, você deve pressionar *Alt* e *F* simultaneamente (que irá ativar

o menu *File* – Arquivo)*; então, mantendo a tecla *Alt* pressionada, pressione A que é letra sublinhada em *Save As...* No SPSS 13.0 (pelo menos quando estiver rodando no Windows XP), essas letras sublinhadas não são visíveis, no entanto, você ainda pode utilizar os atalhos de teclado. Se você pressionar *Alt* então as letras sublinhadas se tornarão visíveis, e uma vez pressionada a tecla *Alt* você pode apenas pressionar as letras sublinhadas para navegar pelos menus. Desta forma, ao longo do livro, eu incluirei as letras sublinhadas para os leitores que quiserem utilizar os atalhos de teclado, mas para aqueles, exceto você, que não querem, tentem não ser desestimulados pelas letras sublinhadas no livro!

Abaixo um guia breve de referência para cada um dos menus e algumas das opções que eles contêm. Isso é meramente um resumo e descobriremos as características de cada menu conforme formos progredindo ao longo do texto.

- **File (Arquivo):** Esse menu permite fazer coisas gerais tais como salvar arquivos, grá-

* N. de T. Evidentemente, se o software estiver traduzido, para se ativar o menu *Arquivo*, será necessário pressionar *Alt* + A.

ficos ou saídas de análise. Da mesma forma, pode-se abrir arquivos já salvos e imprimir gráficos, dados ou saídas. Em resumo, ele contém todas as opções costumeiramente encontrados no menu *File* (Arquivo).

- **Edit (Editar):** Esse menu contém funções de edição para o editor de dados. No SPSS para Windows é possível recortar e colar blocos de números de uma parte do editor para outro (que pode ser bastante útil quando você percebe que digitou montes de números no lugar errado). Você pode, ainda, utilizar *Options* (Opções) para selecionar várias preferências tais como as fontes que são utilizadas nas saídas. As preferências por omissão (*default*) são boas para a maioria das situações; a única coisa que você pode querer mudar é o tamanho da página de saída do visualizador colocando a opção como infinito (isso salvará muitas árvores quando você começar a imprimir coisas). Para fazer isso você precisa selecionar a orelha *Viewer* (Visualizador) no início da janela.
- **View (Exibir):** Esse menu lida com especificações do sistema tais como se o editor apresentará linhas de grade, ou se você apresentará os rótulos (*labels*) dos valores (exatamente o que são rótulos se tornará claro mais tarde).
- **Data (Dados):** Esse menu permite que se façam alterações no editor de dados. As características importantes são *insert variable* (inserir variável), que é utilizada para inserir uma nova variável no editor de dados (isto é, adicionar uma nova coluna); *insert case* (inserir caso), que é utilizado para inserir uma nova linha de dados entre duas já existentes, *split file* (dividir arquivo), que é utilizado para dividir um arquivo por uma variável de agrupamento (veja a Seção 3.4) e *select cases* (selecionar casos), que é utilizada para rodar uma análise utilizando somente uma amostra dos casos (valores).
- **Transform (Transformar):** Você deve utilizar esse menu se quiser manipular uma das variáveis de alguma forma. Por exem-

plo, você pode utilizar *recode* (recodificar) para mudar os valores de certas variáveis (por exemplo, se você quiser atribuir nomes diferentes por alguma razão) – veja o Quadro 3.2. A função *compute* (calcular) é também útil para transformar dados (por exemplo, para criar uma nova variável que é a média de duas existentes). Essa função permite que se execute qualquer número de modificações nas variáveis existentes (veja a Seção 3.3.4).

- **Analyze (Analisar):** A diversão começa aqui, porque todos os procedimentos estatísticos estão nesse menu.⁴ A seguir, há um breve guia das opções disponíveis que serão utilizadas durante o curso deste livro (isso é apenas uma pequena porção do que está disponível):

(a) **Descriptive Statistics (Estatística Descritiva):**⁵ Esse menu serve para

determinar medidas descritivas (média, mediana, moda, etc.), frequências e para uma exploração geral dos dados. Existe ainda um comando denominado de *crosstabs* (tabulação cruzada) que é útil para explorar frequências de dados e executar testes tais como o qui-quadrado, o exato de Fisher e o *kappa* de Cohen.

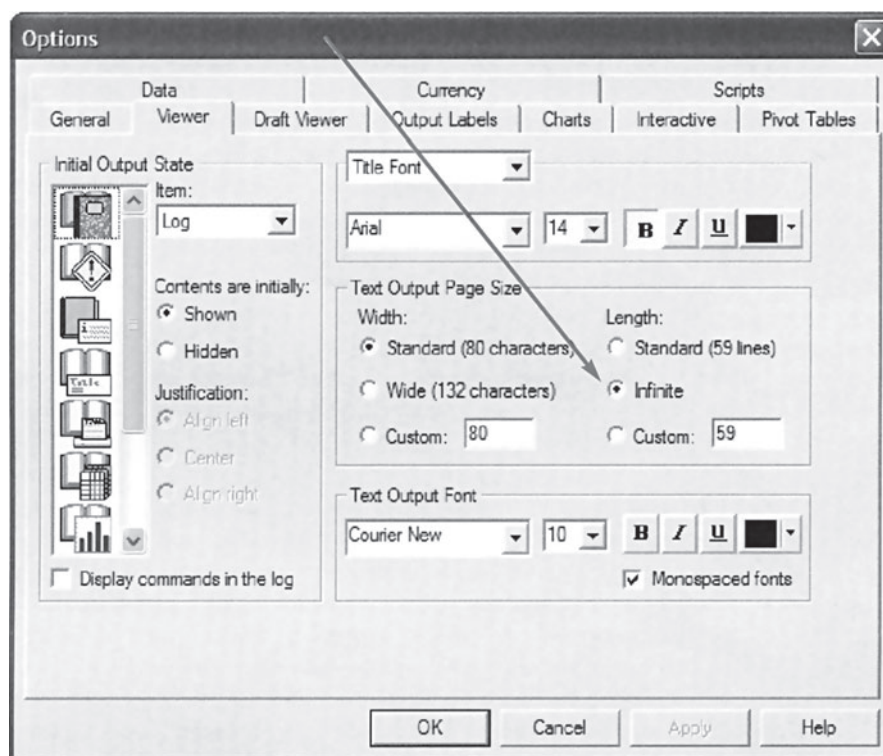
- (b) **Compare Means: (Comparar Médias):** Aqui é onde se podem encontrar os testes *t* (para amostras relacionadas e independentes – Capítulo 7) e a ANOVA independente (Capítulo 8).

- (c) **General Linear Model: (Modelo Linear Generalizado):**⁶ Esse menu é para realizar Análises de Variâncias complexas tais como a de dois fatores (relacionada, não-relacionada ou mista), a de um fator com medidas repetidas e a análise de variância multiva-

⁴ Esse menu é denominado de **Statistics** (Estatística) na versão 8.0 ou anterior.

⁵ Esse menu é denominado de **Summarize** (Resumir) na versão 8.0 ou anterior.

⁶ Esse menu é denominado de **ANOVA Models** (Modelo ANOVA) na versão 6 do SPSS.



riada (MANOVA) – veja os capítulos 9, 10, 11, 12 e 14.

- (d) **Correlate: (Correlação):** Não precisa ser um gênio para perceber que é aqui que as técnicas de correlação são encontradas! Você poderá executar correlações bivariadas tais como o cálculo do r de Pearson, o ρ de Spearman e o τ de Kendall bem como as correlações parciais (veja o capítulo 4).
- (e) **Regression: (Regressão):** Existe uma variedade de técnicas de regressão disponíveis no SPSS. Você poderá executar uma regressão linear simples, uma múltipla (Capítulo 5) e técnicas mais avançadas tais como a regressão logística (Capítulo 6).
- (f) **Loglinear: (Loglinear):** A análise loglinear está escondida nesse menu, esperando por você, e pronta para atacar como uma tarântula de sua toca (Capítulo 16).
- (g) **Data Reduction: (Redução de dados):** Você encontrará a Análise de Fatores aqui (Capítulo 15).
- (h) **Scale: (Escala):** Aqui será encontrada a Análise de Confiabilidade (Capítulo 15).
- (i) **Nonparametric Tests: (Testes não-paramétricos):** Existem uma variedade de estatísticas não-paramétricas disponíveis tais como o teste de aderência pelo qui-quadrado, o teste binomial, o teste de Mann-Whitney, o de Kruskal-Wallis, Wilcoxon e a ANOVA de Friedman (Capítulo 13).
- **Graphs (Gráficos):** O SPSS vem com seu próprio módulo gráfico bastante versátil. Entre os tipos de gráficos disponíveis podem ser incluídos: diagramas de barras, histogramas, diagramas de dispersão, dia-

gramas de caixa-e-bigodes, diagramas de pizza e diagramas de barras de erros para enumerar apenas alguns. Existe também a facilidade para editar qualquer tipo de gráfico para fazê-lo parecer bastante atrativo, que, na minha opinião, é bem formulado.

- **Window (Janela):** Permite que se mude de uma janela para outra. Assim, se você estiver olhando para uma saída e quiser retornar para a planilha, você poderá fazê-lo utilizando esse menu. Existem ícones como atalhos para a maioria das opções desse menu assim ele não é particularmente útil.
- **Help (Ajuda):** Esse é um menu inestimável pois ele oferece a você ajuda on-line tanto sobre o próprio sistema quanto sobre testes estatísticos. Embora os arquivos de auxílio estatístico possam ser, às vezes, praticamente inúteis (afinal de contas o programa não foi projetado para ensinar estatística), eles certamente não são um substituto para a aquisição de um bom conhecimento, mas poderá livrar você do aperto em algumas situações.

Assim como os menus, existe ainda um conjunto de ícones no topo da janela do editor de dados (veja a Figura 2.2) que são atalhos para recursos específicos que são utilizados com frequência. Todos esses recursos podem ser acessados pelo sistema de menus, mas utilizando os ícones você poupará tempo. Abaixo uma lista breve desses ícones e para que servem:



Esse ícone fornece a opção de abrir arquivos salvos previamente (se você estiver no editor de dados o SPSS assume que você quer abrir um arquivo de dados; se você estiver no visualizador de saídas, ele assumirá que você quer abrir um arquivo desse tipo).



Esse ícone permite que um arquivo seja salvo. Ele irá salvar o arquivo corrente, isto é, em que você estiver trabalhando quando clicar no ícone (seja ele de dados ou de saída). Se o arquivo ainda não tiver sido previamente salvo ele abrirá a caixa de diálogos *save data as* (salvar dados como).



Aqui será aberta a caixa de diálogos de impressão não importa o tipo de arquivo que você estiver trabalhando (dados ou saídas). As opções de impressão irão depender da impressora que você estiver utilizando. Uma dica útil é selecionar partes da saída clicando em porções da janela de saída (veja a Seção 2.5). Quando a caixa de diálogos de impressão aparecer lembre-se de clicar na opção de impressão apenas do texto selecionado. Selecionando partes do texto irá salvar muitas árvores pois a opção do SPSS é imprimir tudo o que estiver na janela de saída.



Clicando nesse ícone ativará a lista das últimas 12 caixas de diálogo que foram utilizadas. A partir dessa lista você poderá selecionar qualquer caixa da lista que ela aparecerá na tela. Esse ícone facilita a repetição de partes de uma análise.



Esse ícone permite que se vá direto a um caso (isto é, um participante). Ele será útil se estivermos trabalhando com grandes arquivos de dados. Por exemplo, se você quiser analisar um levantamento com 3000 respondentes será bastante tedioso ir rolando a tela até encontrar a resposta de um participante em particular. Esse ícone pode ser utilizado para pular diretamente para o caso (por exemplo, caso 2407). Clicando nesse ícone ativará uma caixa de diálogos que irá requerer que seja digitado o número do caso procurado.



Clicar nesse ícone fornecerá informações sobre uma determinada variável do editor de dados (uma caixa de diálogos permite a escolha de qual variável se poderá solicitar um resumo das informações).



Esse ícone não é para permitir que você faça observação de pássaros, mas ele permite que se procure por palavras ou números no arquivo de dados ou na janela de saída.



Ao clicar nesse ícone um novo caso será inserido no editor de dados (assim, ele

cria uma linha em branco no ponto em que ele estiver destacado – cursor estiver posicionado). Essa função é muito útil se for necessário adicionar novos dados ou se você esqueceu-se de colocar os dados de um participante em particular no editor de dados.



Clicar aqui irá criar uma nova variável à esquerda da variável que está ativa no momento (para ativar uma variável, simplesmente clique uma vez no nome dela no topo da coluna).



Esse ícone é um atalho para o comando **Data⇒Split File...** (Dados⇒Dividir Arquivo) (veja a Seção 3.4). Cientistas sociais às vezes conduzem experimentos com diferentes grupos de pessoas. No SPSS diferenciamos grupos de pessoas pela utilização de variáveis de código (veja a Seção 2.4.4) e essa função permite que se divida nossa saída de acordo com essa variável. Por exemplo, podemos querer testar a habilidade estatística de homens e mulheres. Podemos codificar cada participante com um número que representa o gênero (por exemplo, 1 = mulher, 0 = homem). Se quisermos então saber a habilidade estatística média de cada sexo nós simplesmente dividimos o arquivo utilizando a variável **gênero**. Qualquer análise subsequente será executada para homens e mulheres separadamente.



Esse ícone é um atalho para a função **Data⇒Weight Cases...** (Dados⇒Ponderar Casos). Essa função é necessária para quando trabalharmos com dados na forma de frequências (veja a Seção 16.4.2) e é útil para algumas situações avançadas de amostragem.



Esse ícone é um atalho para a função **Data⇒Select Cases...** (Dados⇒Selecionar Casos). Se você quer analisar apenas uma parte dos dados, essa é a opção a ser utilizada. Essa função permite que se especifique que casos devem ser incluídos na análise.



Ao clicar nesse ícone os rótulos dos dados para qualquer variável que tenha esse tipo de característica aparecerão ou serão ocultados. Algumas vezes agrupamos casos e para isso utilizamos uma variável de grupo ou codificadora para informar o computador a que grupo cada caso pertence. Por exemplo, se codificamos o gênero como 1 = mulher e 0 = homem então o computador saberá que cada vez que o valor 1 aparecer na coluna (variável) **gênero** a pessoa é uma mulher. Se você pressionar esse ícone, a legenda (mulher) irá aparecer no editor de dados ao invés do valor numérico (1); assim, você irá ver as palavras *homem* e *mulher* na coluna **gênero** ao invés de uma série de números (0 ou 1). A ideia se tornará mais clara na Seção 2.4.4.

2.4.1 Entrando com os dados no editor ①

Quando você executa o SPSS pela primeira vez ele vai apresentar uma planilha em branco no editor de dados e essa planilha será nomeada de “sem título” (isso é ridículo pois uma vez que ela foi nomeada de “sem título” ela já terá um título!). Quando entrar com um novo conjunto de dados, você deve entrar com os seus dados de uma forma lógica. O editor do SPSS é organizado de forma que *cada linha representa os dados de um objeto (indivíduo) enquanto que cada coluna representa uma variável*. Não existe uma separação para variáveis dependentes e independentes: os dois tipos devem ser colocados em colunas separadas. O ponto chave é que cada linha representa os dados de uma pessoa. Portanto, qualquer informação sobre cada caso deve ser fornecido com o editor de dados. Por exemplo, imagine que estamos interessados em investigar as diferenças entre os sexos quanto a percepção de dor criada por estímulos de quente e frio. Você pode colocar as mãos de algumas pessoas em um recipiente com água gelada por um minuto e perguntar a elas para atribuírem um valor do quanto doloroso isso foi em uma escala de 0 a 10. Você então pode pedir a elas

para segurar uma batata quente e novamente medir as percepções de dor. Imagine que eu sou um participante. Você terá uma única linha representando os meus dados, assim existirá uma coluna diferente para o meu nome, minha idade, meu gênero, minha percepção de dor com a água gelada e minha percepção de dor com a batata quente: Andy, 31 (quando esse livro apareceu), homem, 7, 10.

A coluna com a informação sobre o meu gênero é uma variável de grupo: posso pertencer ou ao grupo dos homens ao das mulheres, mas não a ambos. Como tal, essa variável é uma variável entre grupos (pessoas diferentes pertencem a grupos diferentes). Em vez de representar pessoas com palavras, no SPSS utilizamos números. Isso envolve atribuir a cada grupo um número e então informar ao SPSS que número representa cada grupo. No entanto, variáveis entre grupos são representadas por uma única coluna na qual o grupo que a pessoa pertence é definido por um número (veja a Seção 2.4.4). Por exemplo, podemos decidir que uma pessoa é homem e daremos a ela o número 0, ou se ela for mulher será atribuído o número 1. Nós teremos então que informar ao SPSS que cada vez que ele encontrar um número 1 em uma dada coluna a pessoa é mulher e se o valor for 0 essa pessoa é homem. Variáveis que especificam a que grupo entre vários uma pessoa pertence pode ser utilizada para dividir arquivos (assim, no exemplo da dor você pode executar uma análise com os homens e as mulheres separadamente – veja a Seção 3.4).

Finalmente, as duas medidas da dor são medidas repetidas (todos os participantes foram sujeitos aos estímulos de quente e frio). Desta forma, os níveis dessa variável podem ser fornecidos em colunas separadas (uma para a dor ao estímulo quente e outra para o frio).

O editor de dados é formado de células, que são apenas quadros em que os valores dos dados podem ser colocados. Quando uma de tais células está ativa ela se torna mais clara com as margens em preto destacadas (veja a Figura 2.2). Você pode se movimentar pelo editor de dados, de uma célula para outra, utilizando as teclas de setas $\leftarrow \uparrow \downarrow \rightarrow$ (encontradas à direita no teclado) ou clicando com o mouse na célula que deseja ativar. Para entrar com um valor no editor simplesmente mova-se para a célula que deseja colocar o valor, digite o valor, e então pressione a tecla com a seta para o lado que deseja continuar indo. Assim, para entrar com uma linha de dados, mova-se para a esquerda na mesma linha, digite o valor e então pressione \rightarrow (esse processo entra com o valor e então movimenta o cursor para a próxima célula à direita).

2.4.2 Criando uma variável ①

O primeiro passo para entrar com dados é criar alguma variável utilizando o painel “Variáveis” do editor de dados e então entrar com os dados no painel “Dados” do editor. Nós seguiremos por esses dois passos trabalhando com um exemplo. Imagine que estamos interessados em verificar a diferença entre professores e alunos. Tomamos ao acaso uma amos-

Dica da Samanta Ferrinho



Para resumir, qualquer variável medida sobre os mesmos participantes (uma medida repetida) deve ser representada por várias colunas (cada coluna representando um nível da variável de medidas repetidas). Entretanto, qualquer variável que define grupos diferentes de pessoas (como ocorre quando o delineamento entre grupos é utilizado e diferentes participantes são atribuídos a diferentes níveis da variável independente) é definida utilizando uma única coluna. Essa ideia se tornará mais clara a medida que você aprender como executar procedimentos específicos.

tra de cinco professores e alunos de psicologia da universidade de Sussex e verificamos quantos amigos cada um tem, o consumo semanal de álcool (em alguma unidade), a renda anual e quão neuróticos eles são (escores mais altos significa maior neurose). Esses dados estão na Tabela 2.1.

2.4.3 O visualizador de variáveis
(Variable View) ①

Antes de realmente entrarmos com os dados no editor nós precisamos criar as variáveis. Para criar as variáveis utilizamos o painel variáveis do editor de dados. Para acessar esse painel clique na orelha *Variable View* no fundo do editor de dados (◀▶\Data View Variable View/); a janela irá mudar e a orelha do visualizador ficará branco (Figura 2.3).

Cada linha do editor de variáveis representa uma variável, e para particularizar uma variável basta configurar a coluna adequada. Você pode alterar várias características tais como, a largura da coluna, o número de casas decimais após a vírgula que será mostrado e se os valores ficarão alinhados à direita, à esquerda ou no centro (mexa um pouco com esses valores e você pegará o jeito). As características mais importantes são as seguintes:

Name Você pode colocar nessa coluna o nome para cada variável. Esse nome irá aparecer no início da coluna quando

você estiver no painel de dados e serve para ajudar você a identificar as variáveis no painel dos dados. Existem algumas regras gerais sobre nomes de variáveis tais como o de eles só poderem ter oito caracteres ou menos, as letras só podem ser as minúsculas e você não poderá utilizar espaços em branco. Se for violada qualquer uma dessas regras o computador dirá que o nome da variável é inválido quando você clicar em uma célula diferente, ou tentar se movimentar para fora da célula utilizando uma tecla de seta.

Type Você pode ter diferentes tipos de dados. O tipo mais comum é *numeric* (numérico) (esse é o padrão, isto é, atribuído por omissão e apenas informa que essa é uma variável que contem números). Outro tipo é a variável *string* (texto), que é uma contração do “coleção de letras”. Assim se você quiser digitar o nome das pessoas, por exemplo, você precisará alterar o tipo de variável para *string* (texto) em vez de *numeric* (numérica).

Label O nome de uma variável (veja acima) é restrito a oito caracteres, assim o SPSS fornece uma maneira para que possamos identificar as variáveis de uma forma mais conveniente (esse rótulo poderá ter maiúsculas e mesmo espaços!). Isso pode não parecer importante mas é provavelmente um dos melhores hábitos que você poderá adquirir. Se você tiver uma variável chamada “número

Tabela 2.1 Alguns dados para brincar

	Amigos	Álcool	Renda	Neurose
Professor	5	10	20000	10
Professor	2	15	40000	17
Professor	0	20	35000	14
Professor	4	5	22000	13
Professor	1	30	50000	21
Aluno	10	25	5000	7
Aluno	12	20	100	13
Aluno	15	16	3000	9
Aluno	12	17	10000	14
Aluno	17	18	10	13

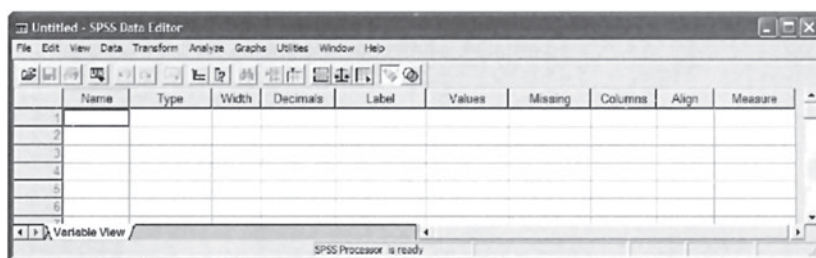


Figura 2.3 O painel “visualizador de variáveis” do editor de dados do SPSS.

de vezes que eu quis me dar um tiro durante uma aula de estatística do Andy Field”, você poderá denominá-la de “tiro”. Se você não adicionar um rótulo, o SPSS irá utilizar o nome da variável (nesse caso tiro) em todas as saídas de uma análise. Isso está bem e poderá ser bom, mas acredite, algum tempo depois quando você tentar rever os dados e as análises provavelmente irá pensar “que raios significa tiro?” ou “que diabos é essa variável sftg45c?”. Assim, mantenha o bom hábito de rotular as suas variáveis!

Values Essa coluna é para atribuir números para representar grupos de pessoas (veja a Seção 2.4.4).

Missing Essa coluna é para atribuir valores a dados *missing* (que faltam ou não foram obtidos) (veja a Seção 2.4.6).

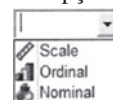
Measure Aqui é onde se define o nível na qual a variável foi mensurada (*Nominal*, *Ordinal* ou *Scale*) (Nominal, Ordinal ou Numérica).

Para entrar com os dados anteriores no editor do SPSS você precisará criar várias variáveis. Se começarmos com a variável **número de amigos**, devemos seguir esses passos:

1. Clicar com o mouse na primeira célula em branco da coluna denominada *Name* (Nome).
2. Digite a palavra *friends* (amigos) lembre-se que só podemos utilizar nomes com oito ou menos caracteres assim não será possível digitar *number of friends* (número de amigos).

3. Saia da célula utilizando as teclas com as setas do teclado (ou você pode apenas clicar em uma célula diferente, mas com as setas é mais rápido).

Você acaba de criar sua primeira variável! Você deve ter notado que uma vez que digitou um nome o SPSS cria alguns valores padronizados (por omissão) para a variável (tal como assumir que ela é numérica e colocar duas casas após a vírgula). Agora em virtude de que você quer manter bons hábitos, mova para a célula na coluna chamada *Label* (Rótulo) e digite ‘número de amigos’. Finalmente, podemos especificar o nível em que a variável foi mensurada (veja o Quadro 2.1) indo para a coluna rotulada *Measure* (Medida) e selecionando uma das opções *Nominal*, *Ordinal* ou *Scale* (Nominal, Ordinal ou Numérica) da lista suspensa (veja o Quadro 2.1)



Uma vez que uma variável foi criada, você pode retornar para o painel dos dados clicando na orelha *Data View* (visualizador de dados) na parte de baixo do editor de dados (◀▶ Data View Variable View /). O conteúdo da janela irá mudar e você poderá notar que a primeira coluna apresenta agora o rótulo *friends* (amigos). Para entrar com os dados, clique na célula no topo da coluna denominada *friends* (amigos) e digite o primeiro valor, 5. Para registrar esse valor na célula, você precisa se movimentar para uma célula diferente e em virtude de estarmos inserindo dados em colunas, a maneira mais fácil de fazer isso é pressionando a tecla ↓. Essa ação moverá você uma posição abaixo e o número 5.00 deverá

Quadro 2.1

Níveis de Medida ①

Existem vários níveis nas quais as variáveis podem ser medidas: nominal, ordinal, intervalar e de razão. O mais baixo é o nominal, onde os números apenas representam nomes. Por exemplo, se perguntarmos às pessoas se ler este capítulo as deixa com tédio, as respostas serão *sim* ou *não*. Assim, as pessoas serão classificadas em duas categorias: entediadas ou não-entediadas, obviamente aqui o valor do número é irrelevante e ele meramente representará uma das duas categorias (se você estiver utilizando uma variável codificadora [próxima seção] então os dados são *nominais*). Uma variável ordinal nos fornece mais informações do que uma nominal. Se utilizarmos uma escala ordinal para medir algo, podemos informar não apenas que coisas ocorreram, mas também a ordem em que elas aconteceram. No entanto, esses dados não nos informam o valor das diferenças entre as categorias. Por exemplo, se classificarmos três pessoas de acordo com o quanto elas estão entediadas – muito entediada, tédio médio e pouco entediada. Esses rótulos nos informam sobre o nível de tédio. Ao utilizar categoriais ordinais nós agora sabemos que a pessoa mais entediada apresenta mais tédio do que a pessoa menos entediada! Dados em intervalo são escores medidos por uma escala em que todos os intervalos são iguais. Por exemplo, em vez de perguntar às pessoas se elas estão entediadas, poderíamos medir o tédio com uma escala de 10 pontos (com 0 representando alguém muito interessado e 10 alguém totalmente entediado). Para dados em intervalo um aumento no tédio representado por uma mudança do grau 3 na escala para 4 tem o mesmo valor que um aumento do grau 9 para o 10, mas não é verdade que um entediado de grau 9 apresenta um tédio 3 vezes maior do que um classificado com grau 3. Dados de razão apresentam essa propriedade mas aqui é possível afirmar que alguém com um grau 8 de tédio está duas vezes mais entediado do que alguém que foi classificado com grau 4 de tédio. Esses dois tipos de dados são representados pela opção *Scale* (numérica) no SPSS. Deve parecer óbvio que em algumas ciências sociais, principalmente psicologia, é extremamente difícil estabelecer se um dado é, de fato, intervalar (nós podemos realmente dizer que uma alteração na escala de tédio representa uma mudança genuína na experiência de apresentar tédio?) – veja Field e Hole (2003).

aparecer na célula acima. Entre com o próximo número, 2, e então pressione ↓ novamente para seguir adiante.

2.4.4 Criando variáveis codificadas ①

Nas seções anteriores eu mencionei as variáveis de grupo (codificadoras) e esta seção é dedicada uma descrição completa desse tipo de variável (é um tipo de variável que será muito usada). Uma variável codificadora (também conhecida como variável de grupo ou agrupadora) é uma variável que consiste em uma série de números representada em níveis de uma variável tratamento ou que descreve diferentes grupos de pessoas. Em experimentos, variáveis de grupo são utilizadas para representar variáveis independentes que tenham

sido mensuradas entre grupos (isto é, diferentes participantes foram atribuídos a grupos diferentes). Assim, se você está realizando um experimento com um grupo de participantes em uma condição experimental e um grupo diferente de participantes em um grupo em um grupo-controle, você poderá atribuir ao grupo-experimental o código 1 e ao grupo controle o código 0, por exemplo. Quando você entrar com os dados no editor, então deverá criar uma variável (que poderá chamá-la de **grupo**) e digitar 1 para todos os participantes do grupo-experimental e 0 para qualquer participante do grupo-controle. Esses códigos informam ao computador que todos os casos que foram atribuídos ao grupo 1 devem ser tratados como se pertencentes a um mesmo grupo e da mesma forma para os casos em que foram


Existe uma regra simples de como as variáveis devem ser colocadas no editor de dados do SPSS: níveis de variáveis entre grupos devem variar verticalmente (ocupar uma única coluna), enquanto que variáveis dentro-assuntos (medidas repetidas) devem variar horizontalmente (mesma linha). Veremos como essa regra funciona no Capítulo 6.

atribuídos o valor 0. Em situações que não sejam experimentos, você poderá simplesmente utilizar os códigos para distinguir grupos de pessoas ocorrendo naturalmente (por exemplo, você pode atribuir aos alunos o código 1 e aos professores o código 0).

Nós temos uma variável de código em nossos dados: é aquela descrevendo se uma pessoa é um professor ou um aluno. Para criar essa variável codificadora, seguimos os passos que seriam dados para criar uma variável normal, mas devemos também informar ao computador que códigos numéricos foram atribuídos e para quais grupos. Assim, antes de mais nada, retorne ao painel de visualização das variáveis se você ainda não estiver nele e mova-se para a célula na segunda linha do editor na coluna rotulada como *Name* (Nome). Digite um nome (chame-a de **grupo**). Eu ainda estou tentando fazer com que você adquira bons hábitos, assim mova-se ao longo da segunda linha para a coluna denominada de *Label* (Rótulo) e forneça a variável uma descrição pormenorizada tal como ‘A pessoa é professor ou aluno?’ Então para definir os códigos dos grupos, mova-se ao longo da segunda linha para a coluna denominada de *Values*. Você verá que na célula está escrito *None* (Nada) e que existe um botão (veja a Figura 2.4). Clique nesse botão para acessar a caixa de diálogos *Value labels* (Valores dos rótulos) (veja a Figura 2.4).

A caixa de diálogos *Value labels* (valores dos rótulos) é utilizada para especificar o código dos grupos. Isso pode ser feito em passos simples. Primeiro clique com o mouse no espaço em branco próximo de onde diz *Value* (Valor) (ou pressione *Alt* e *u* ao mesmo tempo) e digite o código (por exemplo, 1). Esses códigos podem ser completamente

arbitrários; por simples convenção as pessoas normalmente utilizam os valores 0, 1, 2, 3, etc., mas na prática você pode colocar um valor de 495 se você estiver se sentindo particularmente arbitrário. O segundo passo é clicar com o mouse no espaço em branco abaixo, próximo de onde diz *Value label* (Valor do rótulo) (ou pressionar *Alt* e ao mesmo tempo *e*) e digite um rótulo apropriado para o grupo. Na figura 2.4, eu digitei 2 como meu código e atribuí um rótulo de *Student* (Estudante). O terceiro passo é adicionar esse código na lista clicando em *Add*. Na Figura 2.4, eu já defini o código 1 para o grupo dos professores; para acrescentar o código do grupo dos estudantes clique em *Add*. Quando você definir todos os valores dos códigos, clique em *OK*; se você clicou no *OK* e esqueceu de adicionar o último código da lista, o SPSS irá exibir uma mensagem avisando que quaisquer alterações pendentes serão perdidas. Em linguagem simples ele está dizendo para voltar e clicar no *Add* antes de continuar. Finalmente variáveis de código sempre representam categorias e assim os níveis em que elas são mensuradas é nominal (ou ordinal se as categoriais tiverem uma ordem), assim é necessário especificar o nível em que a variável foi mensurada indo até a coluna denominada *Measure* (Medida) e selecionando *Nominal* (ou *Ordinal* se os grupos tiverem uma ordem) de uma lista suspensa.

Tendo definido os códigos, você pode então mudar para o painel editor de dados e digitar os valores numéricos na coluna apropriada (se uma pessoa for um professor, digite 1, se for um estudante digite 2). O que é realmente excitante é que você pode solicitar que o computador apresente os códigos ou os valores que você forneceu clicando em 

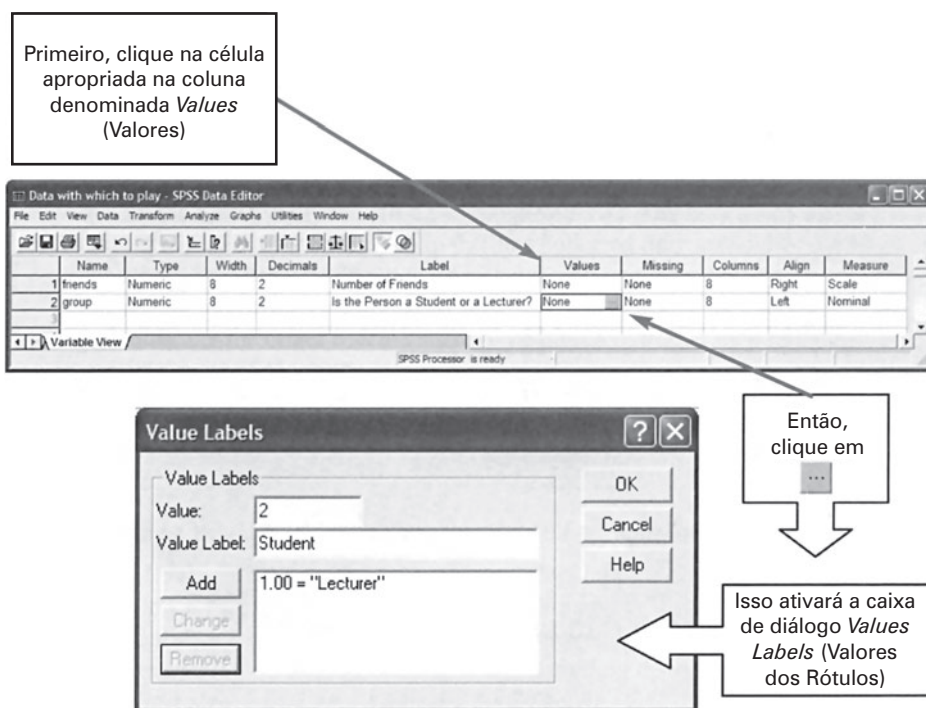


Figura 2.4 Definindo variáveis de código e seus valores no SPSS.

(veja a Figura 2.5). A Figura 2.5 mostra como os dados devem ser arranjados para uma variável de grupo/código. Agora, lembre que cada linha do editor de dados representa um participante e assim, neste exemplo, está claro que os cinco primeiros são professores enquanto que os de 6 a 10 são estudantes. Esse exemplo, também, demonstra por que variáveis de grupo são utilizadas para variáveis que tenham sido mensuradas entre assuntos: porque pela utilização do código é impossível para um participante pertencer a mais do que um grupo. Essa situação deve ocorrer no delineamento entre grupos (isto é, cada participante deve ser testado em apenas um dos grupos: o experimental ou o controle). Entretanto, no delineamento de medidas repetidas (entre conteúdos) cada participante é testado em cada uma das condições e assim nós não iremos utilizar esse tipo de variável de grupo (pois cada participante deve tomar parte em cada uma das condições experimentais).

2.4.5 Tipos de variáveis ①

Existem diferentes tipos de variáveis que podem ser usadas no SPSS. Na maioria dos casos você irá utilizar variáveis **numéricas**. Essas variáveis são aquelas que contêm números e incluem as do tipo código que acabaram de ser descritas. No entanto, uma das outras opções ao se criar uma variável é especificar o seu tipo, o que é feito clicando na coluna denominada *Type* no painel variável e então clicando em ... (exatamente como foi feito na especificação dos rótulos na Figura 2.4). Isso irá ativar a caixa de diálogos *Variable type* (tipo de variável) como na Figura 2.6, que mostra os *Default* (valores por omissão). Por padrão, uma variável é estabelecida como sendo numérica e armazena oito dígitos, mas você pode trocar esse valor digitando um novo em um espaço denominado de *Width* (tamanho) na caixa de diálogos. Sob circunstâncias normais você não necessitará mais do que o SPSS retenha mais

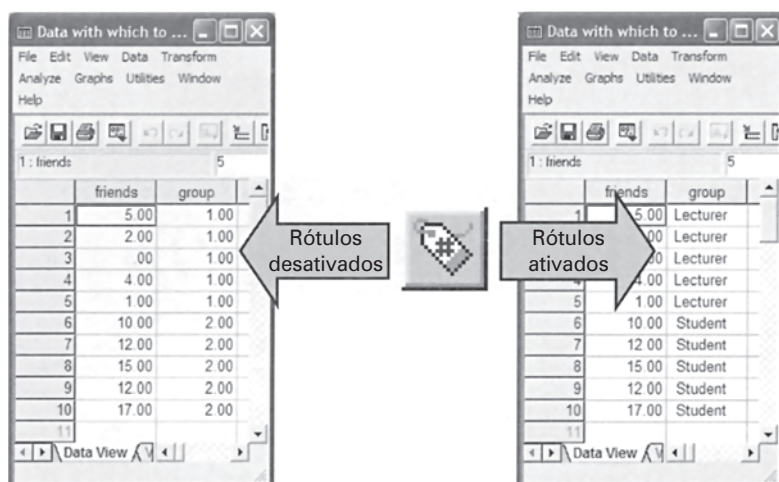


Figura 2.5 Codificando valores no editor de dados com os rótulos dos valores ativados e desativados.

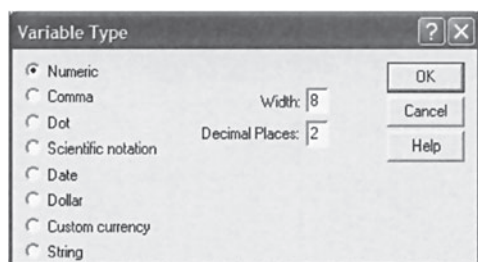


Figura 2.6 Definindo o tipo de variável sendo utilizado.

do que oito dígitos a menos que você esteja fazendo cálculos que sejam particularmente precisos. Outro valor determinado por omissão ou padrão são as duas casas após a vírgula (de fato você irá notar que por padrão sempre que você digitar qualquer número inteiro na planilha o SPSS acrescentará uma vírgula seguida de dois zeros, isso poderá ser desconcertante de início!) É bastante fácil mudar o número de casas decimais para uma dada variável substituindo o 2 por um novo valor que dependerá do grau de precisão que for necessário.

O *variable type* (tipo da variável) da caixa de diálogos também permite que se especifique outros tipos de variáveis, na maior parte

serão utilizados valores numéricos. Outro tipo de variável é a texto (*string*). Uma variável texto é simplesmente uma linha de caracteres e pode representar comentários sobre um determinada participante ou outra informação que não se deseja analisar como uma variável de agrupamento (como o nome do participante). Se for selecionada a opção variável texto (*string*), o SPSS permitirá que se especifique o tamanho da variável texto (*string*), que por omissão é de oito caracteres, de forma que será possível inserir frases longas se necessário.

2.4.6 Valores que faltam (*missing*) ①

Apesar de lutarmos para coletar conjuntos de dados completos ocorre que, muitas vezes, nós temos dados que não podem, por algum motivo, ser obtidos, dados que faltam ou são desconhecidos. A omissão de dados pode ocorrer por vários motivos: em questionários longos os participantes podem acidentalmente esquecer-se de responder a algumas questões; em procedimentos experimentais falhas mecânicas podem fazer que o dado não seja registrado e em levantamento de assuntos delicados (por exemplo, comportamento sexual) os participantes podem exercer o seu direito de

não responder as questões. No entanto, só porque não temos informações sobre alguns detalhes dos participantes isso não significa que devemos ignorar os demais dados que temos (embora isso possa em algumas vezes criar dificuldades estatísticas). Nesses casos precisamos informar ao computador que o dado de algum participante em particular está faltando. O princípio por trás dos valores desconhecidos é totalmente similar aquele sobre variáveis de grupo ou codificadoras, isto é, aqui devemos escolher ou arbitrar um valor para representar o ponto de dados que está faltando. Esse valor simplesmente diz ao computador que não existe um valor registrado para o participante para uma determinada variável. O computador então irá ignorar a célula do editor de dados (ele não utiliza o valor em uma análise). Você deve ter cuidado na escolha desse valor e certificar-se que ele não irá coincidir com qualquer valor existente no conjunto de dados original. Por exemplo, se informarmos ao computador que o valor 9 indica um valor desconhecido ou omitido e vários participantes apresentaram genuinamente esse escore, então o computador irá tratar esses dados como não-existent, quando na realidade, eles existem.

Para especificar um código para os valores desconhecidos basta clicar na coluna denominada de **Missing** no painel das variáveis do editor e então clicar em **...** para ativar a caixa de diálogos *Missing Values* (valores que faltam) na Figura 2.7. Por omissão, o SPSS assume que não existem valores desconhecidos mas se você, de fato, tiver dados com

esse tipo de problema você pode escolher defini-los de três formas: a primeira é selecionar valores isolados (clcando no círculo próximo onde diz *Discrete missing values* – valores desconhecidos isolados) que são valores únicos que representam dados que faltam. A razão pela qual você pode querer escolher ter vários números para representar valores que faltam é que você poderá atribuir diferentes significados para cada um dos valores. Por exemplo, você poderá ter o número 8 representando a resposta ‘não se aplica’, um código 9 para representar uma resposta ‘eu não sei’ e um código 99 dizendo que o participante não respondeu a questão. Para o computador não fará diferença e ele irá ignorar todas as células que contém esses valores; entretanto, a utilização de códigos diferentes pode ser uma forma útil de lembrar porque um determinado valor está faltando. Normalmente, um valor discreto é suficiente e em um experimento em que atitudes são mensuradas em uma escala de 100 pontos (escores variando de 1 a 100) você pode escolher o valor 666 para representar valores que faltam porque (1) esse valor não irá ocorrer nos dados que foram coletados (2) porque valores que faltam criam problemas estatísticos e você poderá lembrar que os que não responderam a questões do questionário são filhotes do capeta! Uma segunda opção é selecionar um intervalo de valores para representar dados que faltam e isto é útil em situações na qual pode ser necessário excluir dados que tem valores entre dois pontos. Assim, podemos querer excluir todos os escores entre 5 e 10. A opção final é ter um intervalo de valores e não valores isolados.

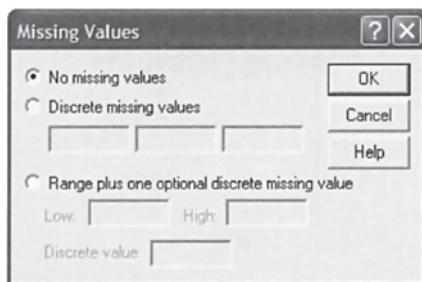




Figura 2.7 Definindo valores que faltam (*missing*).

2.4.7 Mudando o formato da coluna ①

A última opção disponível quando definimos uma variável é ajustar o formato da coluna dentro do editor de dados. A opção por omissão é ter uma coluna que tem oito caracteres de largura com todos os valores alinhados à direita. Esses dois valores podem ser alterados. No painel variáveis você pode aumentar ou diminuir o valor padrão de 8 clicando nas setas abaixo e acima (). Para mudar o alinhamento da coluna

mova-se para a coluna denominada **Align** (alinhar) e clique na célula da variável que você quer alterar. Você pode então ativar da lista apresentada em um menu suspenso clicando  e selecionando umas das opções:  *Left*, *Right* ou *Center* (Esquerda, Direita ou Centro). É particularmente útil ajustar o tamanho da coluna quando você tiver uma variável de grupo (código) com rótulos que excedem os oito caracteres.

2.4.8 Teste rápido ①

Tendo criado as variáveis **amigos** e **tipo** com um pouco de ajuda tente entrar com as demais variáveis da Tabela 2.1. Os dados finais e as variáveis devem parecer com as da Figura 2.8 (mais ou menos!).

2.5 O VISUALIZADOR DE SAÍDAS (OUTPUT VIEWER) ①

Junto à janela de edição de dados existe uma segunda janela conhecida como *viewer* (visualizadora) ou visualizadora de saídas ou simplesmente de saídas. Antigamente, a janela de saída (como era denominada) mostrava somente resultados estatísticos em uma fonte sem graça e os gráficos eram apresentados em uma janela separada.⁷ A nova e melhorada janela de saída dos resultados apresenta gráficos, tabelas e resultados estatísticos no mesmo lugar (com fontes bem melhores e tabelas bem

⁷ Ela se tornou o *output navigator* (navegador de saída) nas versões 7.0 e 7.5, mas mudou o nome para *viewer* (visualizadora) na versão 8.

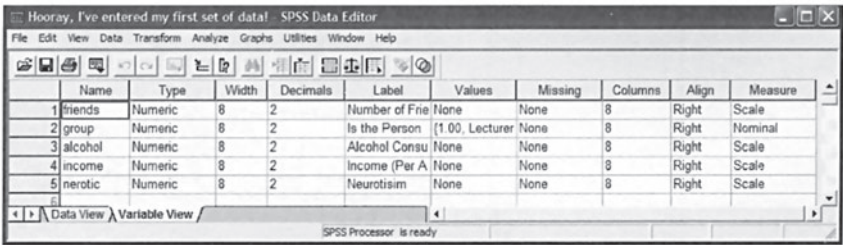
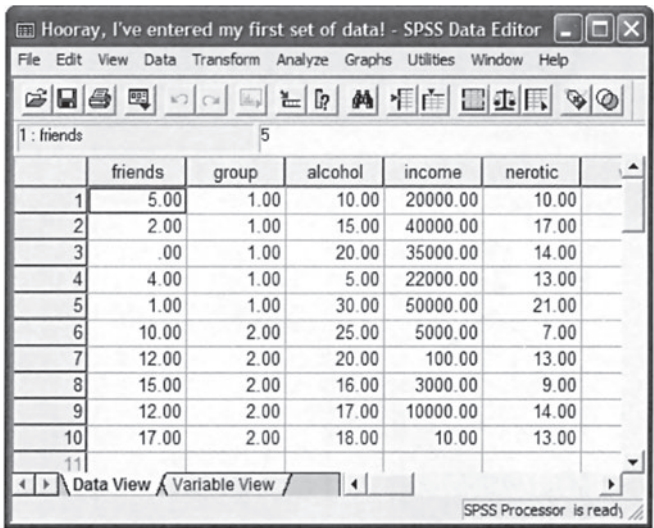


Figura 2.8 Visão final dos dados e variáveis para os valores da Tabela 2.1

apresentadas). Minha previsão na última edição desse livro foi a de que futuras versões do SPSS iriam incluir facilidades para fazer chá na janela de saída, mas infelizmente isso não se realizou (SPSS por favor tome nota!).

A Figura 2.9 mostra a estrutura básica da janela de saídas (ou de apresentação de resultados). No lado direito existe um grande espaço na qual a saída é apresentada. O SPSS mostra tanto gráficos quanto resultados de análises estatísticas nessa parte da janela. É também possível editar os gráficos e para fazer isso você simplesmente deve dar um duplo clique naquele que deseja ver editado. No lado esquerdo da janela de saída existe um diagrama de árvore ilustrando a estrutura da saída.

Esse diagrama é útil para quando for realizada uma série de análises porque ela fornece uma forma fácil de acessar partes específicas da saída. A estrutura de árvore é praticamente autoexplicativa no sentido de que cada vez que for realizado um procedimento (tal como construir um gráfico ou rodar uma análise estatística), o SPSS lista esse procedimento como um cabeçalho principal.

Na Figura 2.9, construí um gráfico e então conduzi uma análise univariada da variância (ANOVA) e assim esses dois nomes aparecem como cabeçalhos principais. Para cada procedimento existe uma série de subprocedimentos e eles são listados como ramos de uma árvore debaixo do cabeçalho principal. Por exemplo,

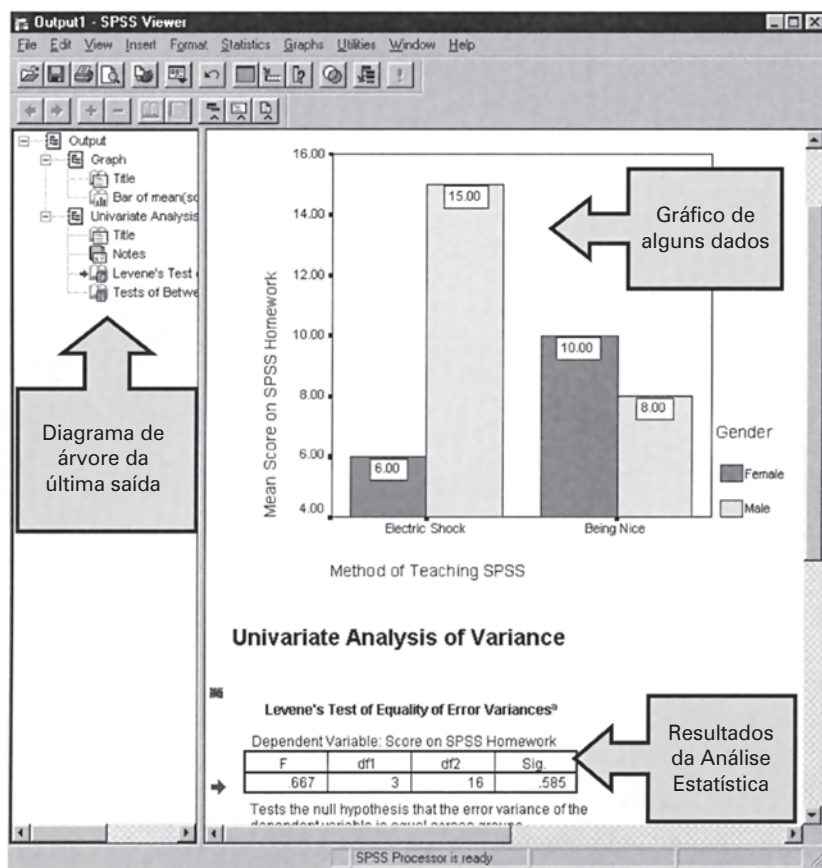


Figura 2.9 O visualizador de saídas.

no procedimento ANOVA existem várias seções de saída tais como o teste de Levene (que testa a hipótese de variâncias homogêneas) e os efeitos entre grupos (isto é, o teste F que verifica se as médias diferem significativamente). Você pode pular para qualquer um destes subcomponentes da análise clicando no ramo apropriado do diagrama de árvore. Assim, se você quiser ir direto para os resultados da análise entre grupos você deve mover o cursor para a porção esquerda da janela e clicar onde diz *Tests of Between-Subjects Effects* (Testes dos resultados entre sujeitos). Essa ação irá ressaltar essa parte da saída na parte maior da janela de visualização. Você também utilizar esse diagrama de árvore para selecionar partes da saída para imprimir. Por exemplo, se quiser imprimir um gráfico apenas e não toda a saída, pode-se clicar na palavra *Graph* (Gráfico) na estrutura de árvore e o gráfico se tornará destacado no painel de saída. Será então possível por meio do menu de impressão determinar que seja impresso somente a parte destacada da janela de saída. Nesse contexto vale a pena notar que se for clicado no cabeçalho principal (tal como *Univariate Analysis of Variance* – Análise Univariada de Variância) o SPSS vai destacar não somente o cabeçalho principal mas também todos os subcomponentes. Isso é extremamente útil quando se quer imprimir os resultados de um único procedimento estatístico.

Existem vários ícones na janela visualizadora de saídas que ajudam a realizar coisas rapidamente sem a necessidade de se recorrer aos menus. Alguns desses ícones são os mesmos que aqueles descritos na janela de edição de dados assim iremos nos concentrar somente naqueles que dizem respeito unicamente a janela de saída.



Como ocorreu com a janela de edição de dados, este ícone ativa o menu de impressão. No entanto, quando esse ícone é pressionado na janela de saída ele ativa um menu para imprimir a saída. Quando o menu de impressão está ativo ele apresentará uma opção por *Default* (omissão) para imprimir toda a saída ou

você poderá escolher uma opção para imprimir a saída no momento visível na tela ou, ainda mais útil, uma opção para imprimir uma seleção (parte) da saída. Para escolher essa última opção você já deve ter previamente selecionado a parte da saída que quiser imprimir (veja o mencionado anteriormente).



Esse ícone leva você rapidamente de volta ao editor de dados.



Esse ícone leva você a última saída da janela visualizadora (assim ele leva você ao último procedimento realizado).



Esse ícone pega a parte ativa atual da estrutura de árvore e avança para um ramo acima. Por exemplo na Figura 2.9 o *Tests of Between-Subjects Effects* (Testes dos resultados entre sujeitos) é um subcomponente sob o título *Univariate Analysis of Variance* (Análise univariada de variância). Se quisermos promover essa parte da saída a um nível mais alto (isto é, fazê-la um título principal) então isso poderá ser feito com a utilização desse ícone.



Esse ícone é o oposto do acima no sentido de que ele rebaixa parte da estrutura de árvore. Por exemplo, na Figura 2.9, se não quisermos que a *Univariate Analysis of Variance* (Análise univariada de variância) seja a única seção, nós podemos rebaixá-la de modo que ela se torne parte do item anterior (o item Gráfico). Esse botão é útil para combinar partes da saída que se relacionam a questões de pesquisa específicas.



Esse ícone recolhe parte da estrutura de árvore o que significa, simplesmente, que ele oculta os subcomponentes sob um título particular. Por exemplo na Figura 2.9, se selecionarmos *Univariate Analysis of Variance* (Análise univariada de variância) e pressionarmos esse ícone, todos os subtítulos irão desaparecer. As seções que desaparecem da estrutura não são eliminadas da própria saída mas existe apenas

uma condensação da estrutura de árvore. Isso poderá ser útil quando existir muitas saídas e o diagrama de árvore tornar-se muito complicado.



Esse ícone expande ou oculta seções. Por omissão todos os títulos principais são mostrados no diagrama de árvore em sua forma expandida. Se, no entanto, você optou por ocultar parte do diagrama (utilizando o ícone acima) então você pode utilizar esse ícone para desfazer a sua besteira.



Esse ícone e o seguinte permitem que se mostre ou oculte partes da própria saída. Assim, pode-se selecionar parte da saída no diagrama de árvore e então utilizar esse ícone para que a própria saída seja ocultada. Ela não é apagada, apenas ficará fora de vista. Assim, esse ícone é semelhante ao que oculta a estrutura de árvores só que ele afeta a própria saída em vez estrutura de árvore. Ele é útil para eliminar detalhes irrelevantes.



Esse ícone desfaz o que foi feito pelo anterior, assim se você ocultou parte da saída e clicar esse ícone ele fará essa parte oculta reaparecer. Por *Default* (omissão) todas as partes da saída são exibidas e assim esse ícone não está ativo: ele se tornará ativo somente quando você tiver ocultado parte da saída.



Embora esse ícone pareça com um rolo de pintura ele infelizmente não pinta a casa para você. O que faz é inserir um novo título ao diagrama de árvore. Por exemplo, se você realizou vários testes estatísticos relacionados a uma ou mais questões de pesquisa você poderá inserir um título principal e então rebaixar os títulos das análises relevantes de modo que todos fiquem sob esse novo título.



Considerando que você tenha feito o colocado acima, você poderá utilizar esse ícone para fornecer ao novo título uma legenda. O que você digitar irá aparecer na saída. Assim, você poderá querer ter um cabeçalho como 'Questão de Pesqui-

sa número 1" que informará que as análises sob esse cabeçalho se relacionam com a primeira questão de pesquisa.



Esse último ícone é utilizado para colocar uma caixa de texto na janela de saída. Você poderá digitar qualquer coisa nessa caixa. No contexto dos dois ícones anteriores você poderá utilizar um quadro de texto para explicar sua primeira questão de pesquisa (por exemplo, "Minha primeira questão de pesquisa é saber se o tédio se instalou ou não ao final do primeiro capítulo do meu livro. A análise seguinte testa a hipótese de que os níveis de tédio serão significativamente mais altos ao final do capítulo do que no início.").

2.6 A JANELA DE SINTAXE ③






Eu mencionei anteriormente que algumas vezes seria útil utilizar a sintaxe do SPSS. Ela é uma linguagem de comandos para executar análises estatísticas e manipulação de dados. Na maioria das vezes nós faremos as coisas que precisamos utilizando as caixas de diálogos; entretanto, a sintaxe do SPSS poderá ser útil. Em algumas situações existirão coisas que só poderão ser executadas com o recurso da sintaxe (admito que muitas dessas coisas são bastante avançadas, mas existirão umas poucas situações nesse livro em que mostrarei a você alguns truques legais utilizando a sintaxe). Uma segunda razão para utilizarmos a sintaxe é se análises semelhantes forem executadas muitas vezes. Nessas situações é frequentemente mais rápido fazer a análise e salvar a sintaxe à medida que se prossegue. Felizmente isso é fácil de fazer porque muitas das caixas de diálogos do SPSS apresentam um botão . Quando você tiver especificado a análise utilizando uma caixa de diálogos, se clicar nesse botão colorá a sintaxe utilizada em uma janela de sintaxe. Para abrir a janela de sintaxe utilize os menus **File**⇒**New**⇒**Syntax...** e uma janela de sintaxe em branco irá aparecer como a da



Figura 2.10. Nessa janela você poderá digitar os comandos de sintaxe que desejar. As regras da sintaxe do SPSS podem ser bastante frustrantes (por exemplo, cada linha deve terminar com um ponto e se ele for esquecido uma mensagem de erro irá aparecer) e estão além dos objetivos deste livro. Entretanto, à medida que formos avançando, mostrarei algumas coisas que darão a você um pouco do sabor de como a sintaxe poderá ser utilizada. Muitos de vocês não precisarão dela, mas para aqueles que desejarem ou precisarem espero que essa pequena amostra seja suficiente para que tenham o interesse despertado e continuem a aprender mais. Uma vez que os comandos tenham sido digitados eles precisam ser executados pela utilização do menu **Run**. O comando **Run⇒All** executará toda a sintaxe digitada na janela (clicando em  isso também ocorrerá), ou você poderá marcar uma seleção da sua sintaxe com o mouse e utilizar o **Run⇒Selection** para processar a sintaxe selecionada. Você poderá rodar o comando corrente utilizando **Run⇒Current** (ou pressionar *Ctrl* e *R* no teclado) ou finalmente executar toda a sintaxe a partir do cursor até o final da janela de sintaxe utilizando **Run⇒To End**.



2.7 SALVANDO ARQUIVOS ①

Embora muitos de vocês estejam familiarizados com a tarefa de salvar arquivos no Windows, ela é um conhecimento vital e assim descreverei brevemente o que é preciso fazer.

Para salvar um arquivo simplesmente utilize o ícone  (ou use os menus **File⇒Save** ou **File⇒Save As...**). Se o arquivo for novo, então clicar nesse ícone ativará a caixa de diálogos *Save Data As* (Salvar Dados como) (veja a Figura 2.11). Se você estiver no editor de dados quando selecionar *Save As...* (Salvar Como) o SPSS irá salvar o arquivo em que você estiver trabalhando no momento, mas se estiver no painel de visualização de saídas ele irá salvar a saída existente no momento.

Existem várias características na caixa de diálogos da Figura 2.11. Primeiro você precisa determinar um local onde armazenar (salvar) o arquivo. Normalmente, existem dois tipos de locais onde se pode guardar dados: o disco rígido e as mídias móveis (disquete, CDs, DVDs, *Pen-Drives*, etc.) ou outras opções que nem imagino que você possa ter no seu computador. A primeira coisa a fazer é selecionar ou uma mídia móvel ou o disco rígido e dar um duplo clique em . Uma vez que tenha escolhido o local principal a caixa de diálogos mostrará todas as pastas disponíveis naquela mídia em particular (você pode ainda não ter pastas no disco ou mídia e nesse caso poderá criar uma clicando em ). Logo que tiver selecionado uma pasta na qual irá salvar o arquivo você precisará dar-lhe um nome. Se você clicar no espaço próximo onde diz *File name* (Nome do Arquivo) um cursor irá aparecer e você poderá digitar um nome de até dez letras. Por omissão o arquivo será salvo no formato do SPSS, assim ele apresentará a extensão *.sav*, se ele for um documento do visualizador de saídas ou janela de saída o arquivo terá a ex-

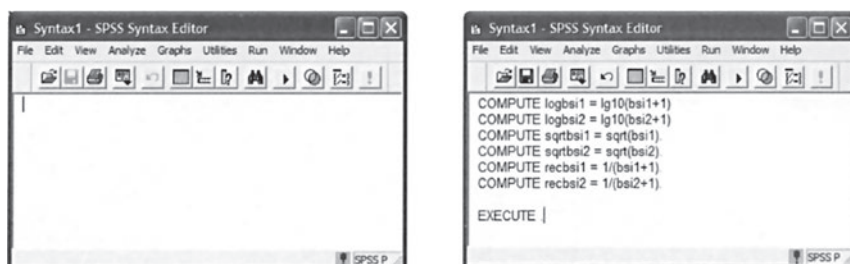


Figura 2.10 Uma nova janela de sintaxe (esquerda) e uma com alguma sintaxe (direita).

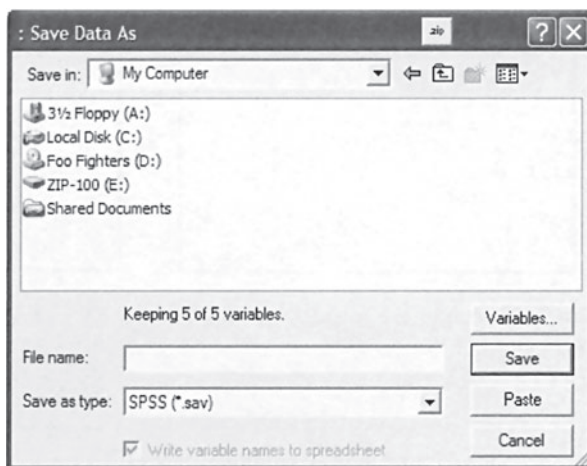







Figura 2.11 A caixa de diálogos *Save Data As* (Salvar Dados Como).

tensão *.spo* e se ele for um arquivo de sintaxe a extensão será *.sps*. Entretanto, você poderá salvar dados em outros formatos tais como os do Microsoft Excel e arquivos textos delimitados por tabulações. Para fazer isso clique em  onde diz *Save as type* (Salvar como tipo) e uma lista dos possíveis formatos será apresentada. Clique no tipo que você desejar. Uma vez que um arquivo tenha sido previamente salvo, ele poderá ser salvo novamente (atualizado) clicando em . Esse ícone aparece tanto no editor de dados quanto no visualizador de saídas e o arquivo a ser salvo dependerá da janela que estiver ativa no momento. O arquivo será salvo no local em que já estiver armazenado.

2.8 RECUPERANDO UM ARQUIVO ①

Ao longo do livro você irá trabalhar com arquivos de dados que estão disponíveis no *site* www.artmed.com.br. Dessa forma é importante que você saiba como carregar esses arquivos de dados no SPSS. *O procedimento é bastante simples. Para abrir um arquivo simplesmente utilize o ícone .

(ou utilize o menu *File⇒Open⇒Data...* [Arquivo⇒Abrir⇒Dados]) para ativar a caixa de diálogos como na Figura 2.12. Primeiro você precisa encontrar o local onde o arquivo está armazenado (salvo). Se você está carregando o arquivo a partir do CD então acesse o diretório do CD clicando em , onde diz *Look in* (Olhar em) e uma lista de localização de diretórios será mostrada. Uma vez que o diretório do CD tenha sido acessado (clicando em ) você deve ver uma lista de arquivos e pastas que podem ser abertas. Como ocorreu com o salvamento de um arquivo, você está no editor de dados então o SPSS irá mostrar apenas arquivos de dados do SPSS para serem abertos (se você estiver na janela de visualização de saída então somente arquivos de saída serão mostrados). Se você utilizar os menus e utilizar o caminho *File⇒Open⇒Data...* (Arquivo⇒Abrir⇒Dados) então arquivos de dados serão apresentados, mas se utilizar o caminho *File⇒Open⇒Output...* (Arquivo⇒Abrir⇒Saídas) então arquivos de saídas serão exibidos e se usar *File⇒Open⇒Syntax...* (Arquivo⇒Abrir⇒Sintaxe) então arquivos de sintaxe serão mostrados (você deve ter captado a ideia geral). Você pode abrir uma pasta dando um duplo clique sobre o seu ícone. Uma vez que você tenha chegado ao arquivo desejado você poderá abri-lo selecionando-o

* N de R. Na edição original, os arquivos foram disponibilizados em um CD. Nesta edição, os mesmos arquivos devem ser acessados no *site* www.artmed.com.br e salvos em seu computador (conteúdo em inglês).

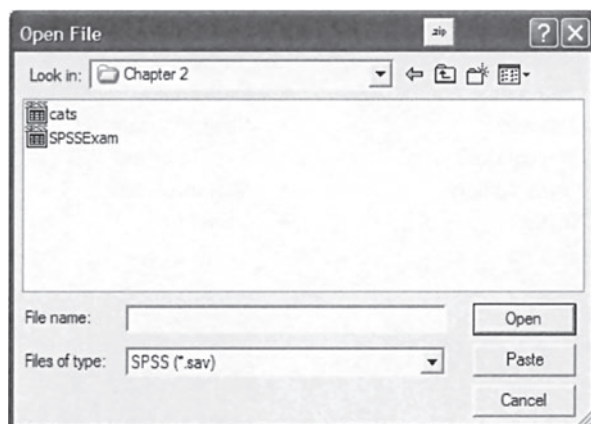


Figura 2.12 Caixa de diálogos para abrir um arquivo.

com o mouse ou então clicando no , ou ainda clicando duas vezes no ícone próximo ao arquivo que deseja abrir (por exemplo, clicando duas vezes em). Os dados do arquivo aparecerão então na janela apropriada. Se você está no editor de dados e quiser abrir um arquivo de saída, clique em onde diz *Files of type* (Arquivos do tipo) e uma lista alternativa de formatos de arquivos será apresentada. Clique no tipo de arquivo apropriado (documento de saída (*.spo), arquivo de sintaxe (*.sps), arquivo do Excel (*.xls), arquivo texto (*.dat, *.txt) e um arquivo de desse tipo estará disponível para ser aberto.

2.9 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Este capítulo forneceu uma introdução básica ao ambiente do SPSS. Vimos que ele utiliza duas janelas principais: um editor de dados e um visualizador de saídas. O editor de dados tem dois painéis um de dados (onde entramos com os valores ou dados) e um de variáveis (onde são definidas as variáveis e suas propriedades). O visualizador é uma janela na qual toda saída aparece, tais como tabelas, estatísticas e gráficos. Nós também criamos nosso primeiro conjunto de dados definindo algumas variáveis e entrando com alguns dados. Ao fazer isso, descobrimos que podemos co-

dificar grupos de pessoas utilizando números (variáveis codificadoras) e descobrimos que linhas no editor de dados representam pessoas (ou casos de dados) e colunas representam diferentes variáveis. Finalmente damos uma olhada na janela de sintaxe e aprendemos como abrir e salvar arquivos.

2.10 TERMOS-CHAVE QUE DESCOBRIMOS

- Dados de razão
- Dados intervalares
- Dados nominais
- Dados ordinais
- Editor de dados
- Editor de sintaxe
- Variáveis numéricas
- Variáveis texto
- Visualizador
- Visualizador de dados

2.11 TAREFA DO ALEX ESPERTO



A tarefa do Alex Esperto para este capítulo é salvar os dados que foram digitados. Salve-os em algum lugar do disco rígido do seu computador (ou em uma mídia móvel se preferir). Dê um título significativo (crie, talvez, uma pasta

chamada “Meus arquivos de dados” em “Meus documentos” um diretório do Windows se você estiver utilizando um PC.①

2.12 LEITURAS COMPLEMENTARES

EINSPRUCH, E. L. *An Introductory guide to SPSS for Windows*. Thousand Oaks (CA): Sage, 1998.

FOSTER, J. J. *Data Analysis using SPSS for Windows: a beginner's guide*. London: Sage, 2001 (2^a ed.).

KINNEAR, P. R., GRAU, C. D. *SPSS for Windows made simple (Release 10)*. Hove: Psychology Press, 2000.

3.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

Como o título sugere, este capítulo ensinará o primeiro passo da análise de dados: explorá-los. Wright (2003), citando Rosenthal, afirma que pesquisadores devem “ficar amigos dos seus dados”. Não, isso não significa que quem utiliza estatística deve ser amigo dos seus dados porque são os únicos amigos que eles têm; em vez disso, Rosenthal quer dizer que os pesquisadores muitas vezes fazem análises apressadas. Wright faz a analogia com um bom vinho; você deve saborear o aroma e os sabores delicados para realmente apreciar a experiência. Isso talvez exagere a beleza da análise de dados; no entanto, apressar a análise é, suponho, parecido com engolir o vinho: o resultado é confuso e incoerente! Este capítulo apresenta algumas formas de exploração de dados. Começaremos verificando como podemos examinar os dados, checar algumas hipóteses básicas sobre eles e então determinar algumas estatísticas descritivas e traçar alguns gráficos.

3.2 DADOS PARAMÉTRICOS ①

3.2.1 Hipóteses “retirar” ①

Muitos dos procedimentos estatísticos descritos neste livro são testes paramétricos baseados na distribuição normal (descrita na Seção 1.5). Um teste paramétrico requer que os dados sejam retirados de um grande catálogo de distribuições descritas por estatísticos e para as quais certas suposições devem ser verdadeiras. Se você utiliza um teste paramétrico quando seus dados não são paramétricos, os resultados talvez não sejam apropriados. Dessa forma, é importante que você verifique as hipóteses antes de decidir qual teste estatístico é apropriado.

Ao longo deste livro você conhecerá minha obsessão com hipóteses e sua checagem. A maioria testes paramétricos baseados na distribuição normal deve preencher quatro hipóteses básicas para que seja preciso. Muitos estudantes acham que testar suposições é uma tarefa entediante e frequentemente ficam confusos sobre como saber se uma hipótese é ou não-verdadeira. Assim, este capítulo foi projetado para levá-lo passo a passo em uma turnê pelo mundo das suposições

Quais são as hipóteses dos testes paramétricos?



paramétricas (oh, que excitante!). Até o momento, você pode achar que essas suposições não são muito atraentes, mas elas podem ter grandes vantagens: pelo menos você pode impressionar seus professores/orientadores listando todas as hipóteses dos

testes que eles violaram ao longo da carreira. Então você pode esquecer, no campo da estatística, as teorias que eles levaram a vida desenvolvendo – e eles não poderão censurá-lo.¹ Bem, como sempre digo, se você precisa encarar a agonia de trabalhar com estatística, deve fazê-la de forma apropriada. As hipóteses dos testes paramétricos são:

1. **Dados normalmente distribuídos:** Assume-se que os dados foram obtidos de uma ou mais populações normais. O fundamento dos testes de hipóteses baseia-se em ter populações distribuídas normalmente. Assim, se essa hipótese não é satisfeita, a lógica que os sustenta será falha (esses princípios já mencionados no Capítulo 1 e serão explicados novamente no Capítulo 6). Muitos pesquisadores verificam suas amostras (utilizando um histograma) e se a amostra parece aproximadamente normal, eles assumem que as populações também serão. Veremos, neste capítulo, que podemos ir um passo além dessa abordagem e testar se a amostra difere significativamente de uma normal.
2. **Homogeneidade da variância:** Essa hipótese significa que as variâncias devem ser

as mesmas para as diferentes populações sendo consideradas. Em delineamentos nos quais vários grupos de participantes são testados, essa hipótese significa que cada uma das amostras foi retirada de populações com a mesma variância. Em delineamentos correlacionais, essa hipótese significa que a variância de uma variável deve ser estável em relação a todos os níveis da outra variável (veja a Seção 3.6).

3. **Dados por intervalo:** Os dados devem ser mensurados pelo menos ao nível ordinal (veja o Quadro 2.1 no Capítulo 2). Isso significa que a distância entre os pontos da sua escala deve ser igual em todas as partes ao longo da escala. Por exemplo, se você tem uma escala de ansiedade de 10 pontos, a diferença de ansiedade representada por uma alteração de 2 para 3 no escore deve ser a mesma representada por uma alteração do escore de 9 para 10.
4. **Independência:** Essa hipótese considera que os dados de participantes diferentes são independentes, o que significa que o comportamento de um participante não influencia o comportamento de outro. Nos delineamentos de medidas repetidas (na qual os participantes são mensurados em mais de uma condição experimental), esperamos que os escores na condição experimental sejam dependentes para um dado participante, mas que o comportamento entre diferentes participantes deve ser independente. Como exemplo, imagine que duas pessoas, Paulo e Júlia, estão participando de um experimento onde eles devem indicar se lembram de ter visto determinadas fotos anteriormente no experimento. Se Paulo e Júlia pudessem conversar sobre se viram ou não certas fotos, então suas respostas *não* seriam independentes. A resposta de Júlia a uma determinada questão dependerá da resposta de Paulo e isso viola a hipótese de independência. Se Paulo e Júlia não pudessem conversar (se estivessem em salas diferentes), suas respostas seriam independentes (a menos que eles fossem telepatas!): a resposta de Paulo não seria influenciada pela de Júlia e vice-versa.

¹ Quando eu estava fazendo meu doutorado, recebemos a tarefa do nosso professor de estatística de pegar algum artigo publicado e criticar os métodos estatísticos utilizados. Escolhi um do meu orientador e resolvi criticar duramente todos os procedimentos de análise de dados (e eu fui *bastante* pedante). Imagine o meu horror quando vi meu orientador aproximando-se aos pulos pelo corredor com um largo sorriso e declarando que, sem que eu soubesse, ele foi o segundo avaliador do meu ensaio. Felizmente, ele teve senso de humor e eu tirei uma boa nota. ☺

As hipóteses de dados intervalares e medidas independentes são, infelizmente, testadas somente pelo senso comum. A hipótese da homogeneidade da variância é testada de forma diferente por diferentes procedimentos e, assim, embora abordemos esta hipótese na Seção 3.6, também iremos discutir esses testes sempre que for necessário. Isso nos deixa somente com a hipótese de normalidade para checar e em muitos aspectos essa é, de qualquer forma, a hipótese mais importante. A maneira mais simples de testar essa hipótese é olhar para a distribuição da amostra. Se os dados da amostra forem normalmente distribuídos, nossa tendência é assumir que eles vieram de uma população normal. Da mesma forma, se a nossa amostra não for normalmente distribuída, iremos assumir que ela veio de uma distribuição não-normal. Assim, para testar essa hipótese, podemos começar plotando com que frequência cada escore ocorre (*a distribuição de frequências*) e ver se ela parece normal. Vamos ver como se faz isso.

3.3 APRESENTANDO DADOS GRAFICAMENTE ①

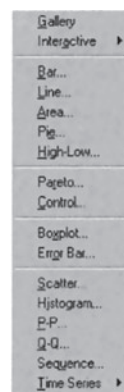
Vamos dar uma olhada em como podemos utilizar uma distribuição de frequências, outros gráficos e estatística descritiva para fazer uma análise completa de nossos dados. Vamos usar um exemplo para ilustrar o que queremos fazer. Uma bióloga está preocupada com os efeitos potenciais dos festivais de música sobre a saúde. Assim, em determinado ano ela foi ao Festival de Música de Glastonbury (os não-britânicos podem fingir que o festival é o Roskilde, Ozzfest, Lollapalooza, Rock in Rio ou outro qualquer!) e verificou a higiene de 810 fãs durante os três dias de duração do festival. Em teoria, cada pessoa foi avaliada diariamente, mas em virtude da dificuldade de fazer isso, faltaram alguns dados nos dias dois e três. A higiene foi mensurada com a utilização de uma técnica padrão (não se preocupe: não foi lamber a axila da pessoa) que resultou em um escore variando de 0 (você está

cheirando como um cadáver em decomposição em cima de um gambá) a 5 (você cheira a rosas em um dia de primavera). Hoje, eu sei, por amargas experiências, que a higiene nem sempre é o forte nesses lugares (o Festival de Reading parece ser particularmente ruim...) e, assim, essa pesquisadora previu que a higiene pessoal diminuiria muito durante os três dias do festival. O arquivo de dados denominado **GlastonburyFestival.sav** pode ser encontrado no *site* www.artmed.com.br (veja a Seção 2.8 para lembrar como abrir um arquivo)

3.3.1 Passo 1: Apontando os erros óbvios utilizando histogramas ①

A primeira coisa a fazer é construir o histograma para ver a forma dos dados. O SPSS tem recursos para construir diferentes tipos de gráficos. Se você clicar no menu **Graphs** (Gráficos), uma lista aparecerá apresentando todas as opções disponíveis. Você já deve conhecer algumas delas (por exemplo, diagramas de *pizza*, barras, etc.):

- **Bar** (Barras): Normalmente utilizado para representar médias de grupos diferentes de pessoas (*Summaries for groups of cases* – resumos para grupos de casos) ou média de variáveis diferentes (*Summaries of separate variables* – resumos de diferentes variáveis).
- **Line** (Linha): Também utilizado para representar médias.
- **Pie** (Torta): Utilizado para apresentar frequências e percentagens.
- **Boxplot** (Caixa e Bigode): Utilizado para mostrar a mediana, espalhamento e intervalo interquartil dos escores.
- **Error Bar** (Barra de Erros): Mostra a média e o intervalo de 95% de confiança em torno dessa média (veja o Capítulo 7).
- **Scatter** (Dispersão): Mostra relações entre duas variáveis (veja o Capítulo 5).



- **Histogram** (Histograma): Mostra a frequência de diferentes escores (útil para visualizar melhor a distribuição dos escores)

Você irá notar que existe também uma opção denominada *Interactive* (Interativa), que abre outra lista com praticamente os mesmos gráficos desse menu. **Os gráficos interativos são na verdade esteticamente mais agradáveis e supostamente devem mudar quando os dados forem alterados (por isso o nome!). Assim, para fazer um histograma você deve selecionar *Graphs⇒Interactive⇒Histogram...*** (Gráficos⇒Interativo⇒Histograma) a fim de abrir uma caixa de diálogo como a da Figura 3.1.

Note que as variáveis do editor de dados estão listadas no painel da esquerda e você pode arrastar cada uma delas para qualquer um dos espaços em branco. Os dois eixos são representados por setas e você simplesmente arrasta a variável para o espaço adequado. No eixo vertical já existe uma variável denominada *Count* (Contagem). Isso, porque queremos contar a frequência de cada escore e o SPSS fará isso. Para começar, vamos fazer o histograma dos dados da higiene para o primeiro dia do festival. Selecione essa variável da lista e arraste para o espaço do eixo horizontal. A caixa de diálogo final deve ser parecida com a da Figura 3.1. Você notará um espaço deno-

minado *Panel variables* (Painel de variáveis) e, se você arrastar a variável para o espaço, o SPSS irá criar um histograma para cada nível dessa variável. Assim, se você tiver uma variável de grupo (codificadora) para gênero e você a colocar no espaço, obterá dois histogramas: um para as mulheres e outro para os homens. As opções padrão (por omissão) são adequadas e, dessa forma, você pode apenas clicar no **OK** (se quiser ajustar uma curva normal ao histograma resultante, deverá selecionar na guia *Histogram* (Histograma) a opção *Normal curve* (Curva normal)).

O histograma resultante é mostrado na Figura 3.2 e você logo notará que existe um caso que é bem diferente dos demais. Todos os escores parecem estar concentrados em um lado da distribuição porque eles são menores do que 5 (formando uma distribuição leptocúrtica) com exceção de um, que apresentou um valor de 20. Ele é um valor atípico (*outlier*): um valor bem diferente dos demais. Valores atípicos distorcem a média e inflacionam o desvio padrão (veja Field e Hole, 2003) e uma análise criteriosa dos dados é uma forma de detectá-los. O estranho desse valor é que ele está acima do valor máximo da escala utilizada (lembre-se: a escala de higiene varia de 0 a 5) e ele deve ser um erro (ou a pessoa tem um transtorno obsessivo-compulsivo e se esfregou tanto que

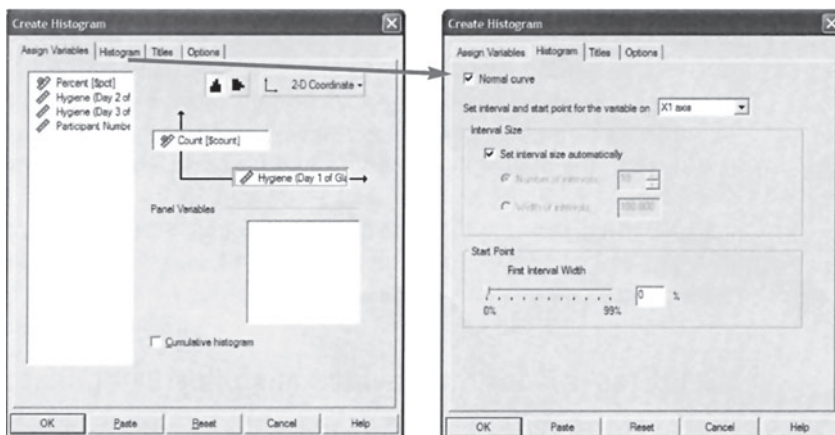


Figura 3.1 Caixa de diálogo para um histograma interativo (mudar as guias no topo da caixa de diálogo altera o conteúdo da mesma).

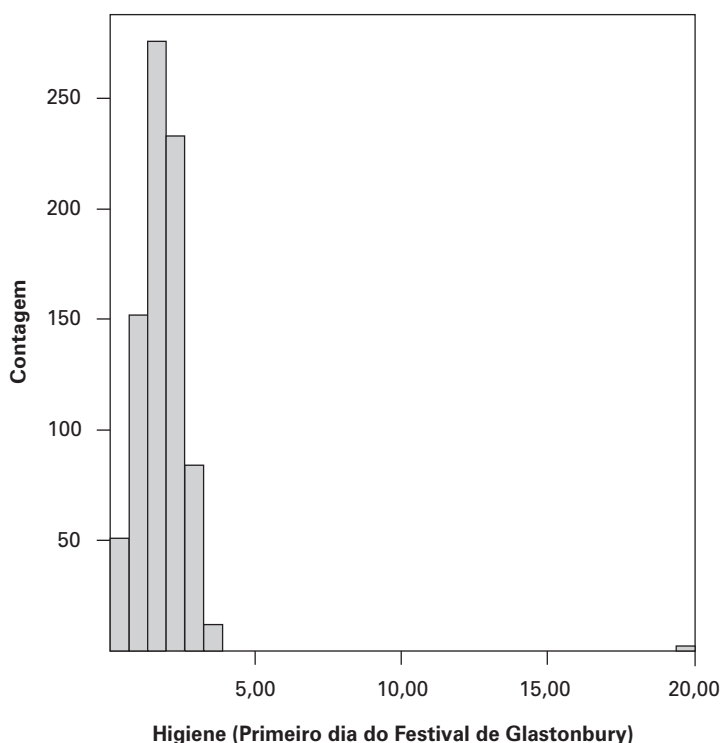


Figura 3.2 Histograma dos escores da higiene (dia 1) do Festival de Glastonbury.

atingiu um estado de limpeza extrema). Entretanto, com 810 casos, como iremos encontrar qual caso foi digitado errado? Podemos vasculhar os dados, mas certamente teremos uma dor de cabeça; em vez disso, podemos traçar um diagrama de **caixa e bigodes**.

Falarei mais sobre diagramas de caixa e bigodes posteriormente neste capítulo, assim, por ora, apenas acredite em mim e selecione **Graphs⇒Boxplot** (Gráficos⇒Diagrama de Caixa e Bigodes) para acessar a caixa de diálogo. Dentro da janela, selecione *Summaries of separate variables* (Diagrama para variáveis separadas) e clique em **Define** (Definir). Isso abrirá outra caixa de diálogo denominada *Boxes Represent* (Os quadros representam). Selecione as três variáveis (representando os três dias do festival) da lista e transfira todas para o painel denominado *Boxes Represent* (Os quadros representam) clicando em **►**. Agora, clique em **Options...** (Opções) e, na caixa

que abrir, selecione *Exclude cases variable by variable* (Exclua casos variável por variável). Precisamos selecionar essa opção porque em algumas das nossas variáveis faltam valores e, se excluirmos os casos por listagem (*Exclude cases listwise*), então se um caso tem um único valor omitido em qualquer uma das três variáveis, ele não será utilizado. Obviamente, não queremos isso; queremos excluir somente os casos para a variável na qual aquele valor esteja faltando. Para tanto, basta selecionar a opção *Exclude cases variable by variable* (excluir casos variável por variável). A caixa de diálogo final será semelhante à Figura 3.3. Clique em **OK** para obter os gráficos.

O diagrama de caixa e bigodes resultante é apresentado na Figura 3.4. É importante notar que o que foi detectado no histograma é mostrado acima do diagrama como um asterisco com um número ao lado. Esse número é a posição do caso que produziu o valor atípico. Ele é

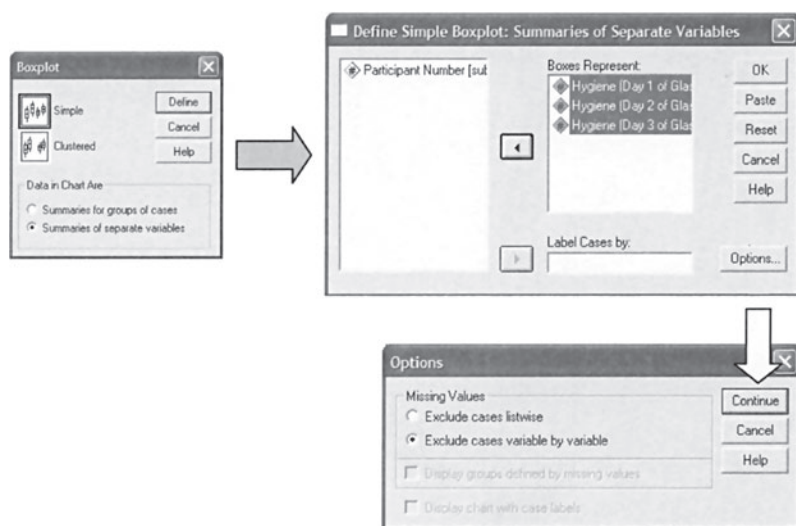


Figura 3.3 Caixa de diálogo para definir um diagrama de caixa e bigodes para diversas variáveis.

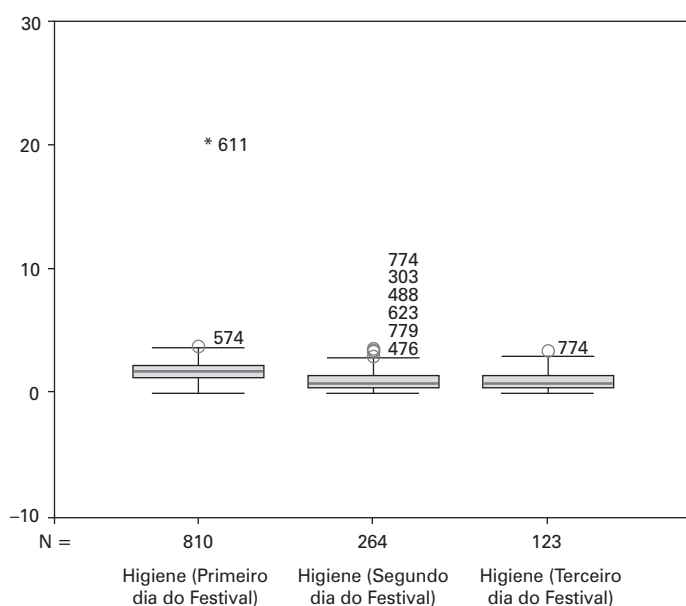



Figura 3.4 Diagrama de caixa e bigodes dos escores da higiene nos três dias do Festival de Glastonbury.

o caso de número 611. Se você for para o editor de dados poderá localizar esse caso rapidamente clicando em  e digitando 611 na caixa de diálogo que aparece. Isso levará você direto

para o caso 611. Olhando para esse caso, você verifica que o valor é 20,02, provavelmente um erro de digitação de 2,02. Você poderá consultar as anotações (questionários ou outro instru-

mento de coleta) para verificar. Vamos assumir que você fez isso e que o valor é de fato 2,02, assim, substitua o valor 20,02 por 2,02 antes de continuar com esse exemplo.

Agora que corrigimos o erro, vamos refazer o histograma. Enquanto fazemos isso devemos obter os histogramas para os dados dos dias dois e três do festival também. A Figura 3.5 mostra os três histogramas resultantes. A primeira coisa a notar é que os dados do primeiro dia parecem mais saudáveis agora que removemos o que foi digitado errado. De fato, a distribuição tem uma aparência surpreendentemente normal: ela é simétrica e não parece

nem pontiaguda nem achatada – isso é bom! No entanto, o mesmo não acontece com as distribuições dos dias dois e três, que não parecem simétricas. Ambas são positivamente assimétricas. Em geral, parece que os dados dos dias dois e três estão mais concentrados em torno dos valores mais baixos da escala. Lembre que quanto mais baixo o escore, menos higiênica a pessoa é, dessa forma, isso sugere que as pessoas estão mais sujas à medida que o festival progride. A assimetria acontece porque uma minoria insiste em manter seus níveis de higiene (contra todas as chances!) ao longo do festival (lenços umedecidos são indispensáveis, eu acho). No entan-

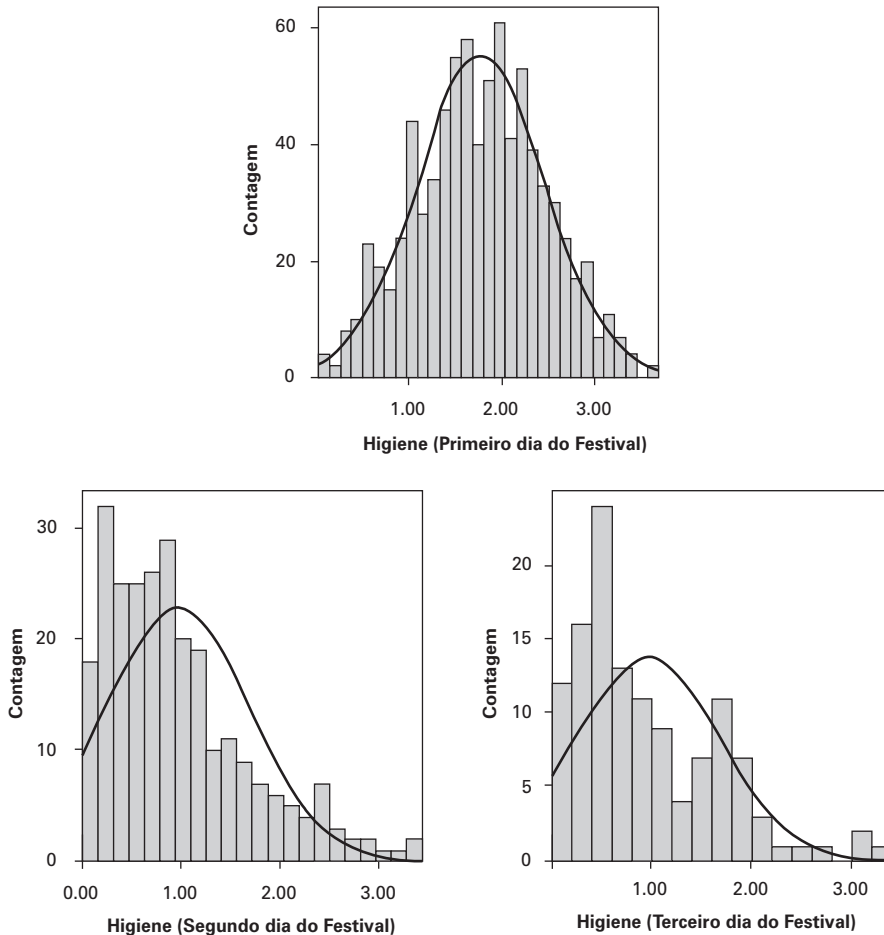





Figura 3.5 Histograma dos escores dos três dias do festival de higiene de Glastonbury.

to, essas distribuições assimétricas podem nos causar problemas se quisermos utilizar testes paramétricos. Na próxima seção, iremos verificar como medir a assimetria e a curtose (achatamento) dessas distribuições.

3.3.2 Passo 2: Estatísticas descritivas e o diagrama de caixa e bigodes ①

Tendo analisado rapidamente as distribuições dos escores da higiene e detectado e corrigido um valor digitado incorretamente, devemos prosseguir procurando uma maneira de quantificar a forma das distribuições e detectar valores atípicos. Para avançar na exploração das distribuições das variáveis, podemos utilizar o comando *frequencies* (frequências), que é acessado utilizando o caminho **Analyze⇒Descriptive Statistics⇒Frequencies...**² (Analisar⇒Estatística Descritiva⇒Frequências...).

² Lembre que esse caminho de menu será **Statistics⇒Summarize⇒Frequencies...** (Estatística⇒Resumir⇒Frequências) na versão 8.0 e anteriores.

A caixa de diálogo principal é apresentada na Figura 3.6. As variáveis no editor de dados devem estar listadas no painel da esquerda e elas podem ser transferidas para o quadro denominado *Variable(s)* (Variável[is]) clicando na variável (ou destacando diversas com o mouse) e depois clicando em . Qualquer análise escolhida será realizada em todas as variáveis listadas no quadro *Variable(s)* (Variável[is]). Se uma variável listada no quadro *Variable(s)* (Variável[is]) é selecionada com o mouse, ela pode ser transferida de volta para a lista de variáveis clicando no botão com a seta (que deve agora estar apontando para a direção oposta). Por padrão, o SPSS produz uma distribuição de frequências de todos os escores no formato de tabela. No entanto, existem duas outras caixas de diálogo que podem ser utilizadas para produzir outras opções. A caixa de diálogo *Statistics* (Estatísticas) é acessada clicando em  e a caixa de diálogo *charts* (gráficos) pode ser ativada clicando em .

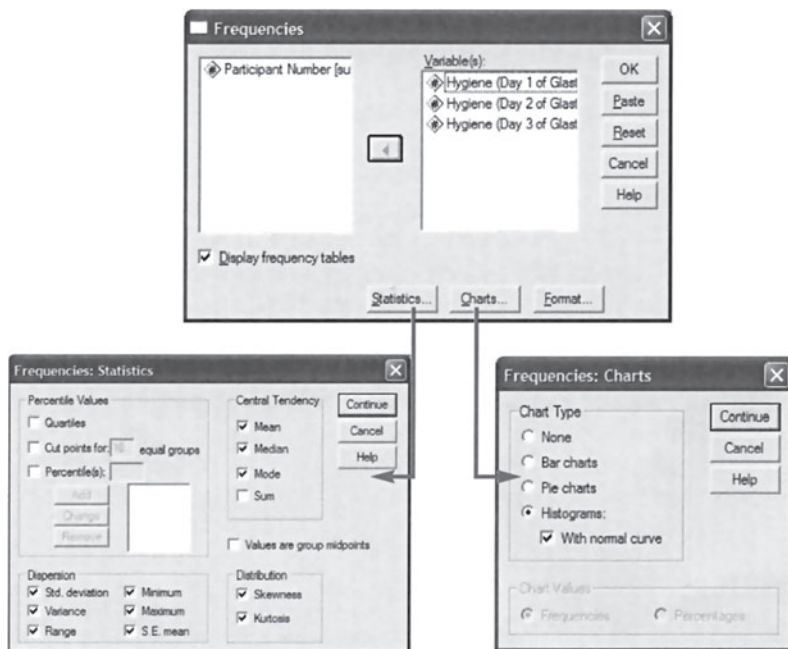


Figura 3.6 Caixa de diálogo para o comando *frequencies* (frequências).

A caixa de diálogo *Statistics* (Estatística) permite a seleção de várias opções de maneiras como a distribuição dos escores pode ser descrita, como medidas de tendência central (média, moda, mediana), medidas de variabilidade (amplitude, desvio padrão, variância, quartis), medidas de forma (assimetria e curtose). Para descrever as características dos dados, devemos selecionar a média, moda, mediana, desvio padrão, variância e amplitude. Para verificar se uma distribuição é normal, precisamos olhar os valores da assimetria e da curtose. A opção *chart* (diagramas) fornece uma maneira simples de representar graficamente a distribuição dos escores (como um gráfico de barras, um de *pizza* ou um histograma). Já obtivemos histogramas dos nossos dados, assim, não precisamos selecionar essa opção, mas poderemos utilizá-la em análises futuras. Quando você tiver selecionado as opções adequadas, retorne à caixa de diálogo principal clicando em **Continue**. Uma vez lá, clique em **OK** para obter a análise.

A saída 3.1 do SPSS mostra a tabela das estatísticas descritivas para as três variáveis desse exemplo. Dessa tabela, podemos verificar que, em média, os escores da higiene foram 1,77 (numa escala de 5) no primeiro dia do festival, mas caíram para 0,96 e 0,98 nos dias dois e três, respectivamente.

As outras medidas importantes para o nosso objetivo são a assimetria e a curtose; ambas apresentam um erro padrão associado. O valor da assimetria e da curtose deverão ser zero em uma distribuição normal. Valores de assimetria positivos indicam uma concentração de valores à esquerda enquanto um valor negativo mostra uma concentração de valores à direita. Valores positivos da curtose indicam uma distribuição pontiaguda e valores negativos indicam uma achatada. Quanto mais distante esses valores estiverem de zero, maior a possibilidade de que os dados não sejam normais. Entretanto, os valores reais da assimetria e curtoses não são, eles próprios, informativos. Precisamos padronizar os valores, isto é, transformá-los em escores-*z*. Vimos na Seção 1.5.3 que um

escore-*z* apresenta média igual a 0 e desvio padrão igual a 1. O motivo para a conversão dos valores para escores-*z* é que dessa forma eles estarão padronizados. Assim, podemos converter qualquer variável mensurada em qualquer unidade em escores-*z*. Com essa transformação, é possível comparar valores mesmo que tenham sido mensurados em unidades diferentes. Para transformar um valor em um escore-*z*, basta subtrair a média da distribuição e dividir o resultado pelo desvio padrão da distribuição. Assimetria e curtose são convertidas em escores-*z* da mesma forma.

$$Z_{\text{Assimetria}} = \frac{S - 0}{EP_{\text{Assimetria}}} \quad Z_{\text{Curtose}} = \frac{K - 0}{EP_{\text{Curtose}}}$$

Nas equações acima, os valores de *S* (Assimetria) e *K* (Curtose) e seus respectivos erros padrões são fornecidos pelo SPSS. Esses escores-*z* podem ser comparados com os valores que esperaríamos serem obtidos apenas por acaso (isto é, valores conhecidos para a distribuição normal mostrados no Apêndice). Assim, um valor absoluto maior do que 1,96 é significativo com $p < 0,05$, e um valor acima de 2,58 é significativo com $p < 0,01$. Valores absolutos acima de 3,29 são significativos com $p < 0,001$. Amostras grandes fornecerão erros padrão pequenos e, desse modo, quando os tamanhos amostrais são grandes, valores significativos serão obtidos mesmo com pequenos desvios da normalidade. Em muitas amostras, valores acima de 1,96 serão suficientes, no entanto, para amostras grandes esse critério deve ser aumentado para 2,58 e com amostras muito grandes, em virtude do problema dos erros padrão pequenos já comentado, nenhum critério deve ser utilizado! Se tivermos uma amostra grande (200 ou mais), é mais importante verificar visualmente a forma da distribuição e procurar os valores da assimetria e da curtose em vez de calcular suas significâncias.

Para os valores da higiene, o escore-*z* da assimetria é $-0,004/0,086 = 0,047$ no primeiro dia, $1,095/0,150 = 7,300$ no segundo dia e $1,033/0,218 = 4,739$ no terceiro dia. Fica

Dica da Samanta Ferrinho



- Para saber se a distribuição dos escores é normal precisamos verificar os valores da assimetria e da curtose na saída do SPSS.
- Valores positivos da curtose indicam uma distribuição pontiaguda e negativos, uma achatada.
- Quanto mais longe os valores estiverem de zero maior a probabilidade de que os dados não sejam normalmente distribuídos.

Saída 3.1 do SPSS

Statistics (Estatísticas)

		Hygiene (Day 1 of Glastonbury Festival (Higiene do Primeiro dia do Festival de Glastonbury)	Hygiene (Day 2 of Glastonbury Festival (Higiene do Segundo dia do Festival de Glastonbury)	Hygiene (Day 1 of Glastonbury Festival (Higiene do Terceiro dia do Festival de Glastonbury)
N	Valid (Válidos) Missing (Desconhecidos)	810 0	264 546	123 687
Mean (Média)		1.7711	0.9609	0.9765
Std. Error of Mean (Erro Padrão da Média)		0.02437	0.04436	0.06404
Median (Mediana)		1.7900	0.7900	0.7600
Mode (Moda)		2.00	0.23	0.44
Standard Deviation (Desvio Padrão)		0.69354	0.72078	0.71028
Variance (Variância)		0.48100	0.51952	0.50449
Skewness (Assimetria)		0.004	1.095	1.033
Std. Error of Skewness (Erro Padrão da Assimetria)		0.086	0.150	0.218
Kurtosis (Curtose)		0.410	0.822	0.732
Std. Error of Kurtosis (Erro Padrão da Curtose)		0.172	0.299	0.433
Range (Amplitude)		3.67	3.44	3.39
Minimum (Mínimo)		0.02	0.00	0.02
Maximum (Máximo)		3.69	3.44	3.41

Existem múltiplas modas. O menor valor é mostrado.

claro então, que embora no primeiro dia os escores sejam simétricos, nos dias 2 e 3 eles apresentam uma assimetria positiva significativa mostrados pelos escores-*z* acima de 1,96 (como já tinha ficado evidente pelos histogramas). No entanto, lembre o que foi dito sobre grandes amostras! Os escores-*z* da curtose são: $-0,410/0,172 = -2,38$ no primeiro dia, $0,822/0,299 = 2,75$ no segundo e $0,732/0,433 = 1,69$ no terceiro dia. Esses valores indicam curtose significativa (ao nível de 5%) em todos os três dias, mas em virtude da amostra grande isso não é uma surpresa e podemos ficar ali-

viados porque os três valores estão abaixo do nosso valor limite de 3,29.

A saída fornece a distribuição de frequências tabelada de cada variável (não reproduzidas aqui). Essas tabelas listam cada escore e o número de vezes que cada um aparece no conjunto de dados. Além disso, cada valor da frequência é expresso como um percentual do total da amostra. Também é fornecido o percentual acumulado, que nos informa quantos casos (em percentual) estão abaixo de um determinado valor. Assim, por exemplo, podemos ver que somente 15,4% dos escores da higiene

estão abaixo de 1 no primeiro dia do festival. Compare isso com a tabela do segundo dia: 63,3% dos escores foram menores do que 1!

Tendo verificado a forma da distribuição, precisamos procurar valores atípicos (*outliers*) (veja o Quadro 3.1). Existem duas formas de

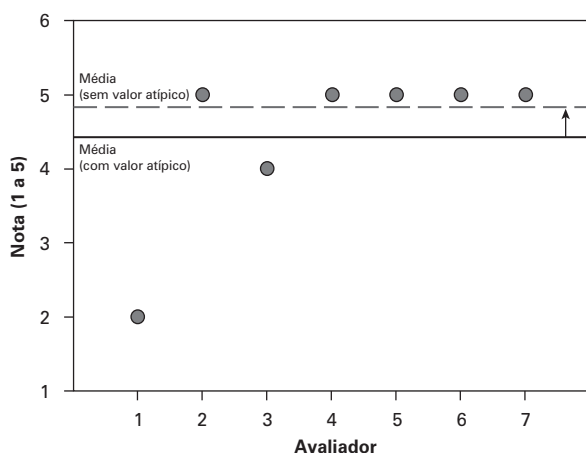
fazer isso: (1) olhar o diagrama de caixa e bigodes e (2) olhar os escores-*z*. Deixarei que o Alex Esperto lide com os escores-*z* no Quadro 3.2 e ficarei com os belos diagramas que são mais adequados ao meu nível de conhecimento. Já mostramos que o SPSS pode criar

Quadro 3.1

Valores atípicos ①



Um valor atípico (*outlier*) é um escore bastante diferente do resto dos dados. Quando analisamos dados devemos estar cientes de tais valores porque eles podem introduzir tendenciosidades no modelo que ajustamos aos dados. Um bom exemplo dessa tendenciosidade pode ser vista em um modelo estatístico simples como a média. Field e Hole (2003) apresentam um exemplo da avaliação da primeira edição deste livro na Amazon.co.uk. Esses valores podem variar de uma a cinco estrelas. Quando ele foi lançado, sete pessoas fizeram uma avaliação e deram uma nota para a primeira edição com valores variando (na ordem em que foram obtidos) de 2, 5, 4, 5, 5, 5, 5. Todas exceto uma dessas avaliações foram semelhantes (principalmente 4 e 5), mas o primeiro valor, de 2, foi bastante diferente dos restantes. No gráfico são mostrados os sete avaliadores no eixo horizontal e as notas atribuídas no eixo vertical. Nesse gráfico existe



uma linha horizontal que representa a média dos sete valores (4,43 como verificado). Deve ficar claro que todos os escores, exceto um, ficam próximos dessa linha. O valor 2 é bastante diferente e está situado bem abaixo da média. Esse escore é um exemplo de valor atípico (*outlier*). A linha horizontal pontilhada representa a média dos escores (4,83) quando o valor atípico não é incluído. A linha é mais alta do que a original, indicando que se ignorarmos o escore atípico, a média aumenta (aumenta em 0,4). Isso mostra como um único valor pode viésar a média; nesse caso, o pri-

meiro escore (de dois) pode baixar a média. Em termos práticos, isso tem uma implicação grande porque a Amazon arredonda para meios valores, assim, esse único valor faz diferença entre o valor registrado pela empresa de 4,5 e o valor de 5 estrelas que seria obtido sem a consideração do valor atípico (embora eu esteja ressentido com isso, agradeço o ótimo exemplo que eles estão me fornecendo um ótimo exemplo de um valor atípico!)

(Os dados desse exemplo foram obtidos no site <http://www.amazon.co.uk/>.)

diagramas de caixa e bigodes na Figura 3.3; repita esse processo para os dados da higiene. O diagrama resultante agora deve parecer diferente (lembre que alteramos o valor incorreto) e ele é mostrado na Figura 3.7.

O que é um diagrama de caixa e bigodes?



Os diagramas de caixa e bigodes na Figura 3.7 nos informam sobre a distribuição dos escores da higiene dos três dias do Festival de Glastonbury. O diagrama de caixa e bigodes nos mostra o menor escore (a linha horizontal inferior em cada figura) e o maior (a linha

superior horizontal de cada figura). A distância entre a linha mais inferior e a aresta inferior da caixa é a amplitude onde os 25% escores mais baixos podem ser encontrados (é denominado *quartil inferior*). A caixa mostra

os 50% dos escores situados no meio do conjunto de valores (denominado *intervalo inter-quartilico*): isto é, 50% dos escores são maiores do que a parte mais baixa da figura, mas menores do que a parte superior da figura. A distância entre a aresta superior da caixa e a linha horizontal superior mostra o intervalo onde os 25% maiores escores poderão ser encontrados (*quartil superior*). No meio da caixa há uma linha horizontal levemente mais grossa que as arestas (bordas) da caixa. Essa linha representa a **mediana**, que seria o escore do meio se todos os escores da higiene fossem colocados em ordem. Assim, esses diagramas nos mostram o intervalo dos dados, o intervalo onde estão situados os 50% dos valores do meio do conjunto e o escore mediano. Assim como os histogramas, eles também nos informam se a distribuição é simétrica ou não. No primeiro dia, em que a distribuição é simétrica, os bigodes em cada lado apresentam o

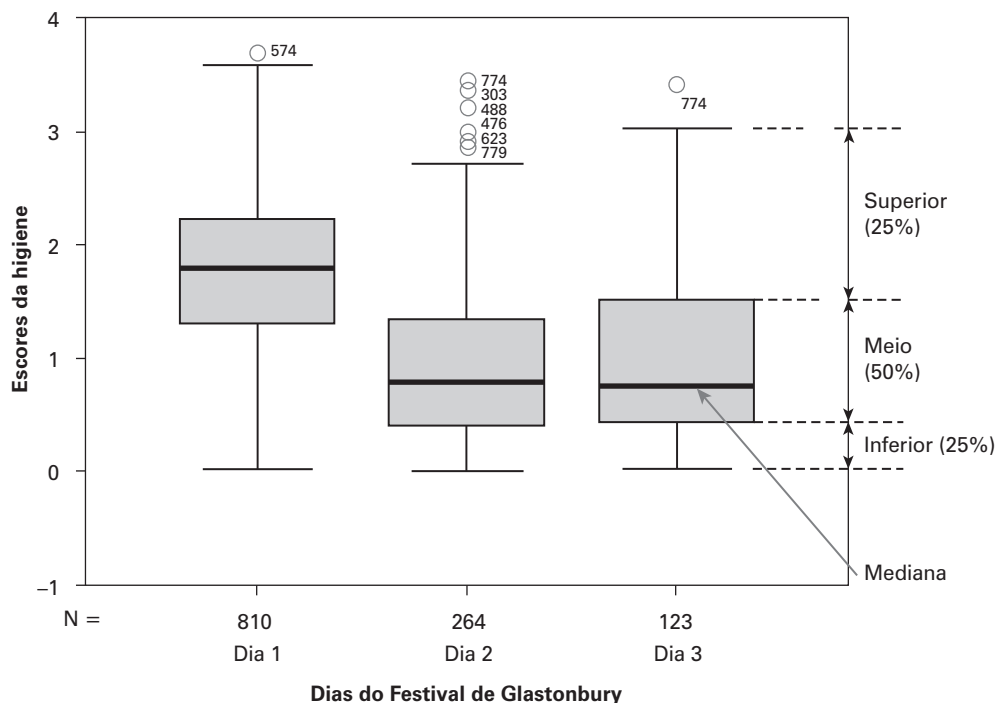


Figura 3.7 Diagrama de caixa e bigodes dos escores de higiene de três dias do Festival de Glastonbury.

Quadro 3.2

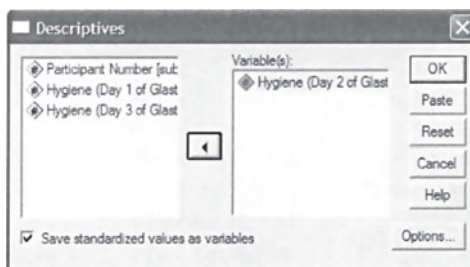
Utilizando os escores-z para encontrar valores atípicos ③



Para verificar a existência de valores atípicos, podemos olhar os escores-z. Eles são simplesmente uma maneira de padronizar os nossos dados. Com isso, quero dizer que os escores são expressos em termos da média e do desvio padrão da distribuição. Especificamente, expressamos os escores em termos de uma distribuição que apresenta média igual a 0 e desvio padrão igual a 1 (veja a Seção 1.5). A vantagem de fazer isso é que é possível estabelecer pontos de referência (*benchmarks*) que podemos aplicar a qualquer conjunto de dados (a despeito da média e do desvio padrão que eles apresentem originalmente). É fácil converter um dado em um escore-z: você simplesmente pega um valor (X), subtrai a média de todos os escores (\bar{X}) dele e divide o resultado pelo desvio padrão de todos os escores (s):

$$z = \frac{x - \bar{x}}{s}$$

Em vez de fazer isso para cada escore manualmente, podemos solicitar que o SPSS faça isso por nós utilizando a caixa de diálogo **Analyze⇒Descriptive Statistics⇒Descriptives...** (Analisar⇒Estatística Descritiva⇒Descritivas...), selecionando a variável (como os dados da higiene do segundo dia conforme diagrama), ou diversas variáveis, e selecionando *Save standardized values as variables* (Salvar valores padronizados como variáveis) antes de clicar em **OK**. O SPSS criará uma nova variável no editor de dados (com o mesmo nome, mas antecedido pela letra z). Podemos utilizar esses escores e contar quantos estão situados dentro de certos limites previamente definidos. Se tomarmos o valor absoluto (isto é, ignorarmos se o escore-z é positivo ou negativo), em uma distribuição normal esperaríamos que aproximadamente 5% tenham valores superiores a 1,96 (às vezes 2 é utilizado por conveniência), 1% tenham valores maiores que 2,58 e praticamente nenhum esteja acima de 3,29.



Alternativamente, é possível utilizar alguma sintaxe do SPSS em uma janela de sintaxe para criar os escores-z e contá-los. No site www.artmed.com.br, há um arquivo de sintaxe denominado **Outliers (Percentage of Z-escores).sps** (*Valores atípicos[Porcentagem de escores-z]*), que escrevi de modo a produzir uma tabela para os dados da higiene do segundo dia do Festival de Glastonbury. Carregue esse arquivo e execute a sintaxe, ou abra uma janela de sintaxe (veja a Seção 2.6) e digite o seguinte (lembre-se de colocar todos os pontos finais) – as explicações do código estão entre * (asteriscos) e não precisam ser digitadas!

DESCRIPTIVES

VARIABLES=day2/SAVE.

*Aqui é utilizada a função *descriptives* (descritivas) do SPSS na variável day2 (em vez de utilizar a caixa de diálogo) para determinar e salvar os escores-z no editor de dados (eles serão apresentados como uma nova variável denominada zday2).*

COMPUTE outlier1 = abs(zday2).

EXECUTE

(Continua)

Quadro 3.2 (Continuação)

*Isso cria uma nova variável no editor de dados denominada *outlier1*, que contém os valores absolutos dos escores-z recém-criados.*

RECODE

Outlier1 (3.29 thru Highest = 4) (2.58 thru Highest = 3) (1.96 thru Highest = 2) (Lowest thru 2 = 1).

EXECUTE.

Isso recodifica a variável denominada outlier1 de acordo com os pontos de referência que descrevi. Assim, se um valor é maior do que 3,29, ele será codificado como 4, se está entre 2,58 e 3,29, seu código será 3, se estiver entre 1,96 e 2,58, o código é 2, e se for menor do 1,98, será codificado como 1.

VALUES LABEL outlier1

1 'Absolute z-score less than 2' 2 'Absolute z-score greater than 1.96' 3 'Absolute z-score greater than 2.58' 4 'Absolute z-score greater than 3.29'.

* Isso atribuirá rótulos apropriados para os códigos definidos acima.*

FREQUENCIES

VARIABLES=outlier1

/ORDER=ANALYSIS.

*Finalmente, essa sintaxe utiliza o recurso *frequencies* (frequências) do SPSS para produzir uma tabela nos mostrando o percentual de uns, dois, três e quatros encontrados na variável outlier1.*

Caso atípico 1

		Frequency (Frequências)	Percent (Percentual)	Valid Percent (Percentual de válidos)	Percent (Percentual)
Valid	<i>Absolute z-score less than 2</i> (Valor absoluto (Válido) do escore-z menor do que 2)	246	30.4	93.2	93.2
	<i>Absolute z-score greater than 1.96</i> (Valor absoluto do escore-z maior do que 1,96)	12	1.5	4.5	97.7
	<i>Absolute z-score greater than 2.58</i> (Valor absoluto do escore-z maior do que 2,58)	4	0.5	1.5	99.2
	<i>Absolute z-score greater than 3.29</i> (Valor absoluto do escore-z maior do que 3,29)	2	0.2	0.8	100.0
	Total	264	32.6	100.0	
Missing	<i>System</i> (Sistema desconhecido)	546	67.4		
Total		810	100.0		

A tabela produzida pela sintaxe é mostrada. Olhe para a coluna denominada “Valid Percent” (Percentual de válidos). Esperaríamos ver 95% dos casos com valores absolutos menores do que 1,96, 5% (ou menos) com um valor absoluto maior do que 1,96 e 1% (ou menos) com um valor absoluto maior do que 2,58. Finalmente, não esperaríamos casos acima de 3,29 (bem, esses casos são valores atípicos significativos). Para os escores de higiene do segundo dia do Festival de Glastonbury, 93,2% dos valores tem escores-z menores do que 1,96. Em outras palavras, 6,8% estão acima (olhando na tabela, obtemos esse valor pela soma de 4,5% + 1,5% + 0,8%). Esse resultado está levemente acima dos 5% esperaríamos fosse normal. Olhando para os valores acima de 2,58, esperaríamos encontrar apenas 1%, mas novamente temos aqui um valor maior, nesse caso, 2,3% (1,5% + 0,8%). Finalmente, descobrimos que 0,8% dos casos estão acima de 3,29 (assim, 0,8% são valores atípicos significativos). Isso sugere que existem muitos valores atípicos nesse conjunto de dados e talvez tenhamos que fazer algo a respeito!



mesmo tamanho (o intervalo entre os 25% valores superiores e os 25% inferiores são iguais). No entanto, nos dias dois e três, os bigodes situados no topo são maiores do que os situados embaixo, mostrando que a distribuição é assimétrica (isto é, 25% dos escores superiores estão espalhados em um intervalo maior do que os 25% escores inferiores). Finalmente, você terá notado alguns círculos acima de cada um dos diagramas. Esses casos são os valores atípicos (*outliers*). Cada círculo apresenta um número próximo a ele, informando em que linha do editor de dados esse caso poderá ser encontrado. Na próxima seção, veremos o que fazer a respeito desses valores atípicos.

3.3.3 Passo 3: Corrigindo problemas nos dados ②



A última seção nos mostrou várias formas de explorar os nossos dados; vimos como olhar para os problemas com a nossa distribuição de escores e como detectar valores atípicos. A próxima questão é o que fazer a respeito desses problemas. Se você detectar valores atí-

picos nos dados, existem várias formas de reduzir o impacto desses valores. No entanto, antes que você execute qualquer uma delas, vale à pena verificar se os dados foram digitados corretamente em cada um dos casos problemáticos. Se os dados estiverem corretos, as três principais opções são:

1. **Remover o caso:** Isso irá obrigá-lo a apagar os dados da pessoa que forneceu o valor atípico. Entretanto, isso só deve ser feito se você tiver uma boa razão para acreditar que esse valor não é um representante da população que você pretende amostrar. Por exemplo, se você está investigando fatores que afetam quanto os gatos ronronam e um gato nunca ronrona, isso será um valor atípico (todos os

gatos ronronam). Após a inspeção, se você descobrir que esse gato era de fato um cachorro (o motivo dele não ronronar) vestido de gato, teremos motivos para excluir esse caso, porque ele vem de uma população diferente (cachorros que gostam de se vestir como gatos) do que a população-alvo (gatos).

2. **Transformar os dados:** Se você tiver uma distribuição não-normal, isso deverá ser feito de qualquer modo (distribuições assimétricas terão, por natureza, valores atípicos, pois são eles os responsáveis pela falta de simetria!). Tais transformações (ver a seguir) devem reduzir o impacto desses valores.
3. **Substituir o valor:** Se a transformação falhar, você pode considerar a substituição do valor. Colocado dessa forma, isso soa como uma falsificação (você está alterando o valor daquilo que foi efetivamente obtido); no entanto, se o valor que você está alterando não é representativo e ainda está distorcendo o modelo, mudar o valor é o menor dos males! Existem várias opções para substituir um valor:

- (a) **O próximo escore mais alto adicionado de um:** mude o dado para uma unidade maior do que o próximo valor mais alto do conjunto de dados.
- (b) **Inverta o valor do escore-z:** Um escore-z de 3,29 é um valor atípico (veja o Quadro 3.2), assim, podemos calcular qual escore forneceu o valor de 3,29 (ou talvez 3) invertendo a equação do Quadro 3.2 que irá fornecer: $x = z s + \bar{X}$. Tudo isso significa que calculamos a média (\bar{X}) e o desvio padrão (s) dos dados, sabemos que z é 3 (precisamente 3,29) e, assim, apenas adicionamos o triplo do desvio padrão à média e substituímos o valor atípico por esse.
- (c) **A média mais dois desvios padrão:** Uma variação do método acima é utilizar a média mais dois desvios padrão (em vez da média mais três desvios padrão).

A próxima seção é de arrepiar os cabelos, portanto, não se preocupe se ela não fizer muito sentido. Muitos cursos de graduação não abordam a transformação dos dados assim, sintam-se à vontade para ignorar esta seção se você quiser!

De todas as opções acima, transformar os dados é talvez a melhor, pois em vez de mudar um único valor, uma alteração é feita em todos eles (assim, você não está apenas selecionando um escore para ser alterado, mas fazendo algo para reduzir o impacto dos valores extremos). Dessa forma, a idéia por trás da transformação é mudar todos os dados para corrigir problemas relacionados ao modelo ou valores atípicos. Embora alguns estudantes muitas vezes (compreensivelmente) pensem que transformar os dados seja errado (a frase “falsificar os resultados” vem à mente de algumas pessoas!), de fato, não é isso, pois você faz a mesma coisa com todos os dados.³ Assim, a transformação dos dados não irá alterar o relacionamento entre variáveis (as diferenças relativas entre pessoas para uma dada variável permanecem as mesmas), mas ela altera as diferenças entre as variáveis (porque altera as unidades de medida). Dessa forma, mesmo se você tiver apenas uma variável que apresenta uma distribuição assimétrica, ainda deve transformar qualquer outra variável do seu conjunto de dados se irá comparar diferenças entre essa variável e outras que pretende transformar. Assim, por exemplo, com os dados da higiene, você pode querer observar como os níveis de higiene se alteraram ao longo dos três dias (isto é, comparar a média do dia 1 com as médias dos dias 2 e 3). Os dados dos dias 2 e 3 são assimétricos e precisam ser transformados, mas em virtude de podermos querer mais tarde comparar os dados aos do

primeiro dia, devemos transformar também os dados do primeiro dia (mesmo que eles não sejam assimétricos). Se não transformarmos os dados do primeiro dia, qualquer diferença nos escores da higiene que encontrarmos entre os dias 1, 2 e 3 será por termos transformado uma variável e não as demais.

Quando você tiver dados com assimetria positiva (como ocorre aqui) que precisam ser corrigidos, existem três transformações que podem ser úteis (como veremos, elas podem ser adaptadas para lidar com dados com assimetria negativa também):⁴

1. Transformação logarítmica ($\log(X_i)$):

Tomar o logaritmo de um conjunto de números esmaga a cauda direita da distribuição. Portanto, é uma boa maneira de reduzir uma assimetria positiva. No entanto, você não pode obter o logaritmo de zero ou de valores negativos, assim, se os seus dados tendem para zero ou produzem números negativos, é necessário adicionar uma constante a todos os valores antes de realizar a transformação. Por exemplo, se você tiver zeros nos dados, faça $\log(X_i + 1)$, ou se tiver valores negativos, adicione qualquer valor que faça o menor dos valores do conjunto se tornar positivo.

2. Transformação por radiciação ($\sqrt{X_i}$):

Tomar a raiz quadrada de valores grandes tem um efeito maior do que extrair a raiz de pequenos valores. Consequentemente, tomar a raiz quadrada de cada um dos valores fará com que cada valor grande fique mais próximo do centro, parecido com a transformação logarítmica. Isso poderá ser útil para reduzir dados com assimetria positiva; no entanto, você ainda terá o mesmo problema com os números negativos (eles não apresentam raízes quadradas).

³ Embora não existam consequências estatísticas em transformar os dados, como Grayson (2004) aponta, podem existir implicações empíricas ou científicas que se sobreponham os benefícios estatísticos. Especificamente, transformar significa que estamos agora trabalhando em um construto diferente do que foi originalmente mensurado e isso acarreta implicações óbvias na interpretação desses dados.

⁴ Você notará, nesta seção, que continuo escrevendo X_i . Vimos no Capítulo 1 que ele se refere ao escore da i -ésima pessoa (assim, o i pode ser substituído pelo nome dessa pessoa em particular, por exemplo, Gregório, $X_i = X_{\text{Gregório}} = \text{Escore do Gregório}$, e para Carol, $X_i = X_{\text{Carol}} = \text{Escore da Carol}$).


3. **Transformação recíproca ($1/X_i$):** Dividir 1 por cada escore reduz o impacto dos grandes valores. A variável transformada terá um limite inferior de 0 (grandes valores ficarão próximos de zero). Uma coisa deve ser levada em consideração para executar esse tipo de transformação: ela reverte os escores, ou seja, valores que eram originalmente grandes no conjunto de dados se tornarão pequenos no conjunto transformado (próximos a zero), mas escores que eram originalmente pequenos se tornarão grandes após a transformação. Por exemplo, imagine dois escores de 1 e 10. Após a transformação, eles se tornarão $1/1 = 1$ e $1/10 = 0,1$: o escore pequeno se torna maior do que o escore grande depois da transformação. No entanto, você pode evitar isso revertendo os escores antes da transformação, encontrando o maior valor e mudando cada escore para o mais alto menos o escore que quer transformar. Assim, a seguinte transformação deverá ser realizada: $1/(X_{\text{Máximo}} - X_i)$.


Cada uma dessas transformações também pode ser utilizada para corrigir dados com assimetria negativa, mas primeiro será necessário reverter os valores (isto é, subtrair cada um dos valores do valor máximo observado). Se você fizer isso, não se esqueça de reverter os valores novamente depois!


3.3.4 Passo 4: Transformando dados utilizando o SPSS ②


O próximo assunto é como realizar essas transformações no SPSS, e isso nos leva ao comando *compute* (calcular). Esse comando permite executar várias funções nas colunas dos dados no editor. Algumas funções típicas são adicionar valores de várias colunas, calcular a raiz quadrada dos valores de uma determinada coluna, ou calcular a média de diversas variáveis (colunas). Para obter a caixa de diálogo do comando, utilize o mouse para especificar **Transform⇒Compute...** (Transformar⇒Calcular...). A caixa de diálogo resultante é mostrada na Figura 3.8. Nessa caixa de diálogo,


existe uma lista de funções no painel do lado direito e um teclado semelhante a uma calculadora no centro. A maioria das funções da calculadora são óbvias e as mais comuns são:

 **Addition** (Adição): Esse botão coloca um símbolo de mais na área de comandos. Por exemplo, com nossos dados da higiene, “day1 + day2” cria uma coluna onde cada linha contém um escore da higiene da coluna denominada *day1* somado ao escore da coluna denominada *day2* (assim, para o primeiro participante: $2,65 + 1,35 = 4$).



 **Subtraction** (Subtração): Esse botão coloca um sinal de menos na área de comandos. Por exemplo, se quisermos calcular a mudança ocorrida na higiene do primeiro dia em relação ao segundo dia, podemos digitar “day2 – day1”. Isso criará uma coluna onde cada linha contém um valor da coluna denominada *day2* subtraído do valor da coluna denominada *day1* (assim, para o participante 1: $2,65 - 1,35 = -1,30$). Dessa forma, a higiene da pessoa caiu 1,30 (na nossa escala de cinco pontos) do primeiro dia para o segundo dia do festival.


 **Multiply** (Multiplicar). Esse botão coloca um sinal de multiplicação na área de comandos. Por exemplo, “day1*day2” cria uma coluna que conterá os escores da coluna denominada *day1* multiplicado pelo escore da coluna denominada *day2* (assim, para o participante 1: $2,65 \times 1,35 = 3,58$).



 **Divide** (Dividir). Esse botão coloca um sinal de divisão na área de comandos. Por exemplo, “day1/day2” cria uma coluna que conterá os escores da coluna denominada *day1* divididos pelos escores da coluna denominada *day2* (assim, para o participante 1: $2,65/1,35 = 1,96$).


 **Exponentiation** (Exponenciação). Esse botão é utilizado para elevar o termo a uma potência especificada. Assim, “day1**2” cria uma coluna que conterá os escores da coluna denominada *day1*


elevada à potência 2 (assim, para o participante 1: $2,65^2 = 7,02$). Da mesma forma “day1**3” cria uma coluna com os valores da variável *day1* ao cubo.

 **Less than** (Menor que). Essa operação é normalmente utilizada pela função *include case* (incluir casos). Se você clicar no botão , uma caixa de diálogo aparece e permite que sejam selecionados certos casos nos quais é possível executar alguma operação. Assim, se for digitado “day1 < 1”, o SPSS executará a função *compute* (calcular) apenas para aqueles participantes cujo escore de higiene no primeiro dia do festival sejam menores do que 1 (isto é, se *day1* for 0,99 ou menos). Dessa forma, podemos utilizá-la para ver quais pessoas já estavam fedendo no primeiro dia do festival!


 **Less than or equal to** (Menor que ou igual a). Essa operação é a mesma que a anterior, mas os casos que são exatamente 1 também seriam incluídos.


 **More than** (Maior que). Essa operação é normalmente utilizada para incluir casos acima de um certo valor. Assim, se você clicar no botão , e digitar “day1 > 1”, o SPSS fará uma análise apenas nos casos em que o escore de higiene no primeiro dia do festival for maior do que 1 (isto é, 1,01 ou acima). Isso pode ser utilizado, caso você queira, para excluir aquelas pessoas que já estavam fedendo quando o festival começou (como estamos interessados nos efeitos do festival na higiene das pessoas, esses casos iriam contaminar os dados – sem mencionar nossas narinas!)


 **More than or equal to** (Maior que ou igual a). Essa operação é a mesma que a anterior, mas irá incluir casos que são exatamente iguais a 1.



 **Equal to** (Igual a). Você pode utilizar essa operação para incluir casos em que os participantes tenham um valor específico. Assim, “day1 = 1” digitado na caixa de diálogo *if* (se) assegura que so-

mente casos que tenham valores iguais a 1 para a variável *day1* serão incluídos. Isso será mais útil quando tivermos uma variável de grupo e quisermos trabalhar com apenas um dos grupos. Por exemplo, se você tiver a variável *música* para agrupar os gêneros musicais dos participantes do festival (veja a Seção 5.10) como 1 = Pagodeiros, 2 = Metaleiros, 3 = *Punks* e 4 = Ecléticos, e digitar “música = 2”, a análise será executada somente com aquelas pessoas que se identificam como “metaleiras” (pessoas que ouvem *heavy metal*).

 **Not equal to** (Não igual a). Essa operação irá incluir todos os casos exceto aqueles com valores específicos. Assim, “música \neq 2” incluirá todos os casos exceto aqueles que se identificaram como metaleiros.

 **Variable type** (Tipo de variável). Esse botão abre uma caixa de diálogo que permite fornecer um nome para uma nova variável e especificar que tipo de variável ela é (por exemplo, numérica).

O resultado de qualquer uma das funções de cálculo irá produzir uma nova coluna no editor de dados, portanto, a primeira coisa a fazer é digitar um nome para essa nova variável (na caixa denominada *Target Variable* [variável alvo]). Se você digitar um nome que já existe, o SPSS irá informar e perguntar se quer alterá-lo ou substituir a variável existente. Se você responder “Yes” (Sim), o SPSS substituirá os dados existentes na coluna com os resultados da função utilizada. Se você responder “No” (Não), nada acontecerá e você precisará renomear a nova variável. Se você clicar em , outra caixa de diálogo aparecerá, em que você pode fornecer à nova variável um rótulo descritivo e especificar se ela é numérica ou texto (veja a Seção 2.4.5). O quadro denominado *Numeric Expression* (Expressão Numérica) é o espaço onde os comandos numéricos são digitados (chamei esse espaço de “área de comandos”). Você poderá entrar com

os nomes das variáveis na área de comandos selecionando a variável requerida da lista de variáveis e clicando em . Da mesma forma, você poderá selecionar certas funções da lista de funções disponíveis e colocá-las na área de comandos clicando em .

Algumas das funções mais úteis estão listadas na Tabela 3.1, que mostra a forma padrão de cada uma, o nome, um exemplo de como ela poderá ser utilizada e o que o SPSS mostraria como saída se a função fosse utilizada no exemplo dado. Existem várias funções básicas para calcular médias, desvios padrão e soma de colunas. Existem também funções, como a raiz quadrada e as logarítmicas, úteis para transformar dados que apresentam assimetria; essas são as funções que iremos utilizar agora. Para o leitor interessado, o guia básico do usuário do sistema SPSS (e os arquivos de ajuda do programa) apresentam detalhes de todas as funções disponíveis na caixa de diálogo *compute* (calcular) (Norušis, 1997, veja também SPSS Inc., 1997)



Agora que verificamos como utilizar as funções básicas do comando *compute* (calcular), vamos utilizá-lo para transformar os nossos dados. Primeiro, abra a caixa de diálogo principal utilizando a sequência de menu: **Transform⇒Compute...** (Transformar⇒Calcular...). Entre com o nome **logday1** (logaritmo do dia 1) no quadro denominado *Target Variable* (variável alvo) e clique em  e forneça à variável um nome que a descreva melhor, como *Logaritmo dos escores de higiene do primeiro dia do Festival de Glastonbury*. Depois, role para baixo a lista de funções até encontrar uma denominada LG10(mumexpr) (Logaritmo decimal[expressão numérica]). Destaque essa função com o mouse e a transfira para a área de comandos clicando em . Quando o comando é transferido, ele aparecerá na área de comando como “LG10(?)” e o ponto de interrogação deve ser substituído pelo nome da variável (que pode ser digitado manualmente ou transferido da lista de variáveis). Assim, substitua o ? com

Tabela 3.1 Algumas funções de cálculo (*compute*) úteis

Função	Nome	Exemplo (Entrada)	Saída
MEAN(?, ?, ..)	Média	Mean(day1, day2, day3)	Para cada linha, o SPSS calcula o valor médio dos escores de higiene ao longo dos três dias do festival
SD(?, ?, ..)	Desvio Padrão	SD(day1, day2, day3)	Para cada linha, o SPSS calcula o desvio padrão dos valores nas colunas rotuladas como day1, day2 e day3.
SUM(?, ?, ..)	Soma	Sum(day1, day2)	Para cada linha, o SPSS adiciona o valor das linhas das colunas denominadas day1 e day2.
SQRT(?)	Raiz Quadrada	SQRT(day2)	Produz uma nova coluna que contém a raiz quadrada de cada valor da coluna denominada day2. Útil para transformar dados assimétricos ou dados com variâncias heterogêneas.
ABS(?)	Valor Absoluto	ABS(day1)	Produz uma variável que contém o valor absoluto dos valores da coluna denominada day1 (valores absolutos são aqueles em que todos os resultados são positivos. Assim, 5 torna-se 5).
LG10(?)	Logaritmo de base 10	LG10(day1)	Produz uma variável que contém os logaritmos (Base 10) dos valores do primeiro dia. É útil para transformar dados com assimetria positiva.
NORMAL(stddev)	Números Aleatórios Normais	Normal(5)	Produz uma variável contendo números pseudo-aleatórios de uma distribuição normal com média 0 e um desvio padrão de 5.

o nome da variável digitando **day1**. Agora, note que nos escores do primeiro dia existe um valor igual a 0. Para resolver esse problema, devemos adicionar uma constante aos escores originais antes de calcular os logaritmos. Qualquer constante poderá ser utilizada, desde que ela faça com que todos os escores sejam maiores do que zero. Nesse caso, o nosso menor escore é zero, então iremos simplesmente adicionar o valor 1 a todos os valores e isso irá assegurar que todos sejam maiores do que zero. Para fazer isso, assegure-se de que o cursor esteja dentro dos parênteses e clique em **+** e depois em **1**. A caixa de diálogo final deve ser semelhante à da Figura 3.8. Note que a expressão que se lê $\text{LG } 10(\text{day1} + 1)$ fará com que o SPSS adicione Day1 a todos os escores do primeiro dia e depois obtenha o logaritmo do valor resultante. Clique em **OK** para criar a nova variável **log-day1** contendo os valores transformados. Agora repita o processo para criar de maneira semelhante às variáveis **logday2** e **logday3** para os dados dos dias 2 e 3!

Agora que sabemos como *log-transformar* dados, podemos experimentar outros tipos de

transformação. Por exemplo, se quisermos utilizar a transformação raiz quadrada, podemos seguir o mesmo processo, utilizando um nome para a variável resultante como **sqrtday1** e selecionando a função **SQRT(numexpr)** da lista. Ela aparecerá no quadro denominado *numeric expression* (expressão numérica) como **SQRT(?)** e você pode simplesmente substituir o ponto de interrogação com a variável que quiser alterar – nesse caso, **day1** (dia1). A expressão final será **SQRT(day1)**. Novamente, tente repetir isso para as variáveis **day2** e **day3** a fim de criar novas variáveis denominadas **sqrtday2** e **sqrtday3**.

Finalmente, se você quiser utilizar a transformação recíproca nos dados do dia 1, utilize um nome como **recday1** e simplesmente clique em **1** e depois em **1**. Normalmente, você selecionará no nome da variável que quiser transformar da lista e irá transferi-la clicando em **→** (ou então digitando o nome da variável). No entanto, o dia dois contém um valor zero e, se tentarmos dividir por zero, obteremos uma mensagem de erro (*you can't divide by 0* – você não pode dividir por 0). Nesse caso, precisa-

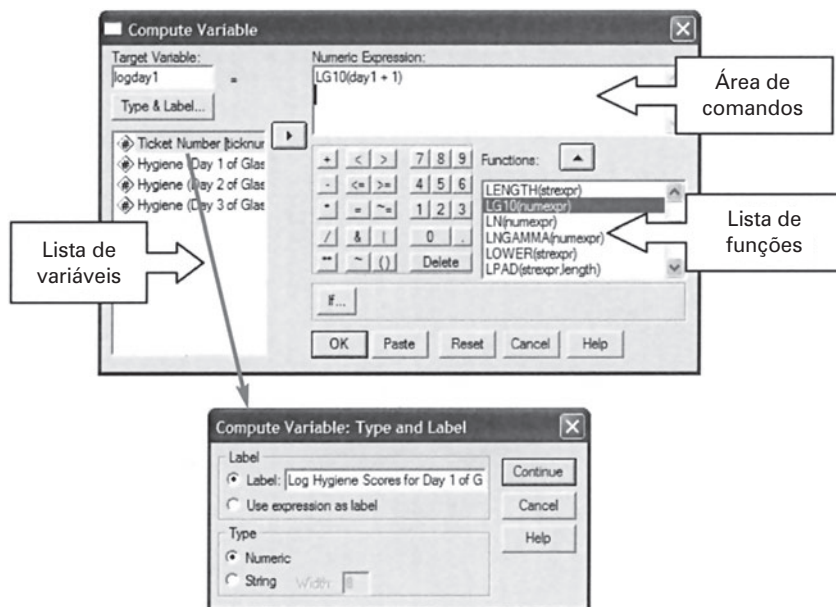


Figura 3.8 Caixa de diálogo para a função *compute* (calcular).

mos adicionar uma constante à nossa variável da mesma forma que foi feito com a transformação logarítmica. Qualquer constante pode ser utilizada, mas 1 é um valor conveniente para esses dados. Assim, em vez de selecionar a variável que queremos transformar, clique em . Isso acrescentará parênteses ao quadro denominado *numeric expression* (expressão numérica), assim, assegure-se de que o cursor esteja entre os parênteses e selecione a variável que você quer transformar da lista e transfira para o painel utilizando a seta (ou digite o nome da variável manualmente). Agora, clique em e então em (ou digite + 1 utilizando o teclado). O quadro denominado *Numeric*

Expression (expressão numérica) deve agora conter o seguinte texto $1/(\text{day1} + 1)$. Clique em para criar a nova variável contendo os valores transformados. Repita esse processo para os dados dos dois dias restantes. Se você estiver calculando muitas variáveis, poderá ser mais rápido utilizar a janela de sintaxe como explicado no Quadro 3.3.

Vamos dar uma olhada nas nossas novas distribuições (Figura 3.9). Compare-as com as distribuições originais (não transformadas) na Figura 3.5. Agora, você pode ver que todas as três transformações limpam os dados dos escores da higiene para o dia 2: a assimetria positiva desapareceu (a transformação pela raiz

Quadro 3.3

Utilizando sintaxe para calcular novas variáveis ③



Se você estiver calculando muitas variáveis, pode ser mais rápido utilizar a sintaxe – por exemplo, para criar os dados transformados no exemplo deste capítulo. No site www.artmed.com.br, há um arquivo de sintaxe, **Transformations.sps**, que escrevi para todas as nove transformações discutidas. Carregue esse arquivo e rode a sintaxe ou abra a janela de sintaxe (veja a Seção 2.6) e digite o seguinte:

```
COMPUTE logday1 = LG10(day1 + 1).
COMPUTE logday2 = LG10(day2 + 1).
COMPUTE logday3 = LG10(day3 + 1).
COMPUTE sqrtday1 = SQRT(day1).
COMPUTE sqrtday2 = SQRT(day2).
COMPUTE sqrtday3 = SQRT(day3).
COMPUTE recday1 = 1/(day1 + 1).
COMPUTE recday2 = 1/(day2 + 1).
COMPUTE recday3 = 1/(day3 + 1).
EXECUTE
```

Cada comando COMPUTE (Calcular) faz o mesmo que você faria utilizando a caixa de diálogo na Figura 3.8. Assim, as três primeiras linhas simplesmente solicitam que o SPSS crie três novas variáveis (**logday1**, **logday2** e **logday3**), que são as transformações logarítmicas das variáveis day1 (dia 1), day2 (dia 2) e day3 (dia 3) mais 1. As próximas três linhas fazem a mesma coisa só que agora utilizando a função SQRT e, assim, calculam a raiz quadrada de **day1**, **day2** e **day3** criando as novas variáveis **sqrtday1**, **sqrtday2** e **sqrtday3**, respectivamente. As próximas três linhas executam a transformação recíproca da mesma forma que as outras 2 funções fizeram. A linha final apresenta o comando EXECUTE (Executar) sem o qual nenhum dos anteriores será processado! Note ainda que cada uma das linhas termina com um ponto final.



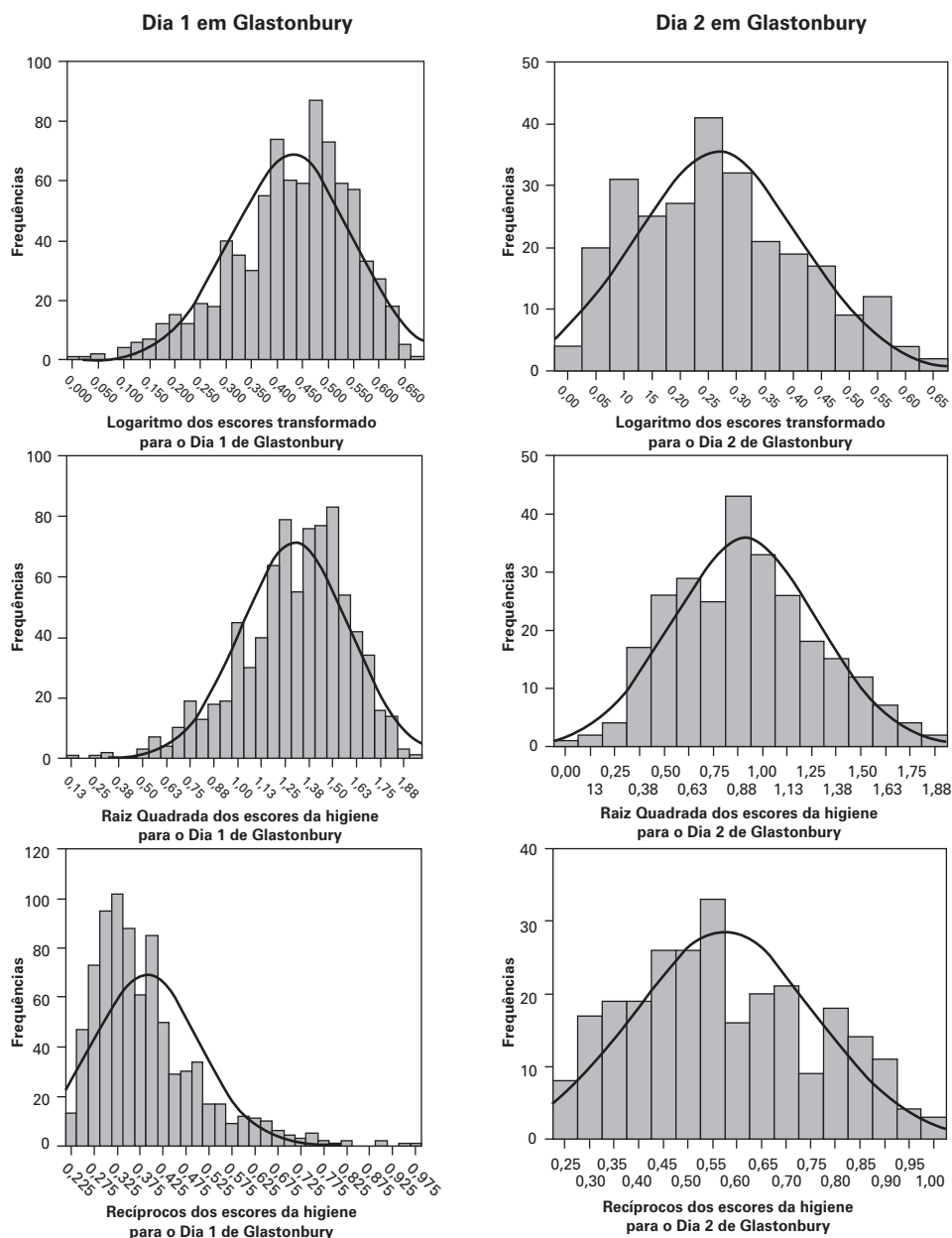


Figura 3.9 Distribuição dos dados da higiene do dia 1 e do dia 2 após as transformações.

quadrada, em particular, foi útil). No entanto, em virtude dos nossos escores da higiene no dia 1 não terem assimetria significativa, eles

agora apresentam uma leve assimetria negativa para a transformação logarítmica e da raiz quadrada e assimetria positiva para a transfor-

mação recíproca!⁵ Se estivermos utilizando apenas os escores do segundo dia, podemos utilizar os dados transformados; no entanto se quisermos verificar os dois dias em conjunto, teremos que ponderar se os benefícios da transformação sobre os dados do segundo dia não são sobrepujados pelos problemas criados com os dados do primeiro dia – a análise de dados pode ser frustrante às vezes!

Se a transformação corrige os problemas, execute qualquer análise estatística que você precise sobre os escores transformados. Se ela não corrigir o problema, você pode considerar a utilização de um teste que não tenha por base a hipótese de dados normalmente distribuídos e, à medida que você percorrer os vários capítulos do livro apresentarei esses testes.⁶

3.4 EXPLORANDO GRUPOS DE DADOS ①

Algumas vezes temos dados em que existem diferentes grupos de pessoas (homens e mulheres, universidades diferentes, pessoas com e sem depressão, por exemplo). Existem várias maneiras de produzir estatísticas básicas para grupos separados de pessoas (e veremos esses métodos na próxima seção). No entanto, pretendo utilizar essa oportunidade para lhe apresentar uma função denominada *split file* (dividir arquivo), que permite repetir qualquer análise sobre vários grupos de casos. A função *split file* (dividir arquivo) permite especificar uma variável de grupo (lembre que essas variáveis são utilizadas para especificar categorias de pessoas). Qualquer procedimento posterior no SPSS irá ser executado em cada uma das categorias especificadas pela variável de grupo.

Você provavelmente está cansado dos dados sobre a higiene do festival de Glastonbury, portanto, vamos utilizar o arquivo de dados **SPSSExam.sav**. Esse arquivo contém dados do desempenho de estudantes em um

exame sobre o SPSS. Quatro variáveis foram mensuradas: **exam** (exame) (escores dos exames sobre o SPSS de alunos do primeiro ano em percentagem), **computer** (computador) (o conhecimento computacional mensurado em percentual), **lecture** (aulas) (percentual de aulas sobre o SPSS frequentadas) e **numeracy** (numerácia) (uma medida da habilidade numérica em uma escala de 1 a 15). Existe ainda uma variável denominada **uni** indicando se o estudante frequenta a Universidade de Sussex ou a de Duncetown*. Para iniciar, abra o arquivo **SPSSExam.sav** (veja a Seção 2.8 se não lembrar como abrir um arquivo). Vamos começar olhando os dados como um todo.



3.4.1 Executando a análise para todos os dados ①

Para olhar a distribuição das variáveis, podemos utilizar o comando *frequencies* (frequências), apresentado na seção anterior (veja a Figura 3.6). Utilize essa caixa de diálogo e coloque todas as quatro variáveis (**exam**, **computer**, **lecture** e **numeracy**) no quadro *variable(s)* (variável [is]). Clique em **Statistics...** para selecionar a caixa de diálogo *statistics* (estatísticas) e selecione algumas medidas de tendência central (média, moda e mediana), medidas de variabilidade (amplitude, desvio padrão, variância e quartis) e medidas de forma (assimetria e curtose). Clique também em **Charts** para acessar a caixa de diálogo dos diagramas e selecione a distribuição de frequências com a curva normal (veja a Figura 3.6 se precisar

⁵ A inversão da assimetria que aconteceu com a transformação recíproca ocorre, como mencionada antes, porque essa transformação inverte o valor dos escores.

⁶ Por conveniência, muitos livros-texto se referem a esses testes como *não-paramétricos* ou *distribuições livres* e os colocam em um capítulo separado. De fato, nenhum desses termos é particularmente preciso (pois nenhum desses testes está livre de suposições), mas, para manter a tradição, eu os coloquei em um capítulo (13) próprio, no ostracismo e solitários.

* N. de T.: É um nome fictício, atribuído pelo autor, que em português significa cidade dos burros.

de ajuda com essa opção). Retorne à caixa de diálogo principal e clique em **Continuar** e, uma vez nela, clique em **OK** para executar a análise.

3.4.2 Saída do SPSS para todos os dados ①

A saída 3.2 do SPSS mostra a tabela das estatísticas descritivas para as quatro variáveis deste exemplo. A partir dessa tabela, é possível ver que, em média, os estudantes frequentaram aproximadamente 60% das aulas, eles obtiveram 58% no exame do SPSS, tiveram somente 51% no teste de conhecimento ou habilidade computacional e apenas 5 em uma escala até 15 em numerácia, isto é, habilidade numérica. Além disso, o desvio padrão para a habilidade computacional foi relativamente pequeno comparado ao do percentual de aulas frequentadas e as notas do exame. Essas últimas duas variáveis apresentaram várias modas (multimodais). As outras medidas importantes são a assimetria e a curtose, ambas apresentando um erro padrão associado. Já discuti-

mos essas medidas anteriormente e verificamos que podemos convertê-las em escores-*z* dividindo-as pelos seus erros padrões. Para os escores do exame, o escore-*z* da assimetria é $-0,107/0,241 = -0,44$. Para a numerácia, o escore-*z* da assimetria foi de $0,961/0,241 = 3,99$. Fica evidente que os escores da numerácia apresentam assimetria positiva, indicando uma concentração de valores na parte esquerda da distribuição (isto é, muitos estudantes pontuaram mal). Tente calcular o escore-*z* para as demais variáveis.

A saída fornece tabelas das distribuições de frequências de cada uma das variáveis (que não são apresentadas aqui). Essas tabelas listam cada escore e o número de vezes que ele ocorreu no conjunto de dados. Adicionalmente, cada valor da frequência é expresso como um percentual da amostra (nesse caso, as frequências e as percentagens são as mesmas, pois o tamanho da amostra é igual a 100). Também, a percentagem acumulada é dada, que nos informa quantos casos estão abaixo

Saída 3.2 do SPSS

Statistics (Estatísticas)

		Computer literacy (Conhecimento Computacional)	Percentage on SPSS Exam (Percentual no Exame do SPSS)	Percentage of lectures attended (Percentual de comparecimento às aulas)	Numeracy (Numerácia)
<i>N</i>	<i>Valid</i> (Válidos)	100	100	100	100
	<i>Missing</i> (Desconhecidos)	0	0	0	0
<i>Mean</i> (Média)		50.7100	58.1000	59.7650	4.8500
<i>Std. Error of Mean</i> (Erro Padrão da Média)		0.8260	2.1316	2.1685	0.2706
<i>Median</i> (Mediana)		51.5000	60.0000	62.0000	4.0000
<i>Mode</i> (Moda)		54.00	72.00 ^a	48.50 ^a	4.00
<i>Standard Deviation</i> (Desvio Padrão)		8.2600	21.3156	21.6848	2.7057
<i>Variance</i> (Variância)		68.2282	454.3535	470.2296	7.3207
<i>Skewness</i> (Assimetria)		-0.174	-0.107	-0.422	0.961
<i>Std. Error of Skewness</i> (Erro Padrão da Assimetria)		0.241	0.241	0.241	0.241
<i>Kurtosis</i> (Curtose)		0.364	-1.105	-0.179	0.946
<i>Std. Error of Kurtosis</i> (Erro Padrão da Curtose)		0.478	0.478	0.478	0.478
<i>Range</i> (Amplitude)		46.00	84.00	92.00	13.00
<i>Minimum</i> (Mínimo)		27.00	15.00	8.00	1.00
<i>Maximum</i> (Máximo)		73.00	99.00	100.00	14.00

a. Existem múltiplas modas. O menor valor é mostrado.

de certo valor. Assim, por exemplo, podemos ver que 66% dos escores da numerácia foram 5 ou menos, 74% foram 6 ou menos, e assim por diante. Olhando em outra direção, podemos perceber que somente 8% (100% – 92%) tiveram um escore maior do que 8.

Finalmente, foram obtidos os histogramas de cada uma das variáveis sobrepostos por uma distribuição normal. Esses gráficos são apresentados na Figura 3.10 e nos mostram várias coisas. Primeiro, parece que o conhecimento computacional tem distribuição aproximadamente normal (poucas pessoas são muito boas e poucos são muitos ruins com computadores, a maioria tem um grau de conhecimento similar). Os escores do exame são bem interessantes porque essa distribuição é claramente não-normal; de fato, ela parece suspeitamente

bimodal (existem dois picos indicando duas modas). Essa observação corresponde a uma anterior da tabela das estatísticas descritivas. O comparecimento às aulas é aparentemente normal, mas as caudas da distribuição são bastante pesadas (isto é, embora muitas pessoas compareçam à maioria das aulas, existe um número de almas dedicadas que comparecem a todas e um grupo maior do que o “normal” que comparecem a poucas). Esse é o motivo dela apresentar altas frequências nas duas caudas. Finalmente, o teste de numerácia produziu dados com uma assimetria positiva acentuada (isto é, a maioria das pessoas foi muito mal nesse teste e poucos foram bem). Isso corresponde ao que a estatística de assimetria já havia indicado.

Estatísticas descritivas e histogramas são boas formas de obter um retrato instantâneo

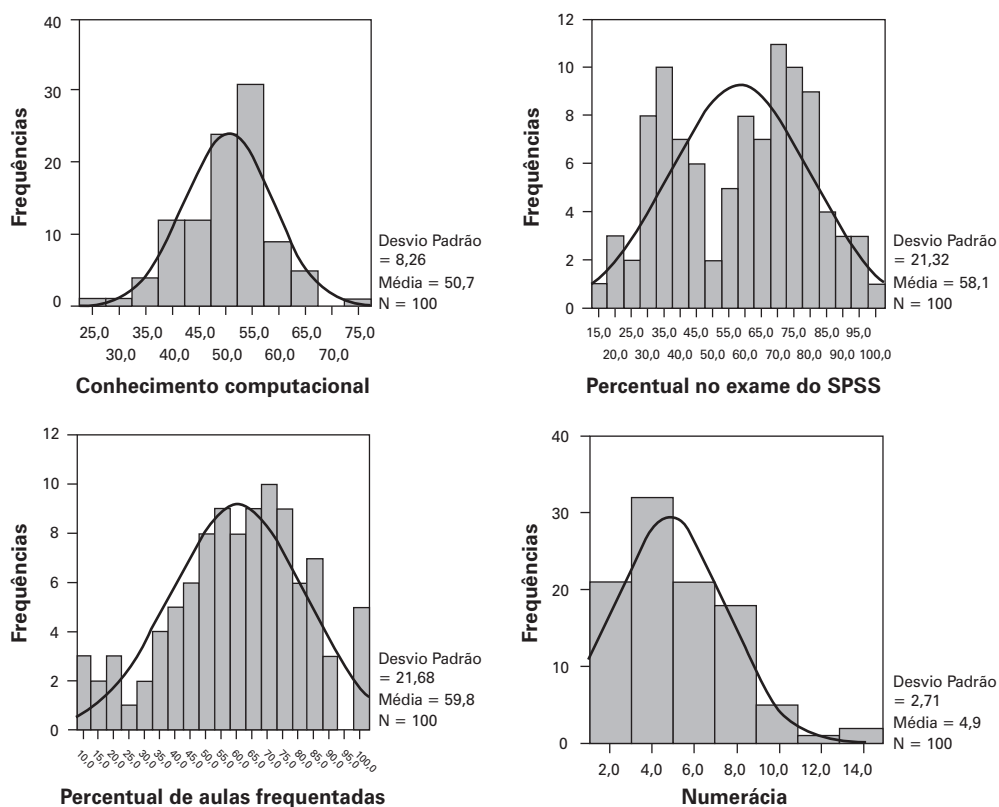




Figura 3.10 Histograma dos dados do exame sobre o SPSS.

da distribuição dos dados. Essa compreensão poderá ser bastante útil: por exemplo, a distribuição bimodal nos escores do exame do SPSS indica que os estudantes são tipicamente muito bons em estatística ou têm dificuldades com ela (existem relativamente poucos que estão entre esses dois extremos). Intuitivamente, esse achado se ajusta à natureza do conteúdo: estatística é muito fácil uma vez que tudo esteja no devido lugar, mas antes que tal encantamento ocorra, ela parece ser desesperadamente difícil! Embora existam muitas informações que podemos obter a partir dos histogramas e informações descritivas sobre a distribuição, existem maneiras mais objetivas na qual podemos verificar o grau de normalidade em um conjunto de dados (veja a Seção 3.5).

3.4.3 Executando a análise para grupos diferentes ①

Se você quiser obter estatísticas separadas para cada uma das universidades, podemos dividir o arquivo e seguir utilizando o comando *frequencies* (frequências) descrito na seção anterior. Para dividir o arquivo, basta utilizar o seguinte caminho de menu: **Data⇒Split File...** (Dados⇒Dividir Arquivo...) ou clicar em . Na caixa de diálogo resultante (Figura 3.11), selecione a opção *Organize output by groups* (Organize a saída por grupos). O quadro *Groups Based on* (Grupos baseados em) será ativado. Selecione a variável contendo

o código dos grupos para o qual você deseja repetir a análise (nesse exemplo, selecione **Uni**) e a transfira para o quadro clicando em . Por padrão, o SPSS irá ordenar o arquivo por esses grupos (isto é, ele listará uma categoria seguida pela outra na janela do editor de dados). Uma vez que você tenha dividido o arquivo, utilize o comando *frequencies* (frequências) (veja a seção anterior). Solicitei estatísticas apenas para as variáveis **numeracy** e **exam**. Você poderá selecionar as outras variáveis se desejar.

3.4.4 Saídas para grupos diferentes ①

A saída do SPSS é dividida em duas seções: primeiro os resultados dos estudantes da universidade de Duncetown e depois os da universidade de Sussex. A saída do SPSS 3.3 mostra as duas principais tabelas. A partir dessas tabelas, fica claro que os alunos da universidade de Sussex tiveram escores maiores tanto no exame do SPSS quanto no teste de numerácia do que os alunos de Duncetown. De fato, olhando para as médias, percebe-se que os estudantes de Sussex tiveram um rendimento 6% maior no exame do SPSS e tiveram escores no teste de numerácia duas vezes maior. Os desvios padrões para as duas variáveis são levemente maiores para a universidade de Sussex, mas não muito.

A Figura 3.12 mostra os histogramas dessas variáveis separados pelas universidades frequentadas. A primeira coisa interes-

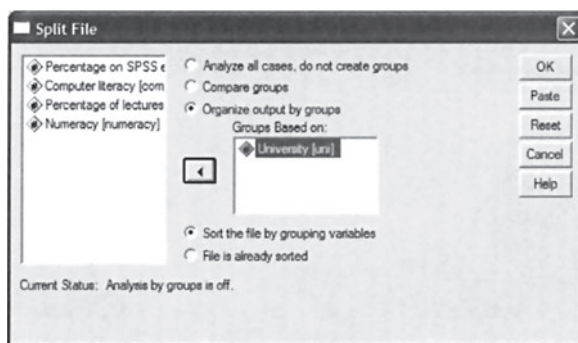


Figura 3.11 Caixa de diálogo para o comando *split file* (dividir arquivo).

Saída 3.3 do SPSS

Estatísticas da Universidade de Duncetown^b

	Percentage on SPSS Exam (Percentual no Exame do SPSS)	Numeracy (Numerácia)
N Valid (Válidos) Missing (Desconhecidos)	50 0	50 0
Mean (Média)	40.1800	4.1200
Std. Error of Mean (Erro Padrão da Média)	1.78032	0.29227
Median (Mediana)	38.0000	4.0000
Mode (Moda)	34.00 ^a	4.00
Standard Deviation (Des- vio Padrão)	12.58877	2.06664
Variance (Variância)	158.47714	4.27102
Skewness (Assimetria)	0.309	0.512
Std. Error of Skewness (Erro Padrão da Assime- tria)	0.337	0.337
Kurtosis (Curtose)	-0.567	-0.484
Std. Error of Kurtosis (Erro Padrão da Curtose)	0.662	0.662
Range (Amplitude)	51.00	8.00
Minimum (Mínimo)	15.00	1.00
Maximum (Máximo)	66.00	9.00

a. Multiple modes exist. The smallest value is shown (Ex-
istem múltiplas modas. É mostrado o menor valor.)
b. University Duncetown University

Estatísticas da Universidade de Sussex^b

	Percentage on SPSS Exam (Percentual no Exame do SPSS)	Numeracy (Numerácia)
N Valid (Válidos) Missing (Desconhecidos)	50 0	50 0
Mean (Média)	76.0200	5.5800
Std. Error of Mean (Erro Padrão da Média)	1.44321	0.43433
Median (Mediana)	75.0000	5.0000
Mode (Moda)	72.00 ^a	5.00
Standard Deviation (Des- vio Padrão)	10.20502	3.07120
Variance (Variância)	104.14245	9.43224
Skewness (Assimetria)	0.272	0.793
Std. Error of Skewness (Erro Padrão da Assime- tria)	0.337	0.337
Kurtosis (Curtose)	-0.264	0.260
Std. Error of Kurtosis (Erro Padrão da Curtose)	0.662	0.662
Range (Amplitude)	43.00	13.00
Minimum (Mínimo)	56.00	1.00
Maximum (Máximo)	99.00	14.00

a. Multiple modes exist. The smallest value is shown (Ex-
istem múltiplas modas. É mostrado o menor valor.)
b. University Sussex University

sante a ser notada é que para as notas dos exames, as distribuições são praticamente normais. Isso parece estranho porque a distribuição é bimodal. No entanto, isso começa a fazer sentido quando você considera que para Duncetown a distribuição está centrada em torno do valor de 35%, mas para Sussex a distribuição está centrada em torno do valor 75%. Isso mostra como é importante verificar a distribuição dentro dos grupos. Se estamos interessados em comparar Duncetown e Sussex, não importa que a distribuição global dos valores seja bimodal; tudo o que importa é que cada grupo seja originário de uma distribuição normal e, nesse caso, isso aparenta ser verdadeiro. Quando as duas amostras são combinadas, essas duas distribuições normais criam uma distribuição bimodal (com uma das modas estando em torno do centro da distribuição de Duncetown e a outra

em torno do centro dos dados de Sussex!). Para os escores da numerácia, a distribuição apresenta uma assimetria positiva leve no grupo de Duncetown (existe uma grande concentração nos escores mais baixos); os escores de Sussex apresentam também uma leve assimetria positiva, mas os escores são em geral mais altos do que os de Duncetown. Dessa forma, a assimetria positiva observada para todos os dados anteriormente é uma contribuição dos escores dos alunos das duas universidades. Quando você tiver terminado com o comando *split* (dividir), lembre-se de desligá-lo ou voltar atrás (senão o SPSS irá fazer todas as demais análises em cada grupo separadamente). Para voltar atrás com esse comando, retorne à caixa de diálogo *splitfile* (dividir arquivo) e selecione *Analyze all cases, do not create groups* (Análise todos os casos, não crie grupos).

Notas do exame do SPSS

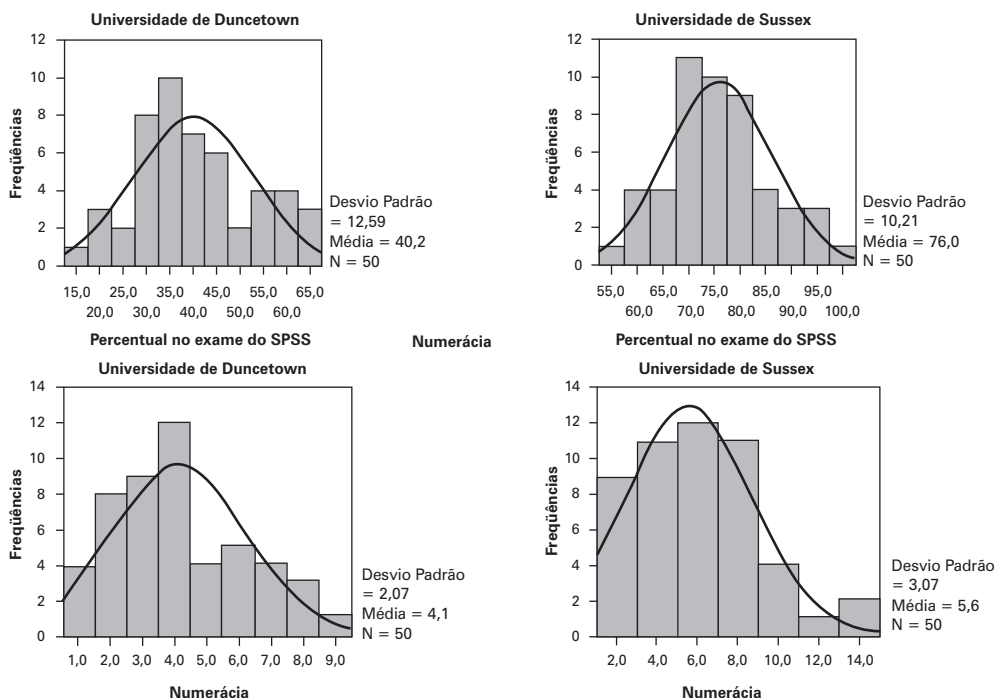


Figura 3.12 Distribuições dos escores no exame e na numerácia para os estudantes das universidades de Duncetown e Sussex.

3.5 TESTANDO SE UMA DISTRIBUIÇÃO É NORMAL ①

Alguém disse Smirnov? Legal! Eu preciso de uma bebida mesmo depois de toda essa análise de dados!



Olhar para os histogramas é uma opção, mas eles nada informam sobre a distribuição estar próxima o suficiente da normalidade. A análise do histograma é subjetiva e aberta a abusos (Pos-

so imaginar pesquisadores sentados olhando para uma distribuição completamente distorcida e dizendo “bem Carlos, ela parece normal para mim” e Carlos replicando “com certeza”). O que é necessário é um teste objetivo para decidir se uma distribuição é ou não-normal. As estatísticas de assimetria e curtose que vimos antes

informam um pouco sobre desvios da normalidade, mas cada uma delas lida com apenas um aspecto da não-normalidade. Outra maneira de tratar o problema é ver se a distribuição como um todo se desvia de uma distribuição normal modelo. Os testes de Kolmogorov-Smirnov (Figura 3.13) e de Shapiro-Wilk fazem justamente isso: eles comparam escores de uma amostra a uma distribuição normal modelo de mesma média e variância dos valores encontrados na amostra. Se o teste é não-significativo ($p > 0,05$), ele nos informa que os dados da amostra não diferem significativamente de uma distribuição normal (isto é, eles podem ser normais). Por outro lado, se o teste é significativo ($p < 0,05$), a distribuição em questão é significativamente diferente de uma distribuição normal (isto é, ela é não-normal). Esses testes parecem ótimos: em um procedimento eles nos dizem se os nossos escores são normalmente distribuídos (beleza!).



Figura 3.13 Andrei Kolmogorov, querendo ter uma Smirnov.

Entretanto, eles apresentam limitações porque com amostras grandes é muito fácil obter valores significativos a partir de pequenos desvios da normalidade e, assim, um resultado significativo não necessariamente nos informa se o desvio da normalidade é suficiente para prejudicar os procedimentos estatísticos que serão aplicados aos dados. Acredito que o sensato a fazer é: de qualquer forma utilize esses testes, mas faça diagramas dos dados bem como tente obter uma decisão sobre a extensão da não-normalidade.

3.5.1 Executando o teste de Kolmogorov-Smirnov no SPSS ①

O teste de Kolmogorov-Smirnov (K-S de agora em diante) pode ser acessado pelo comando *explore* (explorar) **Analyze⇒Descriptive Statistics⇒Explore...** (Analisar⇒Estatística Descritiva⇒Explorar...),⁷ A Figura 3.14 mostra a caixa de diálogo para o comando *explore* (explorar). Primeiro, entre com qualquer variável de interesse no quadro denominado *Dependent list* (Lista Dependente) selecionando-as no lado esquerdo e transferindo para o direito clicando em . Para esse exemplo, apenas

selecione os valores das variáveis exame e numerácia. É também possível selecionar um fator (variável de grupo) pelo qual podemos dividir a saída – (assim, se você selecionar **uni** e transferi-la para o quadro denominado *Factor List* [Lista de fatores], o SPSS irá produzir uma análise exploratória para cada um dos grupos – parecido com o comando *split file* [dividir arquivo]. Se você clicar em **Statistics...**, uma caixa de diálogo aparece, mas a opção por omissão é suficiente (produzirá médias, desvios padrão e assim por diante). A opção mais interessante para o nosso objetivo é acessada clicando em **Plots...**, que abrirá uma nova caixa de diálogo onde você pode selecionar a opção ☒ **Normality plots with tests**; isso irá produzir tanto um teste K-S quanto um gráfico denominado *normal Q-Q plots* (diagramas Q-Q normais)* para todas as variáveis selecionadas. Por omissão, o SPSS também irá apresentar diagramas de caixa e bigodes. Separado por grupos se um fator foi especificado) e também um diagrama de caule e folhas. Clique em **Continue** para retornar à caixa de diálogo principal e clique em **OK** para executar a análise.

3.5.2 Saída do procedimento *Explore* (Explorar) ①

A primeira tabela produzida pelo SPSS contém estatísticas descritivas (médias, etc.) e deve ter os mesmos valores que a tabela obtida com a utilização do procedimento frequências (*frequencies*). A tabela importante é a do teste K-S.

Essa tabela inclui o próprio teste, os graus de liberdade (que devem ser iguais ao tamanho da amostra) e os valores da significância do teste. Lembre que um valor significativo (Sig. < 0,05) indica um desvio da normalidade. Tanto para os valores do exame quanto para os da numerácia o teste K-S é altamente

⁷ Esse caminho seria **Statistics⇒Summarize⇒Explore...** (Estatísticas⇒Resumir⇒Explorar...) na versão 8.0 e anteriores.

* N. de T.: Um diagrama Q-Q é um gráfico onde o quantil do primeiro conjunto de dados é apresentado contra o quantil de um segundo conjunto de dados. Por quantil entendemos um determinado percentual de valores abaixo de um dado valor. Alguns quantis especiais são: quartis, decis e percentis.

Tests of Normality (Testes de normalidade)

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic (Estatística)	Df (gl)	Sig. (Significância)	Statistic (Estatística)	Df (gl)	Sig. (Significância)
Percentage on SPSS exam (Percentual no exame do SPSS)	0.102	100	0.012	0.961	100	0.005
Numeracy (Numerácia)	0.153	100	0.000	0.924	100	0.000

a. Correção de significância de Lilliefors.

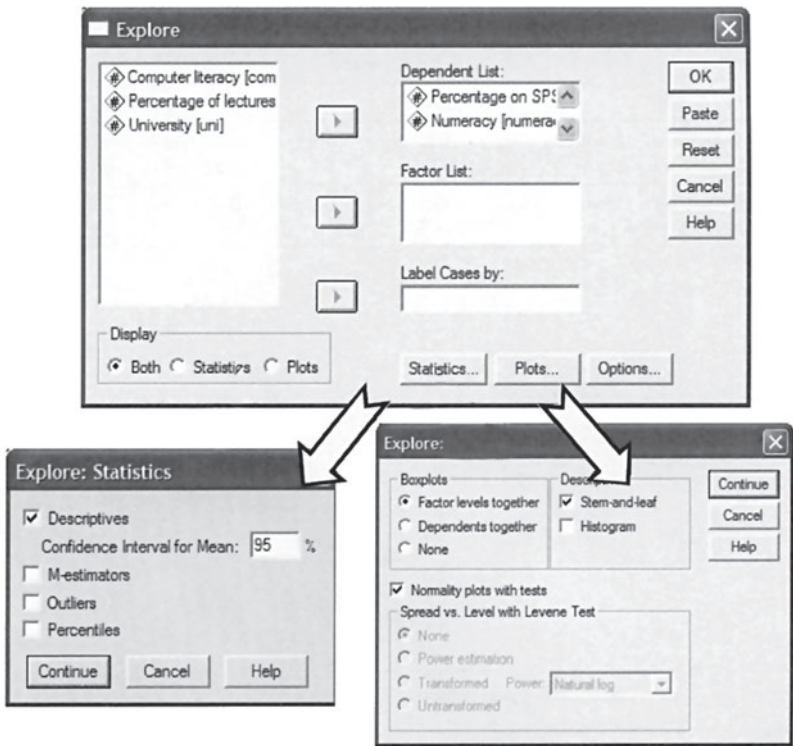


Figura 3.14 Caixas de diálogo para o comando *explore* (explorar).

significativo, indicando que as duas distribuições se desviam da normalidade. Esse resultado provavelmente reflete a distribuição bimodal encontrada para os escores do exame e a distribuição com assimetria positiva observada para os escores da numerácia. Entretanto, esses testes confirmam que esses desvios eram significativos. Essa confirmação é importante porque os histogramas nos informam

apenas que nossas distribuições amostrais se desviam da normalidade; eles não nos dizem se esse desvio é importante. A estatística teste para o teste K-S é representada por D, assim, podemos relatar esses resultados da seguinte forma. A percentagem no exame do SPSS, $D(100) = 0,10$, $p < 0,05$, e para os escores da numerácia, $D(100) = 0,15$, $p < 0,001$, foram ambos significativamente não-normais. Os

números nos parênteses são os graus de liberdade (gl) da tabela.

Concluindo, tenha em mente que quando olhamos os escores dos exames considerando grupos separadamente, as distribuições parecem praticamente normais; mas se executarmos testes separados para as duas universidades (colocando **uni** na caixa denominada *Factor List* [Lista de Fatores], como na Figura 3.17) o teste K-S poderá ser não-significativo. De fato, se você fizer isso, obterá a saída 3.4 do SPSS na forma de tabela, que mostra que os percentuais do exame sobre o SPSS são, de fato, normais entre os dois grupos* (os valo-

res na coluna denominada *Sig.* são maiores do que 0,05). Isso é importante porque se nossa análise envolve a comparação entre dois grupos, o que importa não é a distribuição global, mas a distribuição de cada grupo.

O SPSS também produz um diagrama Q-Q normal para qualquer variável especificada (veja a Figura 3.15). O diagrama Q-Q normal traça os valores que esperaríamos obter se a distribuição fosse normal (valores esperados) contra os valores realmente vistos nos dados (valores observados). Os valores esperados são uma linha reta na diagonal do primeiro quadrante do sistema de coordenadas, onde os valores observados (os pontos no diagrama) devem cair exatamente ou bem próximos da linha (significando que os valores observados são os mesmos que se esperaria obter de um conjunto normalmente distribuído). Qualquer

* N. de T.: Como a hipótese nula não está sendo rejeitada aqui, eu colocaria que “as distribuições são possivelmente normais” em vez de “de fato normais”, como o autor coloca.

Saída 3.4 do SPSS

Tests of Normality (Testes de normalidade)

University (Universidade)		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic (Estatística)	Df (gl)	Sig. (Signf.)	Statistic (Estatística)	Df (gl)	Sig. (Signf.)
Percentage on SPSS exam (Percentual no exame do SPSS)	Duncetown	0.106	50	0.200*	0.972	50	0.238
	Sussex	0.073	50	0.200*	0.984	50	0.715
Numeracy (Numerácia)	Duncetown	0.183	50	0.000	0.941	50	0.015
		0.155	50	0.004	0.932	50	0.007

* Esse é um limite inferior da verdadeira significância.
a. Correção de significância de Lilliefors.

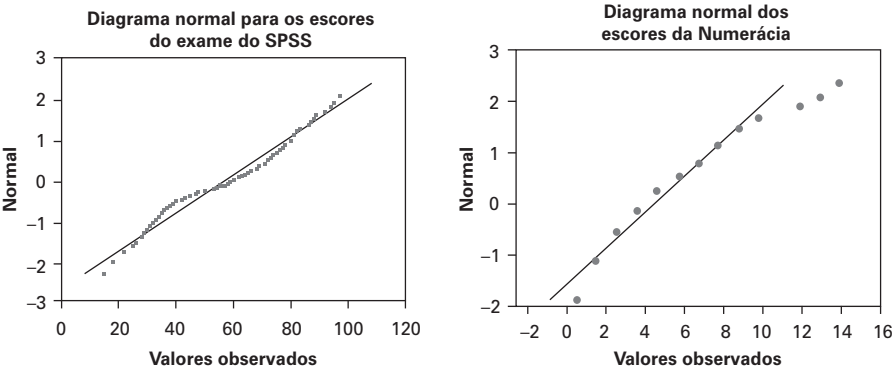


Figura 3.15 Diagramas Q-Q normais das variáveis escores no Exame do SPSS e na numerácia.

desvio dos pontos em relação à linha reta diagonal representa um desvio da normalidade. Assim, se o diagrama Q-Q se parece com uma linha reta com uma longa cobra serpenteando ao redor, você tem alguns desvios da normalidade! Especificamente, quando a linha fica de forma consistente abaixo da diagonal ou acima, isso está mostrando que a curtose é diferente de uma distribuição normal e, quando os pontos apresentarem uma forma de S, o problema será com a assimetria.

Em ambas as variáveis analisadas já sabemos que os dados são não-normais e esses diagramas confirmam essa observação porque os pontos se desviam substancialmente da linha. Vale à pena notar que os maiores desvios ocorrem com a variável *numeracia* e isso está de acordo com o valor altamente significativo da variável no teste K-S. Tal desvio da normalidade nos informa que não podemos utilizar um teste paramétrico porque a hipótese de normalidade não se verifica. Nessas circunstâncias, podemos em alguns casos considerar a utilização de um teste não-paramétrico como forma de verificar a hipótese de interesse (veja o Capítulo 4).

3.6 TESTANDO A HOMOGENEIDADE DA VARIÂNCIA ①

Até agora, me concentrei na hipótese de dados normalmente distribuídos; entretanto, no início do Capítulo 1 mencionei outra suposição, a homogeneidade da variância. Essa hipótese significa que à medida que você avança entre os níveis de uma variável, a variância da outra não deve mudar. Se você coletou grupos de dados, isso significa que a variância da sua variável ou variável de saída deve ser a mesma em cada um desses grupos. Se você coletou dados contínuos (como os delineamentos intercorrelacionais), essa suposição significa que a variância de uma variável deve ser estável em todos os níveis da outra variável. Vamos ilustrar com um exemplo. Uma otologista está interessada no efeito dos altos sons dos concertos sobre a audição dos frequentadores. Assim, ela decidiu enviar 10 pessoas a uma

turnê com a banda mais barulhenta que ela pode encontrar, Motörhead. Essas pessoas foram a *shows* em Brixton (Londres), Brighton, Bristol, Edimburgo, Newcastle, Cardiff e Dublin. Após cada *show*, a otologista media o número de horas que as pessoas ficavam com os ouvidos zumbindo.

A Figura 3.16 mostra o número de horas que cada pessoa teve zumbido no ouvido após cada concerto (cada pessoa está representada por um círculo). As linhas horizontais representam o número médio de horas que existiu zumbido nos ouvidos após cada *show* e essas médias estão conectadas por uma linha de forma que podemos ver a tendência geral dos dados. Lembre-se de que para cada *show*, os círculos são os escores a partir do qual as médias são calculadas. Agora, podemos ver nos dois gráficos que a média aumenta à medida que as pessoas vão a mais *shows*. Assim, após o primeiro *show* os ouvidos zumbem por aproximadamente 12 horas, mas depois do segundo, eles zumbem por aproximadamente 15 a 20 horas e na noite final da turnê, eles zumbem por 45 a 50 horas (dois dias). Assim, existe um efeito acumulado dos *shows* sobre o zumbido nos ouvidos. Esse padrão é visto nos dois gráficos; a diferença entre os gráficos não está nas médias (que são basicamente as mesmas), mas em termos do espalhamento dos escores em torno da média. Se olharmos para o gráfico da esquerda, veremos que a dispersão dos escores permanece praticamente a mesma após cada *show* (os escores estão na maioria agrupados em torno da média). Colocando de outra forma, se for mensurada a distância vertical entre os escores mais alto e o mais baixo após o *show* de Brixton e fizermos o mesmo após os demais, todas essas distâncias serão aproximadamente iguais. Embora a média aumente, o espalhamento dos escores para a perda de audição é a mesma a cada nível da variável *show* (a dispersão dos escores é a mesma após Brixton, Brighton, Bristol, Edimburgo, Newcastle, Cardiff e Dublin). Isso é o que queremos dizer por *homogeneidade da variância*. O gráfico da direita mostra uma situação diferente: se você examinar a dispersão dos escores após o *show*

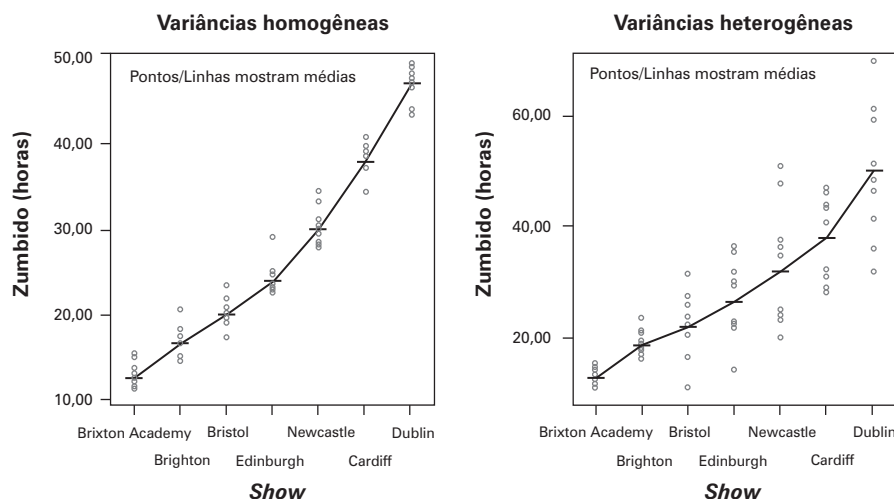


Figura 3.16 Gráficos ilustrando dados com variâncias homogêneas e heterogêneas.






de Brixton, verá que estão bem concentrados em torno da média (a distância vertical entre o maior e o menor escore é pequena), mas após o *show* de Dublin (por exemplo) os escores estão bastante espalhados em torno da média (a distância vertical entre o maior e o menor escore é grande). Esse é um exemplo de *heterogeneidade da variância*; isto é, em alguns níveis da variável *show* a variância dos escores é diferente dos outros níveis (nesse caso, a distância vertical entre o maior e o menor escore é diferente entre diferentes *shows*).

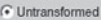



Espero que você tenha captado a idéia do que a homogeneidade da variância significa. Mas como testá-la? Muitos procedimentos estatísticos apresentam particularidades únicas para fazê-lo: na análise correlacional, como a regressão, a tendência é utilizar gráficos (veja a Seção 5.8.7) e para grupos de dados, a tendência é utilizar o teste de Levene. O teste de Levene verifica a hipótese de que a variância nos grupos é a mesma (isto é, a diferença entre as variâncias é zero). Dessa forma, se o teste de Levene é significativo quando $p \leq 0,05$, podemos concluir que a hipótese nula é incorreta e que as variâncias são significativamente diferentes – assim, a suposição de variâncias homogêneas foi violada. Se, no entanto, o teste de Levene não é significativo (isto é, p


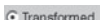
$> 0,05$), aceitamos a hipótese nula de que as diferenças entre as variâncias é zero – as variâncias são aproximadamente iguais e a suposição é mantida. Embora o teste de Levene possa ser selecionado como uma opção entre muitos testes estatísticos, ele também pode ser uma alternativa quando você está explorando os dados (e é melhor fazê-lo agora do que esperar até a análise principal).

Como ocorreu com o teste K-S (e outros de normalidade), quando o tamanho da amostra é grande, pequenas diferenças entre as variâncias dos grupos podem produzir um teste de Levene significativo (em virtude do que vimos no Capítulo 1, o poder do teste é aumentado). Uma dupla checagem útil é olhar a **razão das variâncias**. Isso é a razão das variâncias entre os grupos de pessoas com a maior variância e a do grupo com a menor variância. Assim, se você olhar para as variâncias entre grupos diferentes, pegue o valor mais alto que encontrar e divida pelo valor mais baixo que encontrar. Se o resultado (razão) for menor que 2, é seguro afirmar que há homogeneidade das variâncias.

É possível realizar o teste de Levene utilizando o menu *explore* (explorar) que foi usado na seção anterior. Para esse exemplo, pegaremos os dados do exame do SPSS que foram

utilizados na seção anterior (estão no arquivo **SPSSExam.sav**). Uma vez que os dados tenham sido carregados, isto é, o arquivo tenha sido aberto, utilize **Analyze⇒Descriptive Statistics⇒Explore...** (Analisar⇒Estatística Descritiva⇒Explorar...) para abrir a caixa de diálogo como a da Figura 3.17. Para manter as coisas de forma simples, olharemos apenas os escores do exame do SPSS e o de numeração desse arquivo, portanto, transfira essas duas variáveis da lista à esquerda do quadro denominado *Dependent List* (Lista Dependente) clicando em  próximo ao quadro e, já que queremos dividir a saída agrupando os dados para comparar as variâncias, selecione a variável **uni** e a transfira para o quadro denominado *Factor List* (Lista de Fatores) clicando em . Depois clique em  para abrir uma nova caixa de diálogo como na Figura 3.17. Para realizar o teste de Levene, precisamos selecionar uma das opções em que diz *Spread vs. Level with Levene's test* (Dispersão versus Nível com o Teste de Levene). Se você selecionar , o teste de Levene é realizado com os dados brutos (um bom começo). No entanto, se as variâncias forem diferentes, você pode corrigir a desigualdade utilizando uma das transformações que já foram vistas neste capítulo. Utilizando a mesma caixa de diálogo, se selecionarmos agora , você irá notar que um menu tipo lista suspensa ficará ativo (veja a Figura 3.17) e, se for selecionada uma dessas transformações listadas e rodar a análise, o SPSS irá calcular o que seria o teste de Levene se os dados fossem transformados utilizando esse método. Isso poderá lhe poupar bastante tempo que seria gasto

tentando várias transformações. Comece realizando a análise com  selecionado e então volte para realizá-la com  selecionada (tente uma transformação logarítmica [log] para começar). Quando terminar com essa caixa de diálogo, clique em  para retornar à caixa de diálogo principal do *explore* (explorar) e clique em  para rodar a análise.

A saída 3.5 do SPSS mostra a tabela do teste de Levene primeiro quando a análise foi feita com a opção  selecionada e depois com a opção  selecionada (logaritmo natural). Você deve ler as estatísticas baseadas na média. Para os dados não transformados, parece que o teste de Levene é não-significativo para os escores do exame do SPSS (valores na coluna rotulada como *Sig.* [Significância])(são maiores do que 0,05) indicando variâncias não significativamente diferentes (isto é, elas são similares e a suposição de variâncias homogêneas é mantida). No entanto, para os escores da numeração, o teste de Levene é significativo (valores na coluna denominada *Sig.* [Significância] são menores do que 0,05) indicando que as variâncias são significativamente diferentes (isto é, elas não são as mesmas e a hipótese de homogeneidade das variâncias foi violada). Assim, isso poderá nos fornecer um motivo para transformar os escores de numeração. Vamos ver o que ocorre quando fazemos isso. Bem, para os escores transformados logaritmicamente, o problema foi revertido: o teste de Levene é agora significativo para os escores do exame do SPSS (valores na coluna denominada *Sig* são menores do que 0,05), mas não é mais significativo para os escores de numeração.

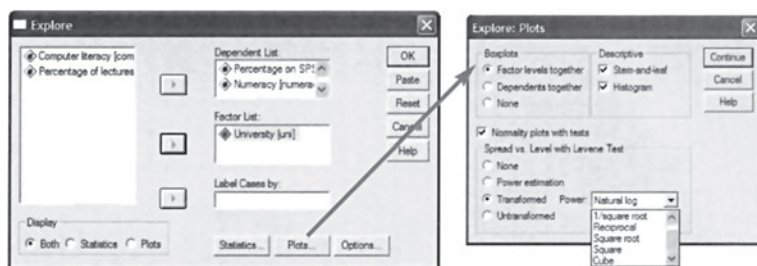


Figura 3.17 Explorando grupos de dados e obtendo o teste de Levene.

Saída 3.5 do SPSS

Dados não transformados

Tests of Homogeneity of Variance (Teste de homogeneidade das variâncias)

		Levene Statistic (Estatística de Levene)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Percentage on SPSS exam (Percentual no exame do SPSS)	Based on Mean (Baseado na média)	2.584	1	98	0.111
	Based on Median (Baseado na mediana)	2.089	1	98	0.152
	Based on Median and with adjusted df (Baseado na mediana com gl ajustado)	2.089	1	94.024	0.152
	Based on trimmed mean (Baseado na média interna)	2.523	1	98	0.115
Numeracy (Numerácia)	Based on Mean (Baseado na média)	7.368	1	98	0.008
	Based on Median (Baseado na mediana)	5.366	1	98	0.023
	Based on Median and with adjusted df (Baseado na mediana com gl ajustado)	5.366	1	83.920	0.023
	Based on trimmed mean (Baseado na média interna)	6.766	1	98	0.011

Dados transformados logaritmicamente

Tests of Homogeneity of Variance (Teste de homogeneidade das variâncias)

		Levene Statistic (Estatística de Levene)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Percentage on SPSS exam (Percentual no exame do SPSS)	Based on Mean (Baseado na média)	25.055	1	98	0.000
	Based on Median (Baseado na mediana)	24.960	1	98	0.000
	Based on Median and with adjusted df (Baseado na mediana com gl ajustado)	24.960	1	64.454	0.000
	Based on trimmed mean (Baseado na média interna)	25.284	1	98	0.000
Numeracy (Nu- merácia)	Based on Mean (Baseado na média)	0.211	1	98	0.647
	Based on Median (Baseado na mediana)	0.279	1	98	0.598
	Based on Median and with adjusted df (Baseado na mediana com gl ajustado)	0.279	1	97.664	0.598
	Based on trimmed mean (Baseado na média interna)	0.238	1	98	0.627

cia (valores na coluna denominada *Sig* não são menores do que 0,05). Nessa situação, considerando que estamos comparando as duas universidades nos escores de numerácia e exame do SPSS separadamente, devemos transformar os escores da numerácia, mas não os do exame. O teste de Levene poderá ser representado pela letra *F* e apresenta dois diferentes graus de liberdade. Como tal, ele poderá ser relatado, de forma geral, como $F(gl_1; gl_2) = \text{valor}$, *Sig*. Assim, para o percentual obtido no exame do SPSS, podemos dizer que as variâncias não são

iguais, $F(1, 98) = 2,58$, *ns*, mas para os escores da numerácia as variâncias são significativamente diferentes, $F(1, 98) = 7,37$, $p < 0,01$. Tente relatar esses resultados para os escores logaritmicamente transformados.

3.7 REPRESENTANDO MÉDIAS GRAFICAMENTE ①

Tendo realizado todas essas experiências com os dados, você pode querer representar algumas médias graficamente. Esta seção mostra

como utilizar alguns dos procedimentos gráficos do SPSS para criar gráficos de médias (isso é útil quando você fizer o relatório ou quando for apresentá-lo a outras pessoas). Como você irá criar esses gráficos no SPSS depende em grande parte da forma que os dados foram coletados (se foram utilizadas as mesmas pessoas ou pessoas diferentes para medir os diferentes níveis de cada variável). Para começar, imagine que o diretor de uma companhia cinematográfica está interessado em verificar se existe algo conhecido como “filme de garotas” (um filme que tipicamente interessa mais às mulheres do que aos homens). Ele pegou 20 homens e 20 mulheres e mostrou à metade de cada amostra um filme supostamente “de garotas” (*O diário de Bridget Jones*) e para a outra metade um filme que não se encaixaria nessa categoria (*Amnésia* – um filme brilhante, por sinal). Em todos os casos ele mensurou a excitação fisiológica como uma medida de quanto o filme foi apreciado. Os dados estão no arquivo denominado **ChickFlick.sav** no *site* www.artmed.com.br. Carregue esse arquivo agora.

Como utilizo
gráficos
interativos?



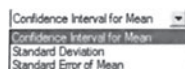
Para traçar um diagrama das médias, iremos utilizar os gráficos interativos do SPSS. A principal vantagem desses gráficos é que é possível colocar barras de erros (veja o Capítulo 8)

e eles podem ser atualizados mesmo quando a base de dados estiver fechada (veja o Quadro 3.4). Você pode também traçar gráficos utilizando o procedimento não interativo do SPSS, mas, para salvar árvores, escrevi isso em um arquivo que está no *site* www.artmed.com.br (**Answers(Chapter 3).pdf**) e no Capítulo 4. Vamos dar uma olhada em um traçado simples: primeiro clique em **Graphs⇒Interactive⇒Bar...** (Gráficos⇒Interativos⇒Barras...) para ativar a caixa de diálogo como a da Figura 3.18. Note que existe uma representação esquemática dos eixos horizontais e verticais e que ambos apresentam um espaço onde é possível arrastar uma variável da lista à esquerda. O eixo

y precisa ser a variável dependente ou o que você mensurou, ou simplesmente a coisa para a qual você quer apresentar a média. Nesse caso, será a *arousal* (**excitação**), assim, selecione-a da lista e arraste-a para o espaço que aparece no eixo vertical. O eixo x deve ser a variável na qual queremos dividir os dados da excitação. Para simplificar, vamos apenas traçar as médias para os dois filmes. Selecione a variável *film* (**filme**) da lista à esquerda e arraste-a para o espaço existente no eixo horizontal.⁸ Na caixa de diálogo embaixo existe uma seção denominada *Bars Represent...* (Barras representam) e ela ficará ativa somente depois que você especificar a variável do eixo vertical (ela é intitulada *Bars Represent Arousal* (Barras representam excitação) porque especificamos a excitação para o eixo vertical). Por padrão, as barras irão representar as médias (geralmente o que desejamos), mas você pode selecionar a partir da lista suspensa clicando em qualquer outra medida descritiva desde as de tendência central (como a mediana e a moda), medidas de dispersão (variância e desvio padrão) e medidas como a assimetria e a curtose.



Se quisermos adicionar as barras de erros (veja a Seção 7.3), devemos clicar na guia (orelha) *Error Bars* (Barras de Erros) no topo da caixa de diálogo e selecionar ☒ *Display Error Bars*. Por omissão, o SPSS irá mostrar um intervalo de confiança para a média (veja a Seção 1.6.2), mas se você escolher pela lista suspensa clicando em próximo de onde diz *Units* (Unidades), poderá escolher traçar o desvio padrão (veja a Seção 1.4.1) ou o erro padrão da média (veja a Seção 1.6). Finalmente, por padrão, você pode escolher uma representação diferente selecionando um dos quatro botões que estão embaixo. Prefiro o último que apresenta as linhas apontando para fora do topo de cada barra, assim, geralmente seleciono esse.



⁸ Para este trabalho é importante que você tenha especificado a variável como nominal (veja a Seção 2.4.3)

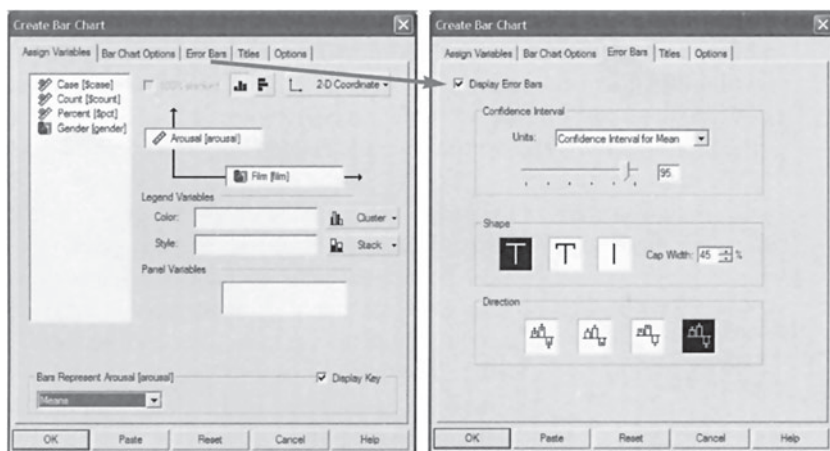


Figura 3.18 Caixas de diálogo para diagramas de barras interativo com barras de erros.

A Figura 3.19 mostra o diagrama de barras resultante. Ele simplesmente apresenta as médias (e o intervalo de confiança dessas médias). Esse gráfico demonstra que, em média, as pessoas ficaram mais excitadas pelo filme *Amnésia* do que por *O Diário de Bridget Jones*. Entretanto, inicialmente queríamos verificar os efeitos do gênero, dessa forma, esse gráfico não está informando o que precisamos

saber. O gráfico que precisamos é o agrupado (*clustered graph*) e ele é facilmente criado utilizando a mesma caixa de diálogo usada antes (veja a Figura 3.18), mas agora especificando uma segunda variável (nesse caso, gênero) no quadro que diz *Legend Variable* (Variável de Legenda) próximo de onde diz *Color* (Cores). Arraste a variável **gender** (gênero) para esse espaço. A caixa de diálogo final é mostrada na

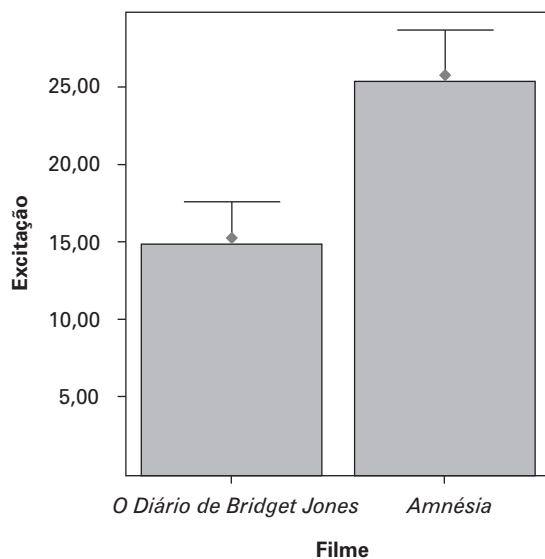


Figura 3.19 Diagrama de barras para a excitação média para cada um dos dois filmes.

Figura 3.20 e o diagrama de barras resultante, na Figura 3.21. Esse gráfico é mais informativo. Ele mostra o mesmo que o último gráfico: isto é, que a excitação foi mais alta para *Amnésia* do que para *O Diário de Bridget Jones*,

mas ele também divide isso por gênero. Olhe primeiro para a excitação média de *O Diário de Bridget Jones*; ele mostra que os homens ficaram de fato mais excitados do que as mulheres. Isso indica que eles gostaram mais do



Figura 3.20 Caixa de diálogo completa para diagrama de barras agrupado.

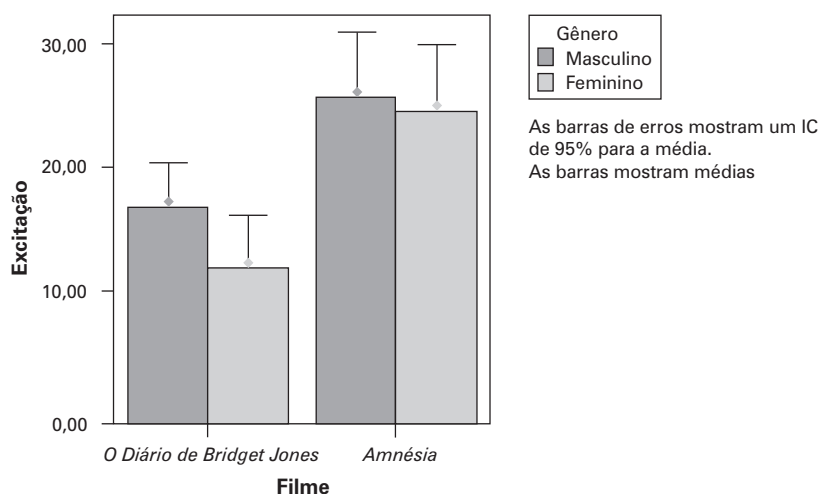
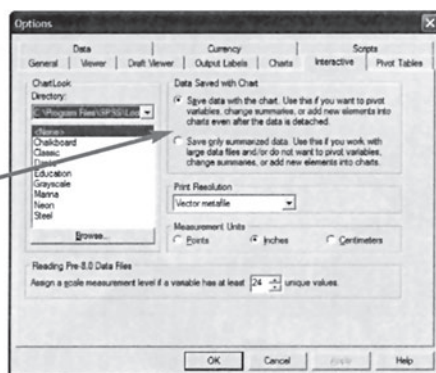


Figura 3.21 Gráfico agrupado mostrando a excitação de homens e mulheres medida durante os dois filmes.

Quadro 3.4

Transformando gráficos interativos em interativos! ①

A principal vantagem de utilizar um gráfico interativo no SPSS (em vez de um não-interativo) é que você pode atualizá-lo mesmo quando o arquivo de dados estiver fechado. No entanto, para que isso aconteça, você precisa marcar essa capacidade interativa que está desligada. Utilize o menu **Edit⇒Options...** (Editar⇒Opções...) para ter acesso à caixa de diálogo *options* (opções). Dentro dessa caixa de diálogo existe uma guia (orelha) no topo denominada *Interactive* (Interativo), assim, clique nessa orelha e selecione a opção *Save data with the chart* (Salve os dados com o gráfico). Isso transformará seu gráfico interativo em, de fato, interativo! Se você estiver utilizando uma grande base de dados talvez ache melhor desligar novamente essa opção e atualizar o gráfico manualmente se os dados forem alterados ou atualizados.



filme que as mulheres! Contraste isso com o filme *Amnésia*, no qual a excitação entre homens e mulheres é semelhante. Em face disso, a idéia do “filme de garotas” não procede: de fato, parece que homens gostam mais de filmes do tipo “de garotas” do que as próprias garotas (provavelmente por que é a única ajuda que temos para entender o complexo funcionamento da mente feminina!).

Você pode editar quase todos os aspectos de um gráfico clicando duas vezes no gráfico e então clicando na característica que deseja alterar. Você pode trocar as cores das barras, os títulos dos eixos, a escala de cada eixo e assim por diante. Pode, também, tornar as barras tridimensionais. Contudo, por mais tentador que isso possa ser (ele pode parecer muito bonito) tente resistir a tentação de construir gráficos 3-D (a menos que você esteja realmente

representando três variáveis): a dimensão extra não informa de útil e normalmente torna apenas o gráfico mais difícil de interpretar (veja Wright 2003).

3.8 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Este capítulo abordou muitas coisas rapidamente! Tentei lhe mostrar a importância de examinar os dados de forma apropriada antes de começar a análise estatística. Primeiro, verificamos a distribuição dos dados e tentamos descobrir possíveis erros que tenham sido cometidos na entrada ou digitação dos dados e também procuramos valores atípicos (*outliers*). Para distribuições não-normais, vimos como transformar os dados a fim de corrigir o problema e como tratar os valores atípicos.

Se você estiver trabalhando com dados que comparam diferentes grupos, deve examinar as distribuições dentro dos grupos. Você pode também aplicar testes estatísticos para verificar a normalidade dos dados (o teste de Kolmogorov-Smirnov) e olhar para outros gráficos como o diagrama *Q-Q*. Também descobrimos alguns usos para a função *transform* (transformar) do SPSS e para o comando *split file* (dividir arquivo). Finalmente, analisamos rapidamente como traçar médias utilizando gráficos interativos.

3.9 TERMOS-CHAVE QUE DESCOBRIMOS

- Bimodal
- Dados normalmente distribuídos
- Diagramas de caixa e bigodes
- Homogeneidade da variância
- Independência
- Intervalo interquartílico
- Mediana
- Multimodal
- Razão de variâncias
- Teste de Kolmogorov-Smirnov
- Teste de Levene
- Teste de Shapiro-Wilk
- Teste paramétrico
- Transformação
- Valor atípico

3.10 TAREFAS DO ALEX ESPERTO



- Utilizando os dados do Capítulo 2 (que você deve ter salvo, mas se não o fez, copie-os novamente da Tabela 2.1), faça alguns gráficos do número médio de amigos, consumo de álcool, renda e neurose dos estudantes e dos professores. Quem tem mais amigos, bebe mais, ganha mais e é mais neurótico? ^①
- Utilizando o arquivo **ChicFlick.sav**, verifique se as distribuições para os dois filmes (ignore o gênero) são normalmente distribuídas. ^①
- Utilizando os dados do arquivo **SPSSEXAM.sav** (lembre que os escores da numeração pareciam ter assimetria positiva – veja a Figura 3.10), transforme esses dados utilizando uma das transformações descritas no Capítulo; os dados se tornaram normais?

Algumas respostas breves podem ser encontradas no arquivo **Answers(Chapter 3).pdf** no site www.artmed.com.br.

3.11 LEITURAS COMPLEMENTARES

- TABACHNICK, B. G., FIDELL, L. S. *Using multivariate statistics*. Boston: Allyn & Bacon, 2001 (4ª ed.). O Capítulo 4 é o guia definitivo para a análise de dados!
- WRIGHT, D. B. *First steps in statistics*. London: Sage, 2002. O Capítulo 2 é uma boa introdução à construção de gráficos.

¹ IC = Intervalo de confiança

4.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①



Geralmente, é interessante para pesquisadores saber qual é o relacionamento que existe, se existe algum, entre duas ou mais variáveis. **Uma correlação é uma medida do relacionamento linear entre variáveis.**

Por exemplo, posso estar interessado no relacionamento entre o tempo gasto lendo este livro e o entendimento do leitor sobre estatística e SPSS. Essas duas variáveis podem estar relacionadas de várias maneiras: (1) elas podem estar *positivamente relacionadas*, o que significa que quanto mais tempo a pessoa gasta lendo este livro, maior o seu entendimento sobre estatística e SPSS; (2) elas podem não estar relacionadas de forma alguma, o que significa que o entendimento da pessoa sobre estatística e SPSS permanece o mesmo independentemente do tempo que ela gasta lendo este livro; ou (3) elas podem estar *negativamente relacionadas*, o que significa que quanto mais uma pessoa ler este livro, menor o seu entendimento sobre estatística e SPSS. Como podemos verificar se duas variáveis estão relacionadas? Este capítulo trata,

em primeiro lugar, de **como podemos expressar estatisticamente os relacionamentos entre variáveis utilizando duas medidas: a covariância e o coeficiente de correlação.** Depois, veremos como é possível representar graficamente os relacionamentos, antes de descobrir como investigar e interpretar correlações com o SPSS. O capítulo termina avaliando medidas mais complexas de relacionamento e, desse modo, introduz o capítulo sobre **regressão múltipla.**

4.2 COMO MEDIMOS RELACIONAMENTOS? ①

4.2.1 Um desvio para o mundo da covariância ①

A maneira mais simples de verificar se duas variáveis estão associadas é ver se elas variam conjuntamente. Para entender o que covariância significa precisamos lembrar o conceito de variância que encontramos no Capítulo 1. Lembre que a variância de uma variável representa a média que os dados se afastam da média. Em números, ela é descrita pela equação (4.1):

$$\begin{aligned} \text{Variância} &= s^2 \\ &= \frac{\sum (x_i - \bar{x})^2}{N - 1} = \frac{\sum (x_i - \bar{x})(x_i - \bar{x})}{N - 1} \quad (4.1) \end{aligned}$$

A média da amostra é representada por \bar{x} , o dado em questão é representado por x_i , e N é o número de observações (veja a Seção 1.4.1). Se estivermos interessados em verificar se duas variáveis estão relacionadas, precisamos ver se as mudanças em uma variável correspondem a mudanças similares na outra variável. **Portanto, quando uma variável se desvia da sua média, esperamos que a outra variável se desvie da sua média de maneira similar.** Para ilustrar o que quero dizer, imagine que pegamos cinco pessoas e as submetemos a certo número de propagandas promovendo as balas toffee e, então, medimos quantos pacotes dessas balas cada pessoa comprou durante a semana seguinte. Os dados são apresentados na Tabela 4.1, assim como a média e o desvio padrão (s) de cada variável.

Se existe um relacionamento entre essas duas variáveis, então enquanto uma variável se desvia da sua média, a outra variável deveria se desviar da sua média da mesma maneira ou de maneira diretamente oposta. A Figura 4.1 mostra os dados para cada participante (os quadrados representam o número de pacotes comprados e os círculos representam o número de propagandas assistidas); a linha pontilhada é o número médio de pacotes comprados e a linha cheia é o número médio de propagandas assistidas. As linhas verticais representam as diferenças entre os valores observados e a média da variável de interesse. A primeira coisa a ser observada na Figura 4.1 é que existe um padrão muito similar nas diferenças de ambas as variáveis. Para os três primeiros participantes, os valores observados estão abaixo da média para ambas variáveis e para as duas últimas pessoas, os valores observados estão acima da média para ambas variáveis. Esse padrão é indicativo de um relacionamento potencial entre duas variáveis (porque parece que se o escore de uma pessoa está abaixo da

média para uma variável, então o escore para a outra estará, também, abaixo da média).

Assim, como calculamos a semelhança exata entre o padrão das diferenças das duas variáveis mostradas na Figura 4.1? Uma possibilidade é calcular a quantidade total de diferenças, mas teríamos o mesmo problema que no caso da variável única (veja a Seção 1.4.1). Igualmente, adicionando simplesmente as diferenças, teríamos uma pequena percepção do relacionamento entre variáveis. Agora, no caso da variável única, elevamos ao quadrado as diferenças para eliminar o problema de diferenças positivas e negativas cancelarem uma a outra. Quando há duas variáveis, em vez de elevar ao quadrado cada diferença, podemos multiplicar a diferença de uma variável pela diferença correspondente da segunda variável. Se ambos os erros são positivos ou negativos, isso nos dará um valor positivo (indicando erros na mesma direção), mas se um erro for positivo e outro negativo, o produto resultante será negativo (indicando erros em direções opostas). Quando multiplicamos as diferenças de uma variável por diferenças correspondentes da segunda variável, conseguimos o que chamamos de *desvios dos produtos cruzados*. Com a variância, se queremos um valor médio das diferenças combinadas para as duas variáveis, devemos dividir pelo número das observações (dividimos, de fato, por $N - 1$). Essa média das diferenças combinadas é conhecida como *covariância*. Podemos escrever a covariância na forma de equação como em (4.2) – você poderá notar que a equação é igual a da variância, exceto que em vez de elevar ao quadrado as diferenças da mesma variável, multiplicamos as diferenças correspondentes de cada variável:

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \quad (4.2)$$

Tabela 4.1

Sujeito	1	2	3	4	5	Média	s
Comerciais Assistidos	5	4	4	6	8	5,4	1,67
Pacotes Comprados	8	9	10	13	15	11,0	2,92

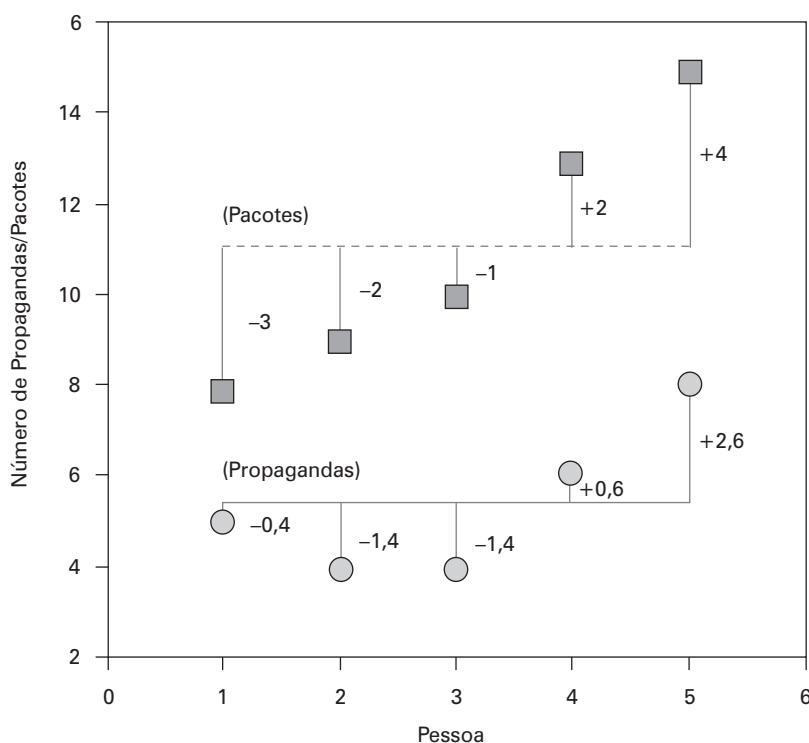


Figura 4.1 Gráfico mostrando as diferenças entre os dados observados e as médias de duas variáveis.

Para os dados da Tabela 4.1 e da Figura 4.1, temos o seguinte valor:

$$\begin{aligned}
 \text{cov}(x, y) &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \\
 &= \frac{(-0,4)(-3) + (-1,4)(-2) + (-1,4)(-1) + (-0,6)(2) + (2,6)(4)}{4} \\
 &= \frac{1,2 + 2,8 + 1,4 + 1,2 + 10,4}{4} \\
 &= \frac{17}{4} \\
 &= 4,25
 \end{aligned}$$

Calcular a covariância é uma boa maneira de avaliar se duas variáveis estão relacionadas entre si. Uma covariância positiva indica que

quando uma variável se desvia da média, a outra variável se desvia na mesma direção. Por outro lado, uma covariância negativa indica que, enquanto uma variável se desvia da média (por exemplo, aumenta), a outra se desvia da média na direção oposta (por exemplo, diminui).

Existe, entretanto, um problema com a covariância como uma medida do relacionamento entre variáveis: ela depende das escalas utilizadas para a medida. Assim, covariância não é uma medida padronizada. Por exemplo, se usarmos os dados acima e assumirmos que eles representam duas variáveis mensuradas em quilômetros, então a covariância é de 4,25 (calculada acima). Se convertermos esses dados em milhas (multiplicando todos os valores por 1,609) e calcularmos a covariância novamente, devemos descobrir que ela aumenta para 11. Essa dependência da escala de medidas é um problema porque significa que não podemos comparar a covariância

de uma maneira objetiva – assim, não podemos dizer se a covariância é particularmente grande ou pequena em relação a outro conjunto de dados a não ser que ambos os conjuntos fossem mensurados nas mesmas unidades.

4.2.2 Padronização e o coeficiente de correlação ①

Para superar o problema da dependência da escala de mensuração, precisamos converter a covariância em um conjunto padrão de unidades. Esse processo é conhecido como *padronização*. Uma forma básica de padronização seria insistir que em todos os experimentos fossem utilizadas as mesmas unidades de mensuração, digamos, metros – dessa maneira, todos os resultados poderiam ser facilmente comparados. Entretanto, se você fosse medir atitudes, por exemplo, seria muito difícil fazê-lo em metros! Portanto, precisamos de uma unidade de medida na qual qualquer escala de mensuração possa ser convertida. A unidade de medida que usamos é o *desvio padrão*. Falamos sobre essa medida na Seção 1.4.1 e vimos que, como a variância, ela é uma medida da média dos desvios a partir da média. Se dividirmos qualquer distância a partir da média pelo desvio padrão, isso nos dará essa distância em unidades de desvios padrão. Por exemplo, para os dados na Tabela 4.1, o desvio padrão para o número de pacotes comprados é aproximadamente 3,0 (o valor exato é 2,91). Na Figura 4.1, podemos ver que os valores observados para o participante 1 foram três pacotes a menos do que a média (assim, houve um erro de -3 pacotes de balas). Se dividirmos essa diferença, -3 , pelo desvio padrão, aproximadamente 3, teremos um valor de -1 . Isso nos mostra que a diferença entre o escore do participante 1 e a média foi de -1 desvio padrão. Assim, podemos expressar o desvio a partir da média para um participante em unidades padronizadas dividindo o desvio observado pelo desvio padrão.

Seguindo essa lógica, se quisermos expressar a covariância numa unidade padrão de medida, podemos simplesmente dividi-la pelo desvio padrão. Entretanto, existem duas variáveis

e, dessa maneira, dois desvios padrão. Quando calculamos a covariância, na verdade calculamos dois desvios (um para cada variável) e, então, os multiplicamos. Por conseguinte, fazemos o mesmo com os desvios padrão: nós os multiplicamos e dividimos pelo produto dessa multiplicação. **A covariância padronizada é conhecida como coeficiente de correlação.** E é definida pela equação (4.3) na qual s_x é o desvio padrão da primeira variável e s_y é o desvio padrão da segunda variável (todas as outras letras são as mesmas da equação definindo a covariância):

$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N - 1) s_x s_y} \quad (4.3)$$

O coeficiente na equação (4.3) é conhecido como o *coeficiente de correlação produto-momento de Pearson* (ou simplesmente *coeficiente de correlação de Pearson*) e foi inventado por Pearson (que surpresa!).¹ Se olharmos para a Tabela 4.1, vemos que o desvio padrão para o número de propagandas assistidas (s_x) foi 1,67 e para o número de pacotes de balas compradas (s_y) foi 2,92. Se multiplicarmos esses valores, temos $1,67 \times 2,92 = 4,88$. Agora, tudo o que precisamos fazer é pegar a covariância, que calculamos algumas páginas atrás como sendo 4,25, e dividir por esse produto. Isso nos dá $r = 4,25/4,88 = 0,87$ (esse é o mesmo valor que você encontrará mais tarde na Saída 4.1 do SPSS).

Padronizando a covariância, encontramos um valor que deve estar entre -1 e $+1$ (se você encontrar um coeficiente de correlação menor de que -1 ou maior do que $+1$, pode estar certo que alguma coisa saiu errada!) Um coeficiente de $+1$ indica que as duas variáveis são perfeitamente correlacionadas de forma positiva, assim, enquanto uma variável aumenta, a outra

¹ Você irá encontrar o coeficiente de correlação momento-produto de Pearson representado por ambos r e R . Tipicamente, a letra maiúscula é usada num contexto de regressão porque ela representa o coeficiente de correlação múltiplo; entretanto por alguma razão, quando elevamos r ao quadrado (como na Seção 4.5.3), é utilizado o R com letra maiúscula. Não me pergunte o porquê – é só para me confundir, eu acho.

Dica da Samanta Ferrinho



- Uma medida grosseira do relacionamento entre variáveis é a *covariância*.
- Se padronizarmos esse valor temos o *coeficiente de correlação de Pearson, r* .
- O coeficiente de correlação deve variar entre -1 e $+1$.
- Um coeficiente de $+1$ indica um relacionamento positivo perfeito, um coeficiente de -1 indica um relacionamento negativo perfeito, um coeficiente 0 indica que não existe relacionamento linear.
- O coeficiente de correlação é uma medida do tamanho de um efeito comumente utilizada: valores de $\pm 0,1$ representam um efeito pequeno, $\pm 0,3$ representa um efeito médio e $\pm 0,5$, um efeito grande.

aumenta proporcionalmente. Inversamente, um coeficiente de -1 indica um relacionamento negativo perfeito: se uma variável aumenta, a outra diminui por um valor proporcional. Um coeficiente 0 indica ausência de relacionamento linear, isto é, se uma variável muda, a outra permanece praticamente igual. Também vimos na Seção 1.8.4 que, como o coeficiente de correlação é uma medida padronizada de um efeito observado, ele é uma medida comumente usada do tamanho de efeito, e que valores de $\pm 0,1$ representam um efeito pequeno, $\pm 0,3$, um efeito médio e $\pm 0,5$, um efeito grande.

4.3 ENTRADA DE DADOS PARA REALIZAR UMA ANÁLISE DE CORRELAÇÃO NO SPSS ①

A entrada de dados para a correlação, a regressão e a regressão múltipla é simples porque cada variável é introduzida em uma coluna separada. Portanto, para cada variável que você mediu, crie uma variável no editor de dados com um nome apropriado, e digite os escores de um participante numa linha do editor de dados. Poderá ocorrer que você tenha uma ou mais variáveis categóricas (como o gênero) e essas variáveis também podem ser digitadas numa coluna (mas lembre-se de definir o valor apropriado dos rótulos). Por exemplo, se quiséssemos calcular a correlação entre as duas variáveis na Tabela 4.1, deveríamos entrar esses dados como na Figura 4.2. Ao longo deste capítulo iremos analisar um conjunto de dados baseado no desempenho em uma prova da faculdade. Os dados para vários

exemplos estão armazenados num único arquivo disponível no *site* www.artmed.com.br chamado de **ExamAnxiety.sav**. Se você abrir esse arquivo, verá que esses dados serão exibidos no editor em colunas separadas e que **gender** (gênero) foi codificado de forma correta. Iremos descobrir a que cada uma das variáveis se refere neste capítulo.

4.4 REPRESENTANDO RELACIONAMENTOS GRAFICAMENTE: O DIAGRAMA DE DISPERSÃO ①

Antes de conduzir qualquer análise de correlação é *essencial* traçar um diagrama de dispersão para ver a tendência geral dos dados.

	adverts	packets	
1	5.00	8.00	
2	4.00	9.00	
3	4.00	10.00	
4	6.00	13.00	
5	8.00	15.00	
6			

Figura 4.2 Entrada de dados para a correlação. O editor de dados mostra que o participante um viu cinco propagandas (**adverts**) e subsequentemente comprou oito pacotes (**packets**) de balas.



Um diagrama de dispersão é simplesmente um gráfico que coloca o escore de cada pessoa em uma variável contra seu escore em outra (e seu escore numa terceira variável pode ser incluído em um diagrama

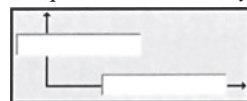
de dispersão 3D). Um diagrama de dispersão nos diz muitas coisas sobre os dados, por exemplo, se parece haver um relacionamento entre as variáveis, que tipo de relacionamento ele é e se quaisquer casos são significativamente diferentes dos outros. Um caso que difere substancialmente da tendência geral dos dados é conhecido como um *valor atípico* e tais casos podem distorcer severamente o coeficiente de correlação (veja o Quadro 3.1 e a Seção 5.6.1.1 para mais detalhes). Podemos usar um diagrama de dispersão para mostrar se algum caso parece um valor atípico.

Traçar um diagrama de dispersão usando o SPSS é fácil demais. Você pode traçá-lo usando o menu dos gráficos interativos ou usando gráficos não-interativos. Vamos examinar os dois tipos de gráficos porque há diagramas de dispersão que você pode criar apenas com gráficos não-interativos.

4.4.1 Diagrama de dispersão simples ①

Esse tipo de diagrama de dispersão serve para relacionar apenas duas variáveis. Por exemplo, uma psicóloga estava interessada nos efeitos do estresse pré-prova no desempenho na prova. Assim, ela planejou e validou um questionário para estimar o estado de ansiedade relacionado às provas (chamado de Questionário de Ansiedade Pré-Prova, ou QAP). Esse instrumento produziu uma medida de ansiedade numa escala de 1 a 100. A ansiedade foi medida antes de uma prova e a percentagem obtida para cada aluno foi usada para avaliar o desempenho na prova. Antes de verificar se essas variáveis estavam correlacionadas, a psicóloga fez um diagrama de dispersão das variáveis envolvidas (os dados estão no arquivo **ExamAnxiety.sav** que você já

deve ter aberto no SPSS). Para representar graficamente essas duas variáveis, você pode usar um gráfico interativo. Para tanto, selecione **Graphs⇒Interactive⇒Scatterplot...** (Gráficos⇒Interativo⇒Diagrama de Dispersão...) para abrir uma caixa de diálogo como a da Figura 4.3. Nessa caixa de diálogo, todas as variáveis do editor de dados são mostradas no lado esquerdo com espaços vazios no lado direito. Você simplesmente clica numa variável da lista à esquerda e a arrasta para o local apropriado. Em qualquer gráfico o eixo vertical é conhecido como o eixo *y* (ou ordenada) do gráfico. Essa variável deve ser a dependente (o resultado que foi medido)², que nesse caso é **exam performance (%) [exam]** (desempenho da prova (%) [prova]). O eixo horizontal é conhecido como o eixo *x* (ou abscissa) do gráfico. Essa variável deveria ser a variável independente, que é o caso em **exam anxiety [anxiety]** (ansiedade pré-prova [ansiedade]). A caixa de diálogo da Figura 4.3 tem uma representação esquemática do eixo *y* (a seta apontando para cima) e do eixo *x* (a seta apontando para a direita) e cada uma tem um espaço. Para especificar essas variáveis, precisamos arrastá-las para o espaço apropriado. Assim, arraste **exam performance (%) [exam]** (desempenho na prova (%) [prova]) para o espaço do eixo *y* e arraste **exam anxiety [anxiety]** (ansiedade pré-prova [ansiedade]) para o espaço do eixo *x*.



Mais abaixo na caixa de diálogo existem outros espaços nos quais você pode arrastar as variáveis. Você pode usar uma variável de grupo para definir diferentes categorias no diagrama de dispersão (ele irá exibir cada

² Na pesquisa experimental, é comum representar a variável independente no eixo horizontal e a variável dependente no eixo vertical. Nessa forma de pesquisa controlada, a implicação é que mudanças na variável independente (a variável que o pesquisador manipulou) causam mudanças na variável dependente. Na pesquisa correlacional, variáveis são mensuradas simultaneamente e, assim, nenhum relacionamento de causa e efeito pode ser estabelecido. Como tal, esses termos não devem ser encarados de forma rígida!

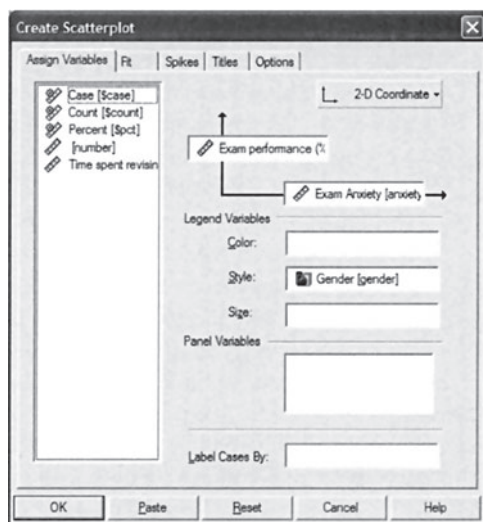


Figura 4.3 Caixa de diálogo para um diagrama de dispersão simples.

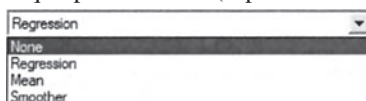
categoria usando uma cor diferente, símbolo ou gráfico). Essa função é útil para ver o relacionamento entre duas variáveis para grupos de pessoas. No exemplo atual, temos dados identificando se o estudante é do sexo masculino ou feminino, assim, podemos usar essas opções da seguinte maneira:

- **Color (Cor):** Se você selecionar **Gender [gender]** (Gênero [gênero]) da lista e arrastá-lo para esse espaço, o diagrama de dispersão resultante exibirá os dados masculinos numa cor diferente dos dados femininos.
- **Style (Estilo):** Se você selecionar **Gender [gender]** (Gênero [gênero]) da lista e arrastá-lo para esse espaço, o diagrama de dispersão resultante usará símbolos diferentes para os dados masculinos e femininos. Isso é o que escolhemos fazer e a caixa de diálogo resultante deverá ser parecida com a da Figura 4.3.
- **Size (Tamanho):** Se você selecionar **Gender [gender]** (Gênero [gênero]) da lista e arrastá-lo para esse espaço, os símbolos dos dados para o sexo masculino e feminino ficarão de tamanhos diferentes no diagrama de dispersão resultante.

• **Panel Variables (Variáveis do Painel):** Se você selecionar **Gender [gender]** (Gênero [gênero]) da lista e arrastá-lo para esse espaço, o SPSS irá produzir dois diagramas de dispersão (lado a lado) – um para o sexo masculino e outro para o feminino.

• **Label Cases by (Rotular Casos por):** Se você selecionar **Gender [gender]** (Gênero [gênero]) da lista e arrastá-lo para esse espaço, o SPSS colocará um rótulo próximo a cada ponto de dados no diagrama de dispersão indicando se aqueles dados são do sexo masculino ou feminino. Com muitos pontos de dados isso irá dificultar a leitura! Um dos usos pode ser se você tiver uma variável com o nome de um participante, ou um código; nesse caso, cada ponto no diagrama de dispersão será rotulado com um código que identifica a pessoa, o que pode ser útil para identificar valores atípicos e, então, ser realmente capaz de descobrir qual caso é um valor atípico!

No topo da caixa de diálogo há várias orelhas que dão acesso a outras opções. Se você clicar em **Titles (Títulos)**, por exemplo, verá uma caixa de diálogo que dá um espaço para digitar um título para o diagrama de dispersão. Uma orelha que pode ser útil (especialmente quando você ler o próximo capítulo)



título) é a **Fit (Ajustar)** (Figura 4.4). Dentro dessa caixa de diálogo você pode escolher uma linha para traçar selecionando um **Method (Método)** da lista suspensa. Se você escolher **regression** (regressão), o SPSS se ajustará à linha reta que melhor representa o relacionamento entre as variáveis no seu diagrama de dispersão (veja o Capítulo 5). Se você escolher **mean** (média), o SPSS traça uma linha horizontal representando a média, e se você escolher **smoother** (suavizar), uma linha curva que melhor descreva o relacionamento entre as variáveis no seu diagrama de dispersão será ajustada. Se você selecionar ☒ **Total**, o SPSS traçará essas linhas para todos os dados, e se você se-

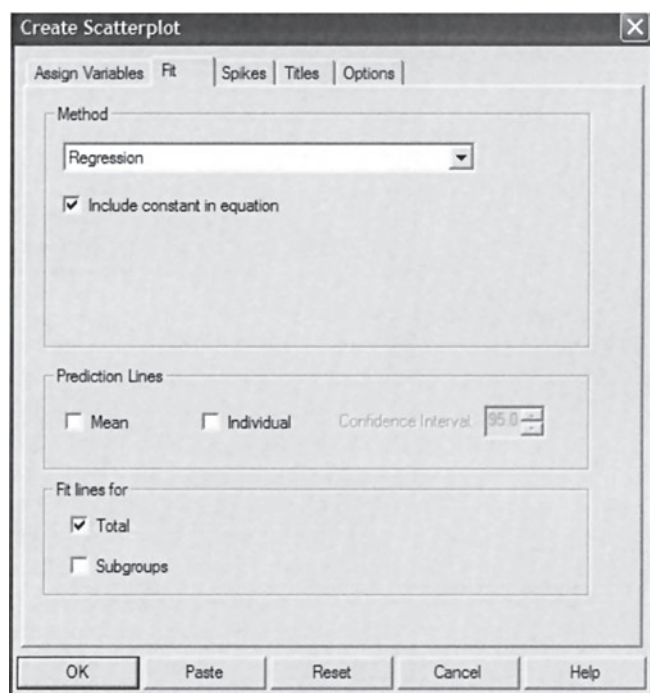


Figura 4.4 Caixa de diálogo para ajustar uma linha de regressão.

lecionar ☒ Subgroups (subgrupos), ele traçará linhas diferentes para diferentes grupos especificados por qualquer variável de grupo que você possa ter usado na caixa de diálogo do diagrama de dispersão original (as opções podem ser selecionadas ao mesmo tempo, assim, você pode ter uma linha geral e linhas de subgrupos no mesmo lote). No nosso exemplo, usamos gênero como uma variável de grupo; dessa forma, obtivemos uma linha para os homens e outra para as mulheres. No momento, não selecione nenhuma dessas opções, mas à medida que ler o Capítulo 5 tenha em mente que é assim que você obtém uma linha de regressão num diagrama de dispersão!

O diagrama de dispersão resultante é mostrado na Figura 4.5. O diagrama de dispersão na sua tela será um pouco diferente, pois os dados de homens e mulheres irão aparecer com símbolos distintos (triângulos e círculos). Substituí os marcadores de masculino e feminino para símbolos diferentes (isso me dá a oportunidade

de me exibir usando símbolos apropriados de gênero!). O diagrama de dispersão mostra que a maioria dos estudantes sofre de níveis altos de ansiedade há (poucos casos de ansiedade com níveis abaixo de 60). Também, não existem valores atípicos óbvios e a maioria dos pontos parece estar próximos uns dos outros. Parece, também, haver uma tendência geral nos dados tal que níveis altos de ansiedade estão associados a notas baixas nas provas e níveis baixos de ansiedade estão sempre associados a notas altas. Os marcadores de gênero mostram que a ansiedade parece afetar homens e mulheres da mesma maneira (porque os diferentes símbolos de homens e mulheres estão intercalados de maneira uniforme). Outra tendência observada nesses dados é que não houve casos de baixa ansiedade e baixo desempenho nas provas – de fato, a maioria dos dados está agrupada na região superior da escala de ansiedade.

Se existissem quaisquer pontos de dados que obviamente não se adequassem à tendên-

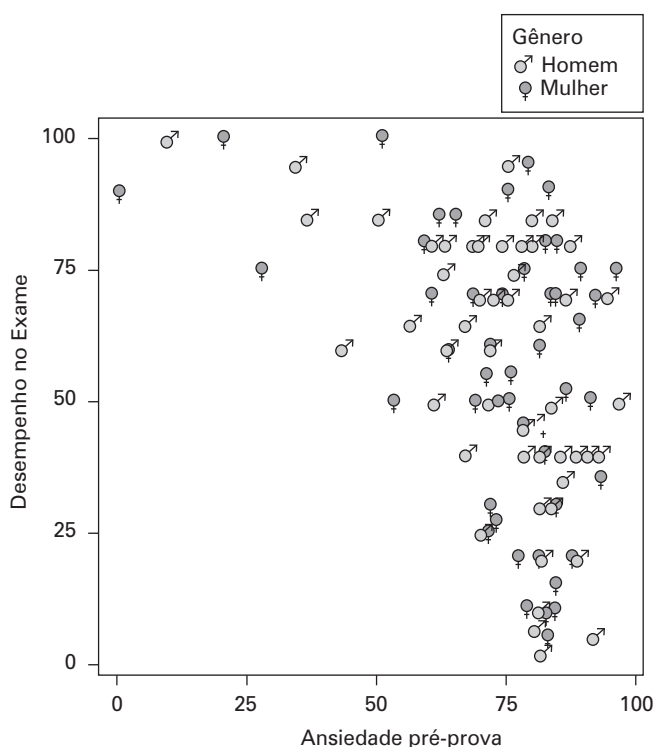


Figura 4.5 Diagrama de dispersão do desempenho na prova *versus* ansiedade pré-prova.

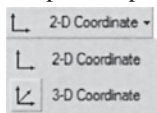
cia geral dos dados, precisaríamos descobrir se houve um bom motivo para que esses participantes respondessem de modo tão diferente. Algumas vezes os valores atípicos são apenas erros na entrada de dados (isto é, você digitou errado um valor), portanto, é bom checar novamente os dados na janela do editor para qualquer caso que pareça incomum. Se um valor atípico não pode ser explicado pela entrada incorreta de dados, é importante tentar descobrir se há uma terceira variável afetando o escore dessa pessoa. Por exemplo, uma pessoa pode estar sofrendo de ansiedade por algo diferente da prova e seu escore no questionário de ansiedade pode ter revelado essa ansiedade, mas pode ser a ansiedade específica com a prova que interfere no desempenho. Portanto, a ansiedade não relacionada dessa pessoa não afetou seu desempenho, mas resultou num escore alto de ansiedade. Se existe uma boa razão para um

participante responder de modo diferente dos outros, você pode considerar a eliminação daquele participante da análise a fim de construir um modelo preciso. Contudo, os dados dos participantes não devem ser eliminados porque eles não se encaixam nas suas hipóteses – somente se existir uma boa explicação para que tenham se portado de maneira tão estranha.

4.4.2 Diagrama de dispersão tridimensional ①

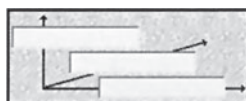
Agora lhe mostrarei uma das poucas vezes em que é apropriado usar um diagrama 3D! Um diagrama de dispersão em 3D é usado para mostrar o relacionamento entre três variáveis. A razão pela qual está correto usar um gráfico em 3D aqui é porque a terceira dimensão está realmente nos dizendo algo de útil (e ele não está lá apenas para parecer

brilhante)³. Por exemplo, imagine que a sua pesquisadora decidiu que ansiedade pré-prova pode não ser o único fator contribuindo para o desempenho na prova. Então, ela também pediu para os participantes manterem um diário de revisão, e ela calculou o número de horas gastas revisando a matéria da prova. A pesquisadora queria ver o re-



lacionamento entre essas variáveis simultaneamente. Para criar um diagrama de dispersão em 3D, use uma caixa de diálogo como a da Figura 4.3. Nessa caixa de diálogo existe uma seta com um menu suspenso rotulado de *2-D Co-ordinate* (Coordenadas 2D). Clique na seta para acessar o menu suspenso e você verá uma segunda opção rotulada de *3-D Co-ordinate* (Coordenadas 3D). Clique nessa opção e uma representação esquemática dos eixos irá mudar para incluir o terceiro eixo: o eixo z.

Usando a Figura 4.3 como um guia, complete a caixa de diálogo como anteriormente (tente colocar a variável do



gênero na caixa colorida dessa vez), mas selecione *3-D Co-ordinate* (Coordenada 3D) e então clique e arraste a variável **Time spent re-**



vising [revise] (Tempo gasto revisando [revisão]) para o espaço do eixo z. O diagrama de dispersão resultante é mostrado na Figura 4.6. Geralmente, os diagramas de dispersão em 3D podem ser difíceis de interpretar e, assim, sua utilidade em explorar dados pode ser limitada; entretanto, eles são ótimos para exibir dados de maneira concisa. Agora, se você clicar duas vezes nesse gráfico, você será capaz de girar os eixos para ter

uma idéia melhor do relacionamento entre as três variáveis – isso será útil para você; contudo, você será obrigado a decidir que ângulo melhor representa o relacionamento quando quiser colocar o gráfico no papel em 2D! Divirta-se com o gráfico e veja se você consegue a mesma figura que eu (a tela próxima desse texto irá ajudar!)

4.4.3 Diagrama de dispersão sobreposto ①

Existem dois tipos de diagramas de dispersão que você não pode fazer usando gráficos interativos: o diagrama de dispersão sobreposto e o diagrama de dispersão matricial. Ambos são muito úteis e vale a pena descobrir o porquê. Um diagrama de dispersão sobreposto é aquele em que muitos pares de variáveis são traçados nos mesmos eixos. Imagine que nossa pesquisadora quisesse verificar o papel da ansiedade pré-prova e do tempo de revisão no desempenho da prova (mas não o relacionamento entre a ansiedade pré-prova e revisão); seria útil traçar os relacionamentos entre tempo de revisão e desempenho na prova e entre ansiedade pré-prova e desempenho simultaneamente. Para tirar o melhor proveito desse tipo de diagrama de dispersão, mantenha uma variável constante e trace-a contra outras. No exemplo da prova, no qual o efeito da ansiedade e do tempo de revisão no desempenho da prova é relevante, você deveria traçar ansiedade (**anxiety**) (X) contra prova (**exam**) (Y) e, então, sobrepor revisão (**revise**) (X) contra desempenho na prova (**exam**) (Y).

Para traçar essas combinações, simplesmente use os menus como segue: **Graphs⇒Scatter....** (Gráficos⇒Espalhamento...). Isso ativa a caixa de diálogo da Figura 4.7, que fornece quatro opções para diferentes tipos de diagramas de dispersão disponíveis. Por omissão, um diagrama simples de dispersão é selecionado como é mostrado pela borda preta em volta da figura. Se você deseja traçar um diagrama de dispersão diferente, mova o cursor sobre as outras figuras e clique com o botão esquerdo do mouse. Depois de selecionar o diagrama de dispersão, clique em **Defina** (Definir).

³ Não é, na verdade, errado usar um diagrama em 3D quando você está representando somente duas variáveis. Quero dizer que nada de ruim irá lhe acontecer (a não ser que você o mostre para certos estatísticos que o trancarão numa sala e farão você escrever 75172 vezes no quadro “eu não devo fazer gráficos em 3D”), apenas não é uma maneira clara de apresentar os dados.

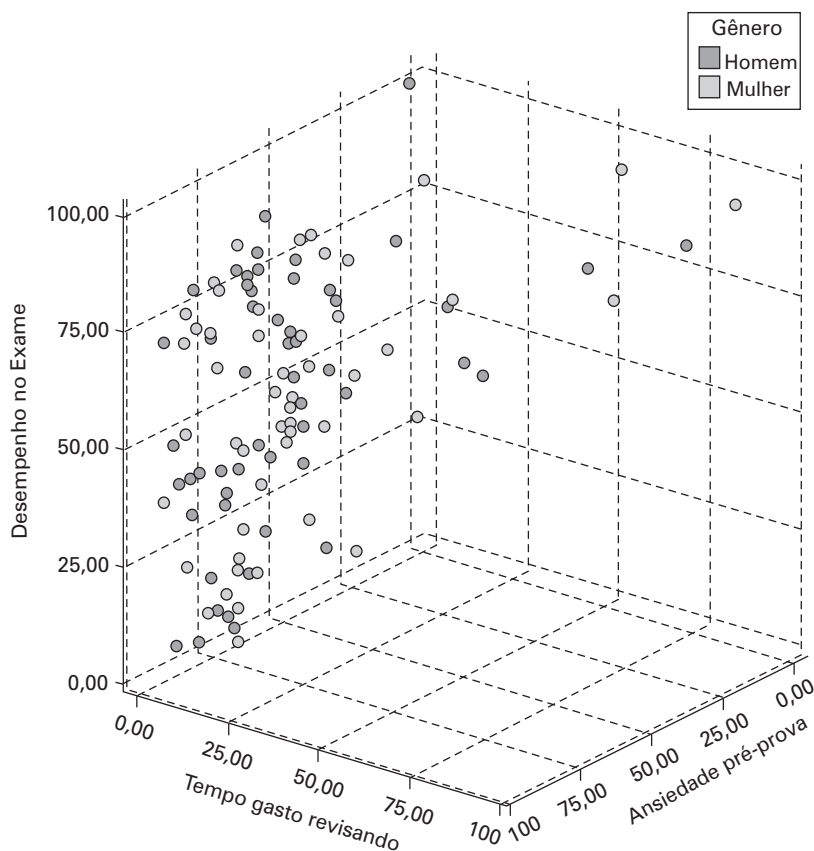


Figura 4.6 Um diagrama de dispersão em 3 D do desempenho da prova traçado contra a ansiedade pré-prova e o tempo gasto revisando.

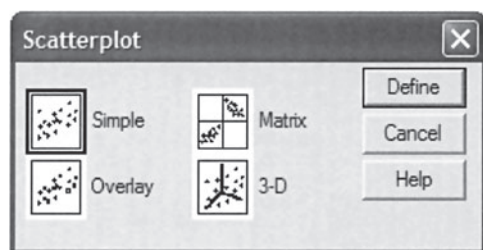



Figura 4.7 Caixa de diálogo do diagrama de dispersão principal.

Para traçar um diagrama de dispersão sobreposto, clique em *Overlay* (Sobreposto) na caixa de diálogo principal do diagrama de dis-

persão (*scatterplot*) (veja a Figura 4.7) e uma caixa de diálogo como a da Figura 4.8 aparece. Para selecionar um par de variáveis, clique em uma variável da lista (isso irá aparecer como *Variable 1* [Variável 1] na seção rotulada de *Current Selections* [Seleções Atuais]) e, então, selecione uma segunda variável (essa estará listada como *Variable 2* [Variável 2]). Transfira o par de variáveis clicando em . O par irá aparecer no espaço rotulado de *Y-X Pairs* (Pares Y-X). A ordem das variáveis está relacionada com o eixo em que elas estarão traçadas, assim, na Figura 4.8 **exam** (prova) irá aparecer no eixo **y** e **anxiety** (ansiedade) no eixo **x**. O segundo par, **exam** e **revise** (prova e revisão) deve, também, ser transferido para

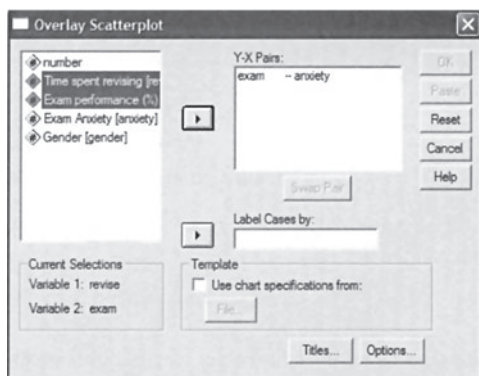


Figura 4.8 Caixa de diálogo para criar um diagrama de dispersão sobreposto. O primeiro par de variáveis já foi especificado e o segundo par foi selecionado, mas não transferido para a lista dos pares.

a lista de pares Y-X, de modo que **exam** (prova) esteja listada em primeiro lugar (e, assim, será traçada no eixo **y**) e **revise** (revisão) em segundo lugar (assim, será traçada no eixo **x**). Se ao transferir duas variáveis, elas não apare-

cerem da forma desejada, você poderá trocar a ordem das variáveis (e, portanto, o eixo no qual elas serão representadas) clicando em **Swap Pair** (trocar par).

Fica claro, a partir da Figura 4.9, que, embora a ansiedade esteja negativamente relacionada ao desempenho na prova, parece que a variável desempenho na prova está positivamente relacionada ao tempo de revisão. Assim, enquanto o tempo de revisão aumenta, o desempenho na prova aumenta também, mas, enquanto a ansiedade aumenta, o desempenho na prova diminui. O diagrama de dispersão sobreposto mostra de forma clara esses diferentes relacionamentos.

4.4.4 Diagrama de dispersão matricial ①

Em vez de traçar muitas variáveis nos mesmos eixos em um diagrama 3D (o que pode ser difícil de interpretar com vários pares de variáveis), é possível traçar uma matriz dos diagramas de dispersão. Esse tipo de diagrama permite ver o relacionamento entre todas as combinações dos diferentes pares de va-

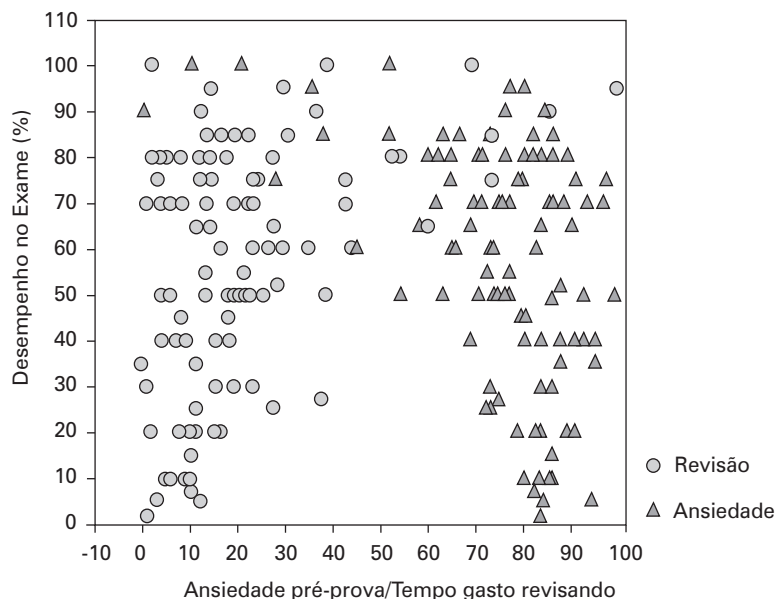



Figura 4.9 Um diagrama de dispersão do desempenho na prova *versus* a ansiedade pré-prova e tempo gasto revisando.

riáveis. Para obter um diagrama de dispersão matricial para os dados usados no diagrama de dispersão sobreposto, selecione a opção diagrama de dispersão Matricial (*Matrix*) na caixa de diálogo principal do diagrama de dispersão (Figura 4.7). A caixa de diálogo completa é mostrada na Figura 4.10: as variáveis estão listadas à esquerda e qualquer uma que você queira representar deverá ser transferida para a caixa rotulada de *Matrix Variables* (Variáveis Matriciais) usando a tecla . Para nossos dados, selecione e transfira **Exam performance (%) [exam]**, **Exam Anxiety [anxiety]** e **Time spent revising [revise]** (Desempenho na Prova (%) [prova], Ansiedade Pré-Prova [ansiedade] e Tempo gasto revisando [revisão]) para o quadro denominado *Matriz Variables* (Variáveis Matriciais). Como ocorreu com o diagrama de dispersão simples, existe uma opção de separar o diagrama de dispersão por uma variável de grupo especificando *Set Markers by...* (Determine marcadores por), mas não há necessidade de usar essa opção para esses dados. O diagrama de dispersão resultante do desempenho na prova contra ansiedade pré-prova e tempo de revisão é mostrado na Figura 4.11.

Os seis diagramas de dispersão na Figura 4.11 representam as várias combinações de cada variável traçada uma contra a outra.

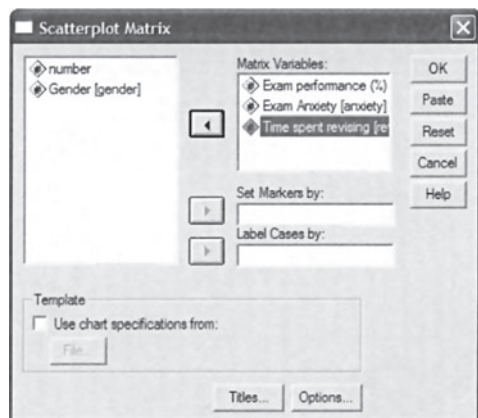


Figura 4.10 Caixa de diálogo para o diagrama de dispersão matricial.

Assim, a grade de referências representa os seguintes diagramas:

- **B1:** desempenho na prova (Y) *versus* ansiedade (X)
- **C1:** desempenho na prova (Y) *versus* tempo de revisão (X)
- **A2:** ansiedade (Y) *versus* desempenho na prova (X)
- **C2:** ansiedade (Y) *versus* tempo de revisão (X)
- **A3:** tempo de revisão (Y) *versus* desempenho na prova (X)
- **B3:** tempo de revisão (Y) *versus* ansiedade (X)

Dessa maneira, os três diagramas de dispersão abaixo da diagonal da matriz são os mesmos diagramas dos acima da diagonal, mas com eixos ao contrário. Dessa matriz podemos ver que o tempo de revisão e a ansiedade estão inversamente relacionados (assim, quanto mais tempo gasto revisando, menos ansiedade o participante teve na prova). Também, no diagrama de dispersão do tempo de revisão contra ansiedade (grade C2 e B3) parece que há um possível valor atípico – há um único participante que gastou pouquíssimo tempo revisando e ficou pouco ansioso em relação à prova. Como todos os participantes que tiveram baixa ansiedade obtiveram notas altas na prova, podemos deduzir que essa pessoa também foi bem na prova (você não odeia um Alex Esperto!). Poderíamos examinar esse caso mais detalhadamente se achássemos que seu comportamento foi provocado por algum fator externo (como tomar estimulantes para o cérebro!). Os diagramas de dispersão matriciais são muito convenientes para examinar pares de relacionamento entre variáveis. Entretanto, não recomendo traçá-los para mais de três ou quatro variáveis porque eles se tornam muito confusos.

4.5 CORRELAÇÃO BIVARIADA ①

Tendo uma visão preliminar dos dados, podemos agora realizar a análise de correlação. Para tanto, iremos retornar aos dados da

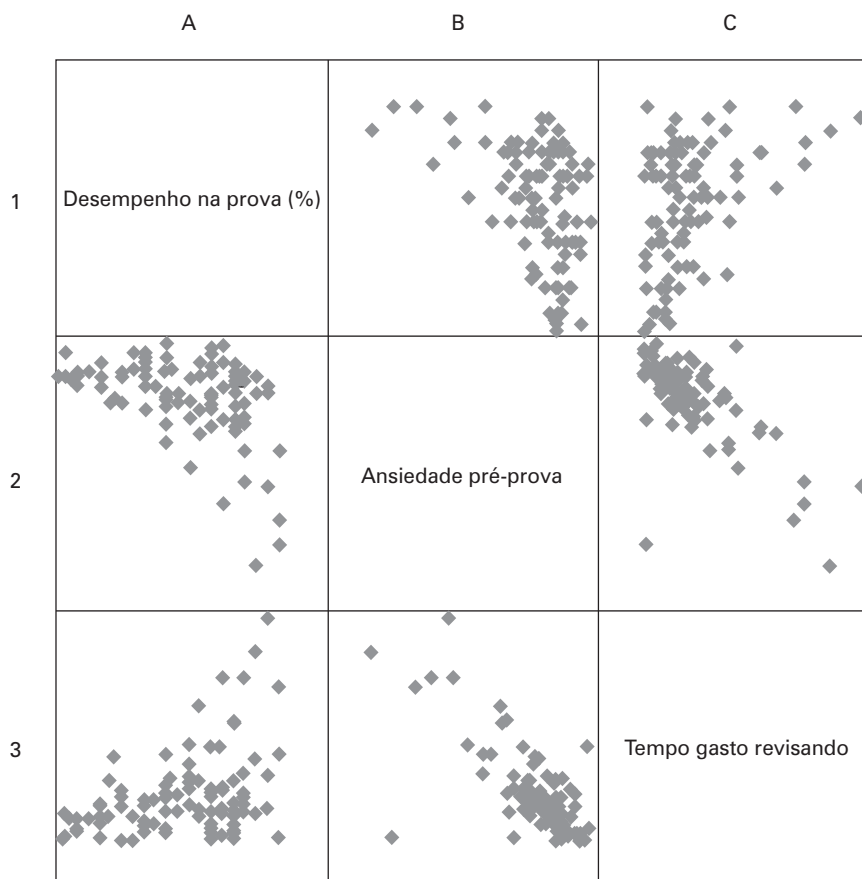


Figura 4.11 O diagrama de dispersão matricial do desempenho na prova, ansiedade pré-prova e tempo de revisão. Referências na grade foram adicionadas para esclarecimentos.

Tabela 4.1, que examinava se havia um relacionamento entre o número de propagandas assistidas e o número de pacotes de balas *toffee* comprados posteriormente. Crie duas colunas no editor de dados e digite esses dados (os preguiçosos podem acessar o arquivo **Advert.sav** no site www.artmed.com.br).

Existem dois tipos de correlação: *bivariada* e *parcial*. Uma correlação bivariada é uma correlação entre duas variáveis (como foi descrito no início deste capítulo), enquanto uma correlação parcial determina o relacionamento entre variáveis “controlando” o efeito de uma ou mais variáveis. **O coeficiente de correlação momento-produto de Pearson**

(descrito anteriormente) e o **rô de Spearman** (veja a Seção 4.5.4) são exemplos de **coeficientes de correlação bivariada**. Para conduzir uma correlação bivariada, você precisa encontrar a opção *Correlate* (Correlacionar) do menu *Analyse* (Analisar). A caixa de diálogo principal é acessada pelo caminho do menu **Analyse⇒Correlate⇒Bivariate...** (Analisar⇒Correlacionar⇒Bivariada...) e é mostrada na Figura 4.12. Usando a caixa de diálogo é possível selecionar quais das três correlações você quer determinar. O padrão estabelecido é a correlação de Pearson, mas você também pode calcular a de Spearman e a de Kendall – veremos as diferenças entre esses coeficientes

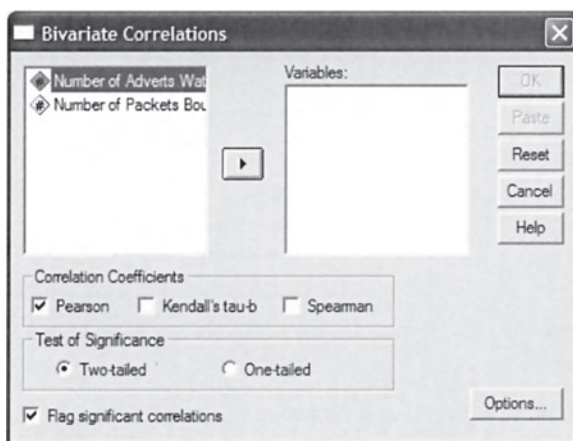


Figura 4.12 Caixa de diálogo ilustrando uma correlação bivariada.

de correlação no momento apropriado. Além disso, é possível especificar se o teste é ou não de uma ou duas caudas (veja a Seção 1.8.2). Um teste unilateral deve ser selecionado quando você tiver uma hipótese direcional (isto é, “quanto mais comerciais a pessoa assiste, mais pacotes de balas ela irá comprar”). Um teste bilateral (por omissão) deve ser usado quando não se pode prever a natureza do relacionamento (isto é, “Eu não tenho certeza se assistir a mais comerciais estará associado com o aumento ou a diminuição de pacotes de balas que uma pessoa compra”).

Tendo acessado a caixa de diálogo principal, você deverá perceber que as variáveis do editor de dados estão listadas no lado esquerdo da caixa de diálogo (Figura 4.12). Existe um quadro vazio rotulado de *Variables* (Variáveis) no lado direito. Você pode selecionar qualquer variável da lista usando o mouse e transferi-la para a caixa das Variáveis (*Variables*) clicando em . O SPSS irá criar uma tabela de coeficientes de correlação para todas as combinações de variáveis possíveis. Essa tabela é chamada de matriz de correlações. Para nosso exemplo atual, selecione as variáveis **Number of Adverts Watched** [adverts] e **Number of Packets Bought** [packets] (Número de Comerciais Assistidos [comerciais] e Número de Pacotes Comprados [pacotes]) e

transfira todas para a lista de variáveis. Uma vez selecionadas as variáveis de interesse, você pode escolher entre três coeficientes de correlação: o de Pearson, o ρ de Spearman e o tau de Kendall. Qualquer um deles pode ser selecionado clicando na caixa de marcar.

Se você clicar em (Opções), outra caixa de diálogo aparece com duas opções de estatísticas (*Statistics*) e duas opções para valores desconhecidos (*missing*) (Figura 4.13). A opção estatísticas estará ativa somente quando a correlação de Pearson for selecionada; se a correlação de Pearson não for selecionada, então essas opções estarão desabilitadas (elas aparecem num cinza claro em vez de preto e você não pode ativá-las). Essa desativação ocorre porque essas duas opções são impor-

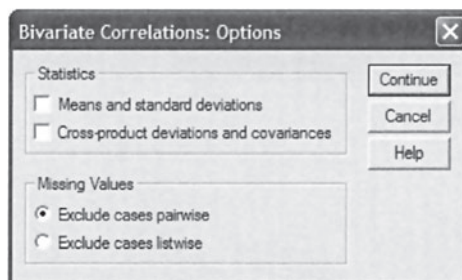


Figura 4.13 Caixa de diálogo para as opções da correlação bivariada.

tantes somente para dados paramétricos e a correlação de Pearson é usada com esse tipo de dados. Se você selecionar a caixa de marcar rotulada de *Means* (Médias) e *Standard deviations* (desvios padrão), o SPSS calculará a média e o desvio padrão de todas as variáveis selecionadas para serem correlacionadas. Se você ativar a caixa de marcar rotulada *Cross-product deviation* (desvio do produto cruzado) e *covariances* (covariâncias), o SPSS fornecerá os valores dessas estatísticas para cada variável sendo correlacionada. O desvio do produto cruzado nos dá a soma dos desvios em relação à medida de cada variável e é simplesmente o numerador (metade acima) da equação (4.2). As opções de covariância nos dão os valores da covariância entre as variáveis, os quais poderiam ser calculados manualmente usando a equação (4.2). Em outras palavras, esses valores das covariâncias são os desvios dos produtos cruzados divididos por $(N - 1)$ e representam o coeficiente de correlação não-padronizado. Em muitos casos, você não precisará usar essas opções, mas, ocasionalmente elas poderão ser úteis!

Para ilustrar o que essas opções fazem, selecione-as para os dados dos comerciais. Deixe as opções selecionadas na caixa de diálogo principal como elas estão. O resultado do SPSS é mostrado na Saída 4.1. A seção rotulada de *Sums of Squares* e *Cross-products*

(Soma dos Quadrados e Produtos Cruzados) nos mostra o produto cruzado (17 nesse exemplo) que calculamos da equação (4.2) e a soma dos quadrados para cada variável. A soma dos quadrados é calculada no topo da equação (4.1). O valor da covariância entre as duas variáveis é 4,25, que é o mesmo valor que foi calculado da equação (4.2). A covariância de uma única variável é o mesmo que a variância da variável (assim, a variância para o número de comerciais é 2,8 e a variância para o número de pacotes comprados é 8,5). Essas variâncias podem ser calculadas manualmente pela equação (4.1). Note, também, que o coeficiente de correlação de Pearson entre as duas variáveis é 0,871, o mesmo valor calculado na Seção 4.22.

4.5.1 Coeficiente de correlação de Pearson ①

O coeficiente de correlação de Pearson foi descrito detalhadamente no início deste capítulo. Para aqueles que não estão familiarizados com a estatística básica, não é relevante falar sobre média a não ser que tenhamos dados mensurados num intervalo ou num nível de razão (para uma revisão, veja o Quadro 2.1 ou Field e Hole, 2003). A correlação de Pearson requer dados intervalares para ser uma medida precisa do relacionamento linear

Saída 4.1 do SPSS Procedimento correlação
Correlations (Correlações)

		<i>Adverts Watched</i> (Comerciais vistos)	<i>Number of Packets</i> (Número de pacotes)
<i>Adverts Watched</i> (Comerciais vistos)	<i>Pearson Correlation</i> (Correlação de Pearson)	1.000	0.871
	<i>Sig. (2 tailed)</i> (Sig. bilateral)	.	0.054
	<i>Sum of Squares and Cross-products</i> (Soma dos quadrados e produtos cruzados)	11.200	17.000
	<i>Covariance</i> (Covariância)	2.800	4.250
	<i>N</i>	5	5
<i>Number of Packets</i> (Número de pacotes)	<i>Pearson Correlation</i> (Correlação de Pearson)	0.871	1.000
	<i>Sig. (2 tailed)</i> (Sig. bilateral)	0.054	.
	<i>Sum of Squares and Cross-products</i> (Soma dos quadrados e produtos cruzados)	17.000	34.000
	<i>Covariance</i> (Covariância)	4.250	8.500
	<i>N</i>	5	5

entre duas variáveis. Entretanto, se você quer determinar se o coeficiente de correlação é significativo, mais suposições são necessárias: para que a estatística teste seja válida os dados devem ser normalmente distribuídos. Embora queiramos que ambas as variáveis sejam distribuídas normalmente, há uma exceção para essa regra: uma das variáveis pode ser categórica contanto que ela tenha somente duas categorias (na realidade, se você olhar a Seção 4.5.6, verá que isso é o mesmo que fazer um teste-*t*, mas estou me apressando um pouco). De qualquer maneira, se os seus dados não são normais (veja o Capítulo 3) ou não são mensurados em intervalos, você não deve selecionar a correlação de Pearson (Figura 4.14).

Acesse novamente os dados do arquivo **ExamAnxiety.sav** para que possamos calcular a correlação entre o desempenho na prova, tempo de revisão e ansiedade pré-prova. Os dados do desempenho na prova estudados são paramétricos e, assim, a correlação de Pearson pode ser aplicada. Acesse a caixa de diálogo principal das *bivariate correlations* (correlações bivariadas) (**Analyze**⇒**Correlate**⇒**Bivariate...**) (Analisar⇒Correlacionar⇒Bivariada...) e transfira **exam**, **anxiety** e **revise** (prova, ansiedade e revisar) para o quadro rotulado de *Variables* (Variáveis) (Figura 4.15). A caixa de diálogo também nos permite especificar



Figura 4.14 Karl Pearson.

se o teste é uni ou bilateral. Testes unilaterais devem ser usados quando existe uma direção específica para a hipótese que está sendo testada e testes bilaterais devem ser usados quando um relacionamento é esperado, mas a direção do relacionamento não é previsível. Nosso pesquisador previu que em altos índices de ansiedade o desempenho da prova seria baixo. Portanto, o teste para essas variáveis deve ser unilateral porque antes dos dados serem cole-

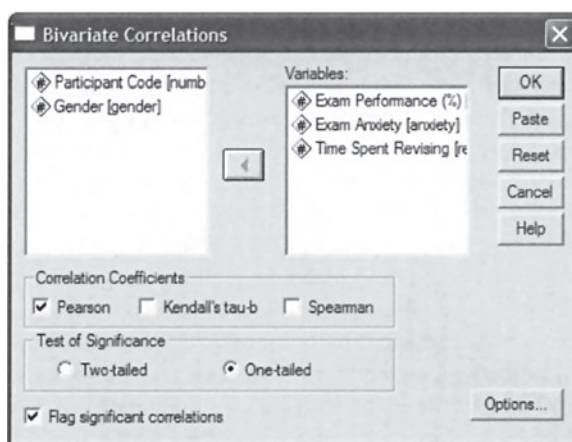


Figura 4.15 Caixa de diálogo completa para os dados do desempenho na prova.

tados o pesquisador previu um tipo específico de relacionamento. E mais, uma correlação positiva entre tempo de revisão e desempenho na prova foi previsto, assim, esse teste também deve ser unilateral. Para assegurar que iremos obter os valores de significância unilaterais selecione **One-tailed** (unilateral) e clique em **OK**.

A saída 4.2 do SPSS fornece uma matriz dos coeficientes de correlação para as três variáveis. Embaixo de cada coeficiente de correlação é apresentado o valor da significância da correlação e o tamanho da amostra (*N*) no qual ele é baseado. Cada variável é perfeitamente correlacionada com ela própria (obviamente) e, assim, $r = 1$ ao longo da diagonal na tabela. O desempenho na prova é negativamente relacionado à ansiedade pré-prova com um coeficiente $r = -0,441$ e existe uma probabilidade de menos de 0,001 que esse coeficiente de correlação tenha ocorrido por acaso numa amostra de 103 pessoas (como está indicado pelos dois asteriscos depois do coeficiente). Esse valor da significância nos diz que a probabilidade dessa correlação ser “um golpe de sorte” é muito baixa (próximo a zero, na verdade). Desde já, podemos ter certeza de que o relacionamento entre desempenho na prova e ansiedade é genuíno. Geralmente, cientistas sociais aceitam qualquer valor de probabilidade abaixo

de 0,05 como sendo estatisticamente significativo e, assim, tal valor é considerado como um indicativo de um efeito genuíno (o SPSS irá marcar qualquer coeficiente de correlação significativo, nesse nível, com um asterisco). A saída mostra, também, que o desempenho na prova está positivamente relacionado com o tempo gasto revisando, com um coeficiente de $r = 0,397$, o qual também é significativo a $p < 0,001$. Finalmente, a ansiedade pré-prova parece ser negativamente relacionada ao tempo gasto revisando ($r = -0,709$, $p < 0,001$).

Em termos psicológicos, isso tudo significa que, à medida que a ansiedade sobre uma prova aumenta, a percentagem de acertos obtida naquele exame diminui. Inversamente, à medida que o tempo de revisão aumenta, a percentagem obtida na prova aumenta. Finalmente, à medida que o tempo de revisão aumenta, a ansiedade do estudante sobre a prova diminui. Assim, há um inter-relacionamento complexo entre as três variáveis.

4.5.2 Um alerta sobre interpretação: Causalidade ①

Devemos ser cuidadosos ao interpretar coeficientes de correlação porque eles não dão indicação da direção da causalidade. Assim,

Saída 4.2 do SPSS Saída do SPSS para a correlação de Pearson

Correlations (Correlações)

		<i>Exam performance (%)</i> (Desempenho na prova)	<i>Exam Anxiety</i> (Ansiedade pré-prova)	<i>Time spent revising</i> (Tempo gasto revisando)
<i>Exam performance (%)</i> (Desempenho na prova)	<i>Pearson Correlation</i> (Correlação de Pearson) <i>Sig. (1 tailed)</i> (Sig. unilateral) <i>N</i>	1.000 . 103	- 0.441** 0.000 103	0.397** 0.00 103
<i>Exam Anxiety</i> (Ansiedade pré-prova)	<i>Pearson Correlation</i> (Correlação de Pearson) <i>Sig. (1 tailed)</i> (Sig. unilateral) <i>N</i>	- 0.441** 0.000 103	1.000 . 103	- 0.709** 0.000 103
<i>Time spent revising</i> (Tempo gasto revisando)	<i>Pearson Correlation</i> (Correlação de Pearson) <i>Sig. (1 tailed)</i> (Sig. unilateral) <i>N</i>	0.397** 0.000 103	- 0.709** 0.000 103	1.000 . 103

Correlation is significant at the .01 level (1-tailed) (Correlação é significativa ao nível de 0,01 (unilateral))

em nosso exemplo, embora possamos concluir que o desempenho na prova diminui à medida que a ansiedade pré-prova aumenta, não podemos dizer que uma alta ansiedade pré-prova *causa* péssimo desempenho na prova. Essa precaução tem dois motivos:

- **O problema da terceira variável:** Em qualquer correlação bivariada a causalidade entre duas variáveis não pode ser dada por certo, porque podem ter outras variáveis, medidas ou não, afetando os resultados. Isso é conhecido como o problema da *terceira variável* ou o *tertium quid*. No nosso exemplo, você pode ver que o tempo de revisão se relaciona significativamente a ambos, desempenho na prova e ansiedade pré-prova, e não há maneira de dizer qual das duas variáveis independentes, se ambas, causa a mudança no desempenho na prova. Assim, se tivéssemos medido somente a ansiedade pré-prova e o desempenho na prova, poderíamos ter assumido que a alta ansiedade pré-prova causou péssimo desempenho na prova. Entretanto, está claro que péssimo desempenho na prova também pode ser explicado pela falta de revisão. Pode haver muitas variáveis adicionais que influenciam as variáveis correlacionadas e essas variáveis podem não ter sido medidas pelo pesquisador. Assim, poderia ser outra variável, não mensurada, que afeta o tempo de revisão e a ansiedade pré-prova. Field e Hole (2003) dão o exemplo da descoberta de Broca de um forte relacionamento entre gênero e tamanho do cérebro: **o cérebro das mulheres são, em geral, menores do que os dos homens. Isso foi, naquela época, usado como evidência para argumentar que as mulheres eram intelectualmente inferiores aos homens. O problema óbvio é que esse relacionamento não leva em consideração o tamanho do corpo: pessoas com corpos maiores têm cérebros maiores independentemente da habilidade intelectual. Gould (1981) analisou novamente os dados e demonstrou que o forte relacionamento entre gênero e tamanho**

do cérebro desapareceu quando se levou em conta o tamanho do corpo!

- **Direção da causalidade:** Coeficientes de correlação nada dizem sobre qual variável causa a alteração na outra. Mesmo se pudéssemos ignorar o problema da terceira variável descrito acima, e pudéssemos assumir que as duas variáveis correlacionadas eram as únicas importantes, o coeficiente de correlação não indica em qual direção a causalidade opera. Assim, embora seja tentador concluir que a ansiedade pré-prova causa a mudança no desempenho na prova, não existe razão estatística de que o desempenho na prova não possa causar a mudança na ansiedade pré-prova. Embora essa última conclusão não faça sentido (porque a ansiedade foi medida antes do desempenho na prova), a correlação não nos diz que isso não é verdade.

4.5.3 Utilizando o R^2 para interpretação ①

Embora não possamos tirar conclusões diretas sobre causalidade, podemos levar o coeficiente de correlação um passo a frente elevando-o ao quadrado. O coeficiente de correlação ao quadrado (conhecido como o coeficiente de determinação, R^2) é uma medida da quantidade de variação em uma variável que é explicada pela outra. Por exemplo, podemos olhar o relacionamento entre ansiedade pré-prova e desempenho na prova. O desempenho na prova varia de pessoa para pessoa devido a diversos fatores (habilidades diferentes, diferentes níveis de preparação e assim por diante). Se adicionássemos todas as variabilidades (como quando calculamos a soma dos quadrados na Seção 1.4.1), teríamos uma estimativa de quanta variação existe no desempenho na prova. Podemos usar, então, o R^2 para nos dizer quanto dessa variabilidade é provocado pela ansiedade pré-prova. Essas duas variáveis tinham uma correlação de $-0,4410$ e, assim, o valor de R^2 será $(-0,4410)^2 = 0,194$. Esse valor nos diz quanto da variação no desempenho na prova pode ser explicado pela ansiedade pré-prova. Se convertermos esse valor em percentagem, podemos dizer que

ansiedade pré-prova tem 19,4% de variabilidade no desempenho na prova. Assim, embora a ansiedade pré-prova estivesse altamente correlacionada com o desempenho na prova, ela contribui somente com 19,4% da variação nos escores da prova. Colocando esse valor em perspectiva, isso nos leva a 80,6% da variabilidade ainda a ser contabilizada por outras variáveis. Devo ressaltar que embora R^2 seja uma medida útil da importância de um efeito, ele não pode ser usado para inferir relacionamentos causais. Embora geralmente falemos em termos de “a variância em y ser responsável por x ” ou mesmo a variação em uma variável explicada pela outra, isso ainda nada diz sobre qual o caminho da causalidade. Assim, mesmo que a ansiedade pré-prova seja responsável por 19,4% da variação em escores de exames, ela não necessariamente causa a variação.

4.5.4 Coeficiente de correlação de Spearman ①

O que acontece se os meus dados não são paramétricos?



O coeficiente de correlação de Spearman, r_s , é uma estatística não-paramétrica e, assim, pode ser usada quando os dados violarem suposições paramétricas, tais

como dados não-normais (veja o Capítulo 3).

O teste de Spearman trabalha classificando os dados em primeiro lugar e então aplicando a equação de Pearson (equação (4.3)) aos dados ordenados. Como professor de estatística, estou sempre interessado em fatores que determinam se um estudante irá bem em um curso de estatística. Um fator consideravelmente importante é seu conhecimento prévio em matemática (no mínimo a experiência prévia irá determinar se uma equação faz sentido!). Imagine que eu peguei 25 alunos e olhei suas notas no curso de estatística no final do seu primeiro ano na universidade. Na Grã Bretanha, o estudante pode obter um conceito de primei-

ra classe, um conceito de segunda classe superior, um conceito de segunda classe inferior, um de terceira classe, passar ou reprovar. Esses conceitos são categorias, mas eles têm uma ordem (um conceito de segunda classe superior é melhor do que um de segunda inferior). Eu poderia, também, perguntar a esses alunos que conceito eles obtiveram nas provas de matemática do GCSE. Na Grã Bretanha, GCSEs são provas escolares feitas com a idade de 16 anos que levam os conceitos A, B, C, D, E ou F. Novamente, esses conceitos são categorias que têm uma ordem de importância (um A é melhor do que todos os outros conceitos). Quando você tem categorias como essas que podem ser ordenadas de uma maneira significativa, os dados são denominados de *ordinais*. Os dados não são intervalares porque um conceito de primeira classe cobre 30% do intervalo (70% – 100%) enquanto um de segunda superior atinge somente 10% do intervalo (60% – 70%). Quando os dados foram medidos somente no nível ordinal eles são chamados de não-paramétricos e a correlação de Pearson não é apropriada. Portanto, o coeficiente de correlação de Spearman é usado. Os dados para esse estudo estão no arquivo **grades.sav**. Eles estão organizados em duas colunas: uma rotulada de **stats** e uma rotulada de **gcse**. Cada uma das categorias descritas acima foi codificada com um valor numérico. Em ambos os casos, o conceito mais alto (primeira classe ou conceito A) foi codificado com o valor de 1, com as categorias subsequentes rotuladas 2, 3 e assim por diante. Note que para cada valor numérico providenciei um rótulo (como fizemos para codificar variáveis).

O procedimento para determinar a correlação de Spearman (Figura 4.16) é o mesmo para a correlação de Pearson exceto que na caixa de diálogo das correlações bivariadas (Figura 4.15), precisamos selecionar ☒ Spearman e não a opção da correlação de Pearson. Nesse estágio, você deve também especificar se quer um teste uni ou bilateral. Para o exemplo acima, previ que as melhores notas no GCSE de matemática seriam correlacionadas com

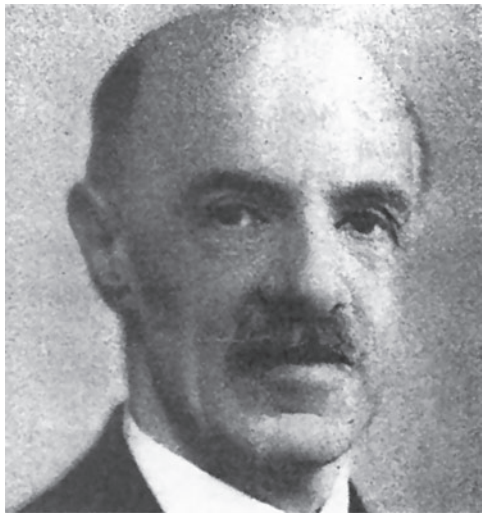


Figura 4.16 Charles Spearman, classificando furiosamente.

melhores notas no meu curso de estatística. A hipótese é direcional e um teste unilateral deve ser selecionado.

A Saída 4.3 do SPSS mostra a correlação de Spearman nas variáveis **stats** e **gcse**. Essa saída é muito similar àquela da correlação de Pearson: uma matriz é mostrada dando o coeficiente de correlação entre duas variáveis (0,455), embaixo está o valor de significância desse coeficiente (0,011) e finalmente o tamanho da amostra (25). O valor da significância

para esse coeficiente de correlação é menor do que 0,05; portanto, pode ser concluído que existe um relacionamento significativo entre os conceitos dos estudantes no GCSE de matemática e o seu conceito no curso de estatística. A própria correlação é positiva: assim, podemos concluir que à medida que os conceitos no GCSE aumentam, existe uma melhoria correspondente nos conceitos de estatística. Dessa maneira, a hipótese foi sustentada. Finalmente, é bom notar que o valor de *N* corresponde ao número de observações que fizemos. Se ele não corresponder, então dados podem ter sido excluídos por alguma razão.

4.5.5 Tau de Kendall (não-paramétrico) ①


A tau de Kendall, τ , é outra correlação não-paramétrica e deve ser usada em vez do coeficiente de Spearman quando você tem um conjunto pequeno de dados com um grande número de postos empatados. Isso significa que se você ordenar todos os escores e muitos deles apresentarem o mesmo posto, o tau de Kendall deve ser usado. Embora a estatística de Spearman seja a mais popular dos dois coeficientes, acreditamos que a estatística de Kendall é realmente uma melhor estimativa da correlação na população (veja Howell, 1997, p. 293). Dessa maneira, podemos estabelecer generalizações mais precisas da estatística de

Saída 4.3

Correlations (Correlações)

			Statistics Grade (Nota em Estatística)	GCSE Maths Grade (Nota GCSE em Matemática)
Spearman's rho (Ro de Spearman)	Statistics Grade (Nota em Estatística)	Correlation Coefficient (Coeficiente de Correlação) Sig. (1 tailed) (Sig. unilateral) N	1.000 . 25	0.455* 0.011 25
	GCSE Maths Grade (Nota GCSE em Matemática)	Correlation Coefficient (Coeficiente de Correlação) Sig. (1 tailed) (Sig. unilateral) N	0.455* 0.011 25	1.000 . 25

*Correlation is significant at the .05 level (1-tailed) (A correlação é significativa ao nível de 0,05 (unilateral))

Kendall do que a de Spearman. Para determinar a correlação de Kendall para os dados das notas de estatística, simplesmente siga os mesmos passos realizados para obter as correlações de Pearson e Spearman, mas selecione  **Kendall's tau-b** e retire a seleção da opção Pearson. A saída é praticamente a mesma da correlação de Spearman.

Você notará da Saída 4.4 do SPSS que o valor real desse coeficiente de correlação é menor do que o de Spearman (diminuiu de 0,455 para 0,354). Apesar da diferença nos coeficientes de correlação, podemos ainda interpretar esse resultado como sendo um relacionamento positivo altamente significativo (porque o valor da significância de 0,015 é menor do que 0,05). Entretanto, o valor do coeficiente de Kendall é uma avaliação mais precisa do que a correlação na população seria. Da mesma forma que com a correlação de Pearson, não podemos assumir que os conceitos no GCSE causaram uma melhoria nos conceitos em estatística.

4.5.6 Correlações bisserial e bisserial por ponto ②



As correlações bisserial e bisserial por ponto apresentam uma diferença conceitual com um cálculo também diferente. Esses coeficientes de correla-

ção são usados quando uma das duas variáveis é dicotômica (isto é, ela é categórica com somente duas categorias). Um exemplo de uma variável dicotômica é a gravidez, porque uma mulher pode ou não estar grávida (ela não pode estar “ligeiramente grávida”). Geralmente, é necessário investigar relacionamentos entre duas variáveis quando uma das variáveis é dicotômica. A diferença entre usar as correlações bisseriais e bisseriais por ponto depende de a variável dicotômica ser discreta ou contínua. Essa diferença é muito sutil. Uma discreta ou dicotômica verdadeira é uma variável para a qual não existe continuidade subjacente entre as categorias. Um exemplo disso é alguém estar vivo ou morto: uma pessoa somente pode estar viva ou morta, ela não pode estar “ligeiramente morta”. Embora você possa descrever uma pessoa como estando “quase morta” – especialmente depois de beber muito – ela ainda está viva se estiver respirando! Portanto, não há continuidade entre as duas categorias. Entretanto, é possível haver dicotomia se existe continuidade. Um exemplo é passar ou reprovar num teste de estatística: algumas pessoas irão simplesmente ser reprovadas enquanto outras irão ser reprovadas por uma margem muito grande; da mesma forma, algumas pessoas irão passar raspando enquanto outras irão gabaritar. Assim, embora os participantes estejam somente entre duas categorias, existe claramente uma continuidade

Saída 4.4

Correlations (Correlações)

			<i>Statistics Grade</i> (Nota em Estatística)	<i>GCSE Maths Grade</i> (Nota GCSE em Matemática)
<i>Kendall's tau_b</i> (Tau_b de Kendall)	<i>Statistics Grade</i> (Nota em Estatística)	<i>Correlation Coefficient</i> (Coeficiente de Correlação) <i>Sig. (1 tailed)</i> (Sig. unilateral) <i>N</i>	1.000 25	0.354* 0.015 25
	<i>GCSE Maths Grade</i> (Nota GCSE em Matemática)	<i>Correlation Coefficient</i> (Coeficiente de Correlação) <i>Sig. (1 tailed)</i> (Sig. unilateral) <i>N</i>	0.354* 0.015 25	1.000 .25

*A correlação é significativa ao nível de 0,05 (unilateral).

subjacente de onde as pessoas estão. Com sorte, está claro que nesse caso existe algum tipo de dicotomia contínua subjacente, porque algumas pessoas passaram ou foram reprovadas mais dramaticamente do que outras. O coeficiente de correlação bisserial por ponto (r_{pb}) é usado quando uma variável é uma dicotomia discreta (por exemplo, gravidez) enquanto o coeficiente de correlação bisserial (r_b) é usado quando uma variável é uma dicotomia contínua (por exemplo, passar ou reprovar numa prova). O coeficiente de correlação bisserial não pode ser calculado diretamente no SPSS: primeiro temos que calcular o coeficiente de correlação bisserial por ponto e, então, usar uma equação para ajustar esse valor.

Imagine que você está interessado no relacionamento entre o gênero de um gato e quanto tempo ele passa longe de casa (o que posso dizer? Eu amo gatos, assim, essas coisas me interessam). Ouvi dizer que gatos machos desaparecem por muito tempo perambulando por longas distâncias na vizinhança (algo sobre hormônios levando-os a encontrar parceiras), enquanto as fêmeas tendem ser mais caseiras. Assim, usei dessa desculpa para visitar muitos dos meus amigos e seus gatos. Anotei o gênero do gato e então pedi ao dono para anotar o número de horas que seu gato ficava ausente de casa por semana. É claro que o tempo gasto longe de casa é medido num nível intervalar – e vamos assumir que ele sa-

tisfaz as outras hipóteses sobre os dados paramétricos – enquanto o gênero do gato é uma dicotomia discreta. Uma correlação bisserial por ponto tem que ser calculada e isso é simplesmente a correlação de Pearson quando a variável dicotômica é codificada com 0 numa categoria e 1 na outra (na verdade, você pode usar qualquer valor e o SPSS irá trocar o mais baixo para 0 e o mais alto para 1 quando fizer os cálculos). Assim, para calcular essas correlações com o SPSS, atribua códigos à variável **gender** (gênero) como foi descrito na Seção 2.4.4 (nos dados já salvos o código 1 é para machos e 0 é para fêmeas). A variável **time** (tempo) simplesmente tem o tempo, em horas, descrito como normal. Esses dados estão no arquivo **pbcorr.sav**. Determine a correlação de Pearson (como na Seção 4.5.1).

A saída do SPSS 4.5 mostra a matriz de correlações de **time** e **gender** (tempo e gênero). O coeficiente de correlação bisserial por ponto é $r_{pb} = 0,378$, o qual tem um valor de significância unilateral de 0,001. O valor da significância dessa correlação é realmente o mesmo que obteríamos se fosse feito um teste t independente para esses dados (veja o Capítulo 7). O sinal da correlação (isto é, se o relacionamento é positivo ou negativo) irá depender inteiramente da maneira que a codificação das variáveis dicotômicas for feita. Para provar isso, nesse caso, o arquivo de dados **pbcorr.sav** tem uma variável extra chamada de

Saída 4.5

Correlations (Correlações)

		Time Away from Home (Hours) (Tempo longe de caso (Horas))	Gender of Cat (Sexo do gato)
Time Away from Home (Hours) (Tempo longe de caso (Horas))	Pearson Correlation (Correlação de Pearson)	1.000	0.378
	Sig. (1 tailed) (Sig. unilateral)	.	0.001
	N	60	60
Gender of Cat (Sexo do gato)	Pearson Correlation (Correlação de Pearson)	0.378*	1.000
	Sig. (1 tailed) (Sig. unilateral)	0.001	.
	N	60	60

*A correlação é significativa ao nível de 0,01 (unilateral).

recode (recodificar) que é a mesma da variável **gender** (gênero), exceto que a codificação é inversa (1 = fêmea, 0 = macho). Se você repetir o cálculo da correlação de Pearson usando **recode** (recodificar) em vez de **gender** (gênero), descobrirá que o coeficiente de correlação agora se torna $-0,378$. O sinal do coeficiente é completamente dependente do código da categoria que você determinar e, assim, temos de ignorar toda a informação sobre a direção do relacionamento. Entretanto, podemos ainda interpretar R^2 como anteriormente. Assim, nesse exemplo, $R^2 = 0,378^2 = 0,143$. Portanto, podemos concluir que gênero contribui por 14,3% da variação do tempo gasto fora de casa.

Imagine agora que queremos converter a correlação bisserial por ponto no coeficiente de correlação bisserial (r_b) (porque alguns dos gatos machos foram castrados e, assim, pode haver um continuum de masculinidade subjacente à variável gênero). Devemos usar a equação (4.4) na qual P_1 é a proporção dos casos que caíram na categoria 1 (o número de gatos machos) e P_2 é a proporção dos casos que caíram na categoria 2 (o número de gatos fêmeas). Nessa equação, y é a ordenada da distribuição normal no ponto onde existe $P_1\%$ da área de um lado e $P_2\%$ no outro (isso ficará mais claro com um exemplo):

$$r_b = \frac{r_{pb} \sqrt{(P_1 P_2)}}{y} \quad (4.4)$$



Para calcular P_1 e P_2 simplesmente usamos o menu **Analyze**⇒**Descriptive Statistics**⇒**Frequencies** (Analisar⇒Estatística descritiva⇒Frequências) e selecionamos a variável **gender** (gênero). Não há necessidade de clicar em mais nenhuma opção, pois os valores por omissão lhe darão tudo o que você precisa saber (isto é, a percentagem dos gatos que são machos ou fêmeas). Isso resulta em 53,3% da amostra são fêmeas (isso é P_2) enquanto os restantes 46,7% são machos (isso é P_1). Para calcular y , usamos esses valores e os valores

da distribuição normal exibidos no Apêndice A.1. Encontre a ordenada onde a curva normal é dividida em 0,467 como a porção menor e 0,533 como a porção maior (na verdade, teremos de usar os valores mais próximos daqueles que são 0,4681 e 0,5319, respectivamente). O valor da ordenada é 0,3977. Se substituirmos esses valores na equação (4.4), obteremos 0,475 (veja abaixo), o que é um valor mais alto do que a correlação bisserial por ponto (0,378). **Portanto, a escolha do coeficiente de correlação pode fazer uma diferença substancial no resultado. Você deve, portanto, ter cuidado ao decidir se sua variável dicotômica tem uma continuidade subjacente ou se ela é uma verdadeira variável discreta.**

$$\begin{aligned} r_b &= \frac{r_{pb} \sqrt{(P_1 P_2)}}{y} \\ &= \frac{(0,378) \sqrt{(0,533 \times 0,467)}}{0,3977} \\ &= 0,475 \end{aligned}$$

4.6 CORRELAÇÃO PARCIAL ②

4.6.1 A teoria por trás da correlação parte e parcial ②



Mencionei anteriormente que existe um tipo de correlação que nos permite ver o relacionamento entre duas variáveis quando o efeito de uma terceira variável é constante. Por exemplo, análises dos dados da ansiedade pré-prova (no arquivo **ExamAnxiety.sav**) mostraram que o desempenho na prova está negativamente relacionado com a ansiedade pré-prova, mas positivamente relacionado ao tempo de revisão e o próprio tempo de revisão estava negativamente relacionado com a ansiedade pré-prova. Esse cenário é complexo, mas dado que sabemos que o tempo de revisão está relacionado a ambos, ansiedade pré-prova e desempenho na prova, então, se queremos uma medida pura do relacionamento entre ansiedade pré-prova e desempenho na prova, devemos levar em

conta a influência do tempo de revisão. Usando os valores de R^2 para esses relacionamentos, sabemos que a ansiedade pré-prova contribui com 19,4% da variância do desempenho na prova, que o tempo de revisão contribui com 15,7% da variância no desempenho na prova e que o tempo de revisão contribui com 50,2% da variância na ansiedade pré-prova. Se o tempo de revisão contribui com metade da variância na ansiedade pré-prova, parece viável que pelo menos alguns dos 19,4% da variância no desempenho na prova debitada à ansiedade é a mesma variância que contribui no tempo de revisão. Como tal, alguma variância no desempenho da prova explicada pela ansiedade pré-prova não é *única* e pode ser atribuída ao tempo de revisão. **Uma correlação entre duas variáveis em que os efeitos de outras variáveis são constantes é conhecida como uma correlação parcial.**

A Figura 4.17 ilustra o princípio por trás da correlação parcial. Na parte 1 do diagrama há uma caixa denominada *Exam Performance* (Desempenho na prova) que representa a variação total nos escores da prova (esse valor seria a variância do desempenho na prova). Existe, também, uma caixa que representa uma variação na ansiedade pré-prova (novamente, essa é a variância daquela variável). Já sabemos que a ansiedade pré-prova e o desempenho na prova compartilham 19,4% da sua variação (esse valor é o coeficiente de correlação ao quadrado). Portanto, as variações dessas duas variáveis se sobrepõem (porque elas compartilham a variância) criando uma terceira caixa (aquela com linhas diagonais). A sobreposição dessas caixas representando desempenho na prova e ansiedade pré-prova é a variância comum. Da mesma forma, na parte 2 do diagrama a variância compartilhada entre o desempenho na prova e tempo de revisão está ilustrada. O tempo de revisão compartilha 15,7% da variação nos escores na prova. Essa variação compartilhada é representada pela área de sobreposição (preenchida com linhas diagonais). Sabemos que o tempo de revisão e a ansiedade pré-prova têm em comum 50% das suas variações, portanto, é muito provável

que alguma variação no desempenho na prova compartilhada pela ansiedade pré-prova sejam as mesmas das variâncias compartilhadas pelo tempo de revisão.

A parte 3 do diagrama mostra a figura completa. A primeira coisa a ser notada é que os quadros (ou caixas) representando ansiedade pré-prova e o tempo de revisão têm uma grande sobreposição (eles dividem 50% da variação). Mais importante, quando olhamos quanto o tempo de revisão e a ansiedade contribuem para o desempenho na prova, vemos que existe uma porção do desempenho na prova que é compartilhada por ambos, ansiedade e tempo de revisão (a área pontilhada). Entretanto, ainda existem pequenas partes da variância no desempenho na prova que são únicas para as outras duas variáveis. Assim, embora na parte 1 a ansiedade pré-prova compartilhe um pedaço grande da variação com o desempenho na prova, alguma dessa sobreposição é também compartilhada com o tempo de revisão. Se removermos o pedaço da variância que é também dividida pelo tempo de revisão, teremos uma medida do relacionamento único entre desempenho na prova e ansiedade pré-prova. **Nós usamos correlações parciais para descobrir o tamanho do pedaço único da variância. Portanto, podemos determinar uma correlação parcial entre ansiedade pré-prova e desempenho na prova enquanto “controlamos” o efeito do tempo de revisão.** Da mesma forma, podemos determinar uma correlação parcial entre o tempo de revisão e o desempenho na prova “controlando” os efeitos da ansiedade pré-prova.

4.6.2 Correlação parcial utilizando o SPSS ②

Para determinar uma correlação parcial com os dados do desempenho na prova, selecione a opção *Correlate* (Correlacionar) do menu *Analyze* (Analisar) e, então, selecione *Partial* (Parcial) (**Analyze⇒Correlate⇒Partial**) (Analisar⇒Correlacionar⇒Parcial...) e uma caixa de diálogo semelhante a da Figura 4.18 será ativada. Essa caixa de diálogo lista todas

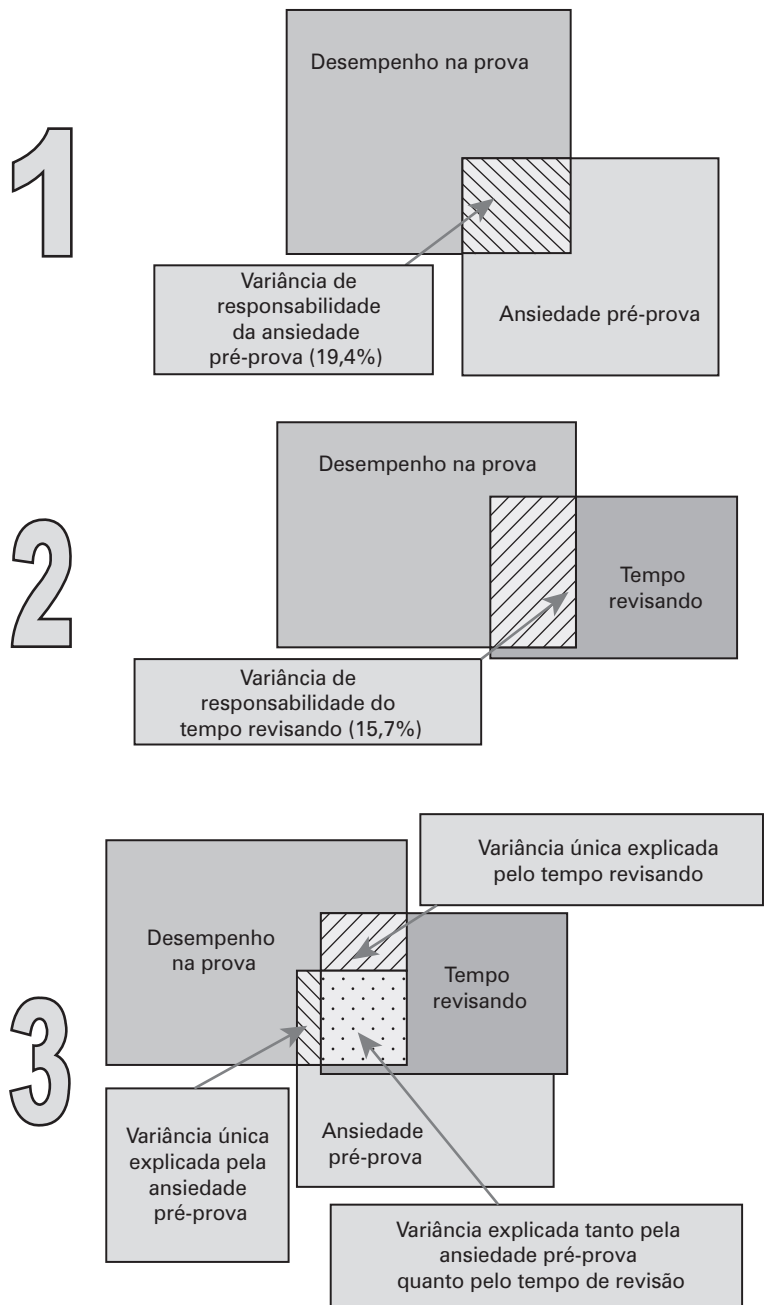


Figura 4.17 Diagrama mostrando os princípios da correlação parcial.

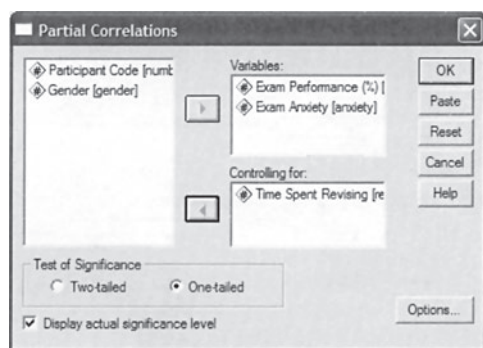


Figura 4.18 Caixa de diálogo principal para determinação de uma correlação parcial.

as variáveis do editor de dados no lado esquerdo deixando dois espaços vazios no lado direito. O primeiro espaço é para listar as variáveis que você quer correlacionar e o segundo é para indicar quais variáveis você quer controlar. No exemplo que descrevi, queremos determinar o efeito único da ansiedade pré-prova no desempenho na prova e, assim, queremos correlacionar as variáveis **exam** (prova) e **anxiety** (ansiedade), enquanto controlamos o tempo de revisão. A Figura 4.18 mostra a caixa de diálogos completa. Se você clicar em **Options...** (Opções) outra caixa de diálogo aparece como mostra a Figura 4.19.

Essas opções são similares às aquelas da correlação bivariada exceto que você pode escolher mostrar correlações de ordem zero. Correlações de ordem zero são os coeficientes de correlação bivariada sem controle de quaisquer outras variáveis (isto é, são o coeficiente de correlação de Pearson). Assim, em nosso exemplo,

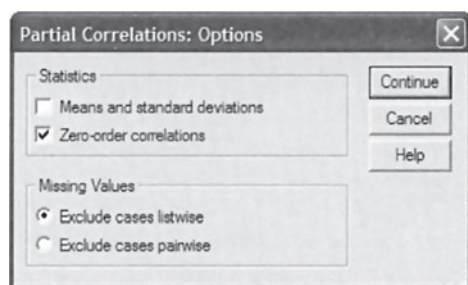


Figura 4.19 Opções para correlação parcial.

se selecionarmos as correlações de ordem zero, o SPSS irá produzir uma matriz de correlações entre **anxiety**, **exam**, e **revise** (ansiedade, prova e revisão). Se você não determinou as correlações bivariadas antes das correlações parciais, essa é uma maneira útil de comparar correlações que não foram controladas com as que foram. Essa comparação dá a você uma percepção da contribuição das diferentes variáveis. Marque a caixa das correlações de ordem zero, mas deixe o restante das opções como estão.

A saída 4.6 do SPSS mostra os resultados para a correlação parcial da ansiedade pré-prova e o desempenho na prova controlada pela variável tempo de revisão. A primeira coisa a notar é a matriz das correlações de ordem zero que foi obtida porque utilizamos a caixa de diálogos *options* (opções). As correlações mostradas aqui são idênticas às aquelas obtidas com o procedimento da correlação de Pearson (compare essa matriz com a da saída 4.2 do SPSS). Abaixo das correlações de ordem zero está a matriz de correlações para as variáveis **anxiety** e **exam** (ansiedade e prova), mas controlada pela variável revisão. **Por ora controlamos uma única variável, assim, obtivemos o que é conhecido como uma correlação parcial de primeira ordem.** É possível controlar os efeitos de duas variáveis ao mesmo tempo (uma correlação parcial de segunda ordem) ou controlar três variáveis (uma correlação parcial de terceira ordem) e assim por diante. Em primeiro lugar, note que a correlação parcial entre desempenho na prova e ansiedade pré-prova é de $-0,2467$, que é consideravelmente menor do que a correlação quando o tempo de revisão não é controlado por ($r = -0,4410$). Na verdade, o coeficiente de correlação é aproximadamente menor do que ele era antes. Embora essa correlação seja estatisticamente significativa (seu valor p está abaixo de $0,05$) o relacionamento diminuiu. Em termos de variância, o valor de R^2 para a correlação parcial é de $0,06$, o que significa que a ansiedade pré-prova pode agora contribuir com somente 6% da variância no desempenho na prova. Quando os efeitos do tempo de revisão não eram controlados, a ansiedade pré-prova compartilhava 19,4% da

Saída 4.6 Saída para uma correlação parcial

PARTIAL CORRELATION COEFFICIENTS (COEFICIENTES DE CORRELAÇÃO PARCIAIS)

Zero Order Partial (Parciais de ordem zero)

	EXAME (Prova)	ANXIETY (Ansiedade)	REVISE (Revisar)
EXAME (Prova)	1.0000 (0)	-.4410 (101)	.3967 (101)
	P = .	P = .000	P = .000
ANXIETY (Ansiedade)	-.4410 (101)	1.0000 (0)	-.7092 (101)
	P = .000	P = .	P = .000
REVISE (Revisar)	.3967 (101)	-.7092 (101)	1.0000 (0)
	P = .000	P = .000	P = .

(Coeficientes / D.F.) / 1-tailed significance)

(Coeficientes / G.L. / Significância unilateral)

PARTIAL CORRELATION COEFFICIENTS (COEFICIENTES DE CORRELAÇÃO PARCIAIS)

Controlling for .. (Controle por ..) REVISE(REVISAR)

	EXAME (Prova)	ANXIETY (Ansiedade)
EXAME (Prova)	1.0000 (0)	-.2467 (100)
	P = .	P = .006
ANXIETY (Ansiedade)	-.2467 (100)	1.0000 (0)
	P = .006	P = .

(Coeficientes / D.F.) / 1-tailed significance)

(Coeficientes / G.L. / Significância unilateral)

". "is printed if a coefficient cannot be computed

(Se um coeficiente não puder ser calculado então ". "é impresso)

variância com os escores da prova e, assim, a inclusão do tempo de revisão diminuiu o montante da variação nos escores da prova compartilhados com a ansiedade. Dessa maneira, uma

medida mais verdadeira do papel da ansiedade pré-prova foi obtida. A execução dessas análises tem mostrado que apenas a ansiedade pré-prova não explica muito da variância nos

escores da prova e descobrimos um relacionamento complexo entre ansiedade e revisão que poderia de outra maneira ter sido ignorado. Embora a causalidade ainda não seja certa, porque variáveis relevantes foram incluídas, o problema da terceira variável é, pelo menos, tratado de alguma forma.

Essas correlações parciais podem ser feitas quando as variáveis são dicotômicas. Assim, usando um exemplo anterior, procuramos um relacionamento entre o tamanho do cérebro e o gênero (que é dicotômico) controlado pelo tamanho do corpo. É, também, bom que a “terceira” variável seja dicotômica, assim, por exemplo, podemos determinar o relacionamento entre uma bexiga frouxa e um grande número de tarântulas subindo nas suas pernas (ambas variáveis contínuas) controladas pelo gênero (uma variável dicotômica).⁴

4.6.3 Correlações semiparciais (ou por parte) ②



No próximo capítulo descobriremos outra forma de correlação conhecida como correlação semiparcial (também chamada de correlação por parte). É válido explicar as diferenças entre esse tipo de correlação e a correlação semiparcial. **Quando fazemos a correlação parcial**

⁴ Esses dois exemplos são, de fato, casos simples de regressão hierárquica (veja o próximo capítulo) e o primeiro é, também, um exemplo de análise de covariância. Isso pode parecer confuso agora, mas, à medida que progredirmos no livro, espero que fique claro que todas as estatísticas que você utiliza são, de fato, a mesma coisa, mas com roupas e nomes diferentes.

entre duas variáveis, controlamos o efeito da terceira variável. Especificamente, o efeito que a terceira variável tem em ambas as variáveis na correlação é controlada. Numa correlação semiparcial, controlamos o efeito que a terceira variável tem em somente uma das variáveis envolvidas na correlação. A

Figura 4.20 ilustra esse princípio para os dados do desempenho na prova. A correlação parcial que calculamos leva em consideração não somente o efeito da revisão no desempenho na prova, mas também o efeito da revisão na variável ansiedade. Se fôssemos calcular a correlação semiparcial para os mesmos dados, teríamos controle apenas sobre o efeito da revisão no desempenho na prova (o efeito da revisão na ansiedade pré-prova seria ignorado). Correlações parciais são mais úteis para observar o relacionamento único entre duas variáveis onde outras variáveis foram descartadas. Correlações semiparciais são, portanto, úteis quando tentamos explicar a variância numa variável particular (uma saída) a partir de um conjunto de variáveis predictoras. Esse será assunto do Capítulo 5.

4.7 COMO RELATAR COEFICIENTES DE CORRELAÇÃO ①

Relatar coeficientes de correlação é muito fácil: você apenas deve dizer quão grande eles são e quais são os valores das significâncias (embora o valor de significância não seja tão importante porque o coeficiente de correlação é ele próprio um tamanho de efeito!). As quatro coisas a serem observadas são: (1) não deve ter zero na frente do ponto decimal

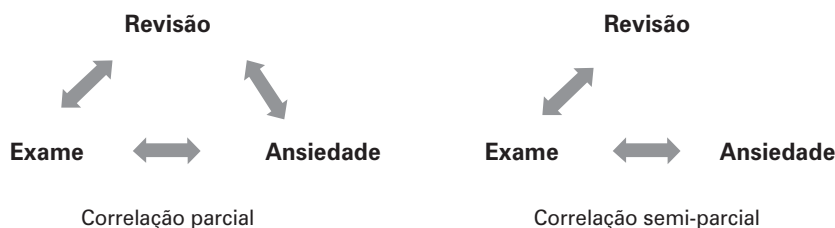


Figura 4.20 A diferença entre uma correlação parcial e uma semiparcial.

para o coeficiente de correlação ou valor de probabilidade (porque nenhum deles deve ser maior que 1)*; (2) se você estiver citando uma probabilidade unilateral, deve dizê-lo; (3) cada coeficiente de correlação é representado por uma letra diferente (e algumas delas são gregas!); (4) existe um critério padrão de probabilidades que usamos (0,05, 0,01 e 0,001). Vamos pegar alguns exemplos deste capítulo:

- ✓ Havia um relacionamento significativo entre o número de comerciais assistidos e os pacotes de balas comprados, $r = 0,87$, p (unilateral) $< 0,05$.
- ✓ O desempenho na prova estava correlacionado significativamente com a ansiedade pré-prova $r = -0,44$ e o tempo gasto revisando, $r = 0,40$; o tempo gasto revisando também foi correlacionado com a ansiedade pré-prova, $r = -0,71$ (todos os $ps < 0,001$).
- ✓ Havia um relacionamento positivo entre os conceitos estatísticos de uma pessoa e seu conceito no GCSE de matemática, $r_s = 0,46$, $p < 0,05$.
- ✓ Havia um relacionamento positivo entre os conceitos estatísticos de uma pessoa e seu conceito no GCSE de matemática, $\tau = 0,35$, $p < 0,05$. (Observe que eu usei o tau de Kendall aqui.)
- ✓ O gênero do gato estava correlacionado significativamente com o tempo que o gato passava fora de casa, $r_{pb} = 0,38$, $p < 0,01$.

Um ponto final a ser observado é que nas ciências sociais nós temos vários níveis *padrão* de significância estatística. O critério mais importante é que o valor da significância é menor do que 0,05; entretanto, o valor exato da significância é muito mais baixo, então podemos ter mais certeza do poder do efeito experimental. Nessas circunstâncias, ficamos contentes que o nosso resultado não seja apenas significativo em 0,05, mas também significativo a um nível muito mais baixo (eba!). Os valores que usa-

mos são 0,05, 0,01, 0,001 e 0,0001. Você raramente será capaz de relatar um efeito que seja significativo a um nível menor do que 0,0001!

4.8 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Este capítulo abordou diferentes maneiras de estudar os relacionamentos entre variáveis. Iniciamos verificando como medir estatisticamente os relacionamentos desenvolvendo o que já sabemos sobre variância (do Capítulo 1) e depois aprendemos a olhar a variância compartilhada entre variáveis. Essa variância compartilhada é conhecida como *covariância*. Após, vimos como representar graficamente os relacionamentos usando o *diagrama de dispersão*. Na sua forma mais simples, esses gráficos traçam valores de uma variável contra outra. Entretanto, você pode traçar mais do que duas variáveis num diagrama de dispersão em 3D ou sobrepor um diagrama de dispersão. É possível também distinguir grupos de pessoas num diagrama de dispersão usando símbolos diferentes e cores. Descobrimos, então, que quando os dados são paramétricos, podemos mensurar a força do relacionamento usando o coeficiente de correlação r de Pearson. Quando os dados violam as suposições dos testes paramétricos podemos usar o r_s de Spearman ou, para pequenos conjuntos de dados, o τ de Kendall pode ser mais preciso. Vimos, também, que as correlações podem ser calculadas entre duas variáveis quando uma delas é dicotômica (isto é, formada por apenas duas categorias); quando as categorias não têm uma continuidade subjacente, usamos a correlação bisserial por ponto r_{pb} , mas quando as categorias têm uma continuidade subjacente, utilizamos a correlação bisserial r_b . Finalmente, verificamos as diferenças entre *correlações parciais*, nas quais o relacionamento entre duas variáveis é medido controlando o efeito que uma ou mais variáveis tem em ambas as variáveis, e *correlações semiparciais*, nas quais o relacionamento entre duas variáveis é medido controlado pelo efeito que uma ou mais variáveis tem em somente uma das variáveis.

* N. de T.: Esse tipo de diferença não é feita no Brasil. Todos os valores, independentemente de poderem ou não ultrapassar um, são representados com o zero inicial.

4.9 TERMOS-CHAVE QUE DESCOBRIMOS

- Correlação bisserial
- Correlação bivariada
- Coeficiente de determinação
- Covariância
- Desvios dos produtos cruzados
- Tau de Kendall
- Padronização
- Correlação parcial
- Coeficiente de correlação de Pearson
- Correlação bisserial por ponto
- Diagrama de dispersão
- Correlação semiparcial
- Coeficiente de correlação de Spearman

4.10 TAREFAS DO ALEX ESPERTO

As respostas para essas tarefas podem ser encontradas no *site* www.artmed.com.br no arquivo **Answers-(Chapter 4).pdf**.



- Um estudante estava interessado em saber se havia um relacionamento positivo entre o tempo gasto escrevendo um artigo e a nota recebida. Ele pegou 45 dos seus amigos e contou quantas horas eles gastavam escrevendo seu artigo (**hours**) e a percentagem que eles tiveram

no artigo (**essay**). Ele também traduziu essas notas na sua classificação de notas (**grade**): primeira, segunda superior, segunda inferior e terceira classe. Usando os dados no arquivo **EssayMarks.sav**, descubra qual foi o relacionamento entre o tempo gasto escrevendo o artigo e a nota final em termos de percentagem e classificação das notas (trace um diagrama de dispersão também!). ①

- Usando os dados do **ChickFick.sav** do Capítulo 3, existe um relacionamento entre gênero e excitação? Usando os mesmos dados, existe um relacionamento entre o filme assistido e excitação? ①

4.10 LEITURAS COMPLEMENTARES

HOWELL, D. C. *Statistical Methods for Psychology*. Belmont (CA): Duxbury, 2002 (5ª edição). O Capítulo 9 fornece uma cobertura mais detalhada sobre correlação e regressão do que Wright, mas a leitura não é tão fácil. O Capítulo 10 é ótimo para correlação bisserial e bisserial por ponto.

WRIGHT, D. B. *First Steps in Statistics*. Londres: Sage, 2002. Esse livro apresenta uma introdução muito clara aos conceitos de correlação e regressão (Capítulo 8).

5

REGRESSÃO

5.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

No último capítulo vimos como medir (e traçar) relacionamentos entre duas variáveis. Podemos avançar um passo nesse processo e aprender como prever uma das variáveis em função da outra. Um exemplo simples é o de tentar prever os níveis de estresse a partir do tempo que falta até a ocorrência de uma palestra. Você esperaria que esse fosse um relacionamento negativo (quanto menor o tempo até a palestra, maior a ansiedade). Podemos estender esse relacionamento básico para responder a questões como “se faltam 10 minutos até que alguém precise falar a um público, quão ansioso ele estará?”. **Essa é a essência da análise de regressão: é uma forma de prever algum tipo de saída (resultado) a partir de uma ou mais variáveis previsoras. Este capítulo aborda a análise de regressão** com profundidade. Começo apresentando alguns princípios básicos utilizando o exemplo de prever um resultado a partir de um único previsor (**regressão simples**). Em seguida, explicarei o que é um modelo, como estimá-lo, como podemos determinar o quão bem ele se ajusta aos dados, como tudo isso é feito no SPSS e como as saídas do SPSS são interpre-

tadas. Depois, analiso o caso mais complexo de quando um resultado é previsto a partir de diversas variáveis previsoras (**regressão múltipla**). Embora eu não aborde como o modelo é estimado, irei descrevê-lo e especificar vários métodos que selecionaram as variáveis que farão parte do modelo. Veremos, ainda, algumas coisas que podem afetar a precisão do modelo e como tratá-las antes de realizar a análise no SPSS, interpretar a saída e descobrir quão bem o modelo se ajusta aos dados e como ele pode ser generalizado. Se você ainda está acordado, o capítulo termina verificando como podemos incorporar variáveis categóricas na análise de regressão utilizando algo chamado de variáveis auxiliares (*dummy*). Ufa!

5.2 UMA INTRODUÇÃO À REGRESSÃO ①

A correlação pode ser uma ferramenta de pesquisa bastante útil, mas ela nada nos informa sobre o poder preditivo das variáveis. Na análise de regressão, ajustamos um modelo preditivo aos nossos dados e então usamos esse modelo para prever valores da variável dependente (VD) a partir de uma ou mais variáveis

independentes (VIs).¹ A regressão simples procura prever uma **variável de saída** a partir de um única variável previsoras, enquanto que a regressão múltipla busca prever um resultado a partir de diversas **variáveis previsoras**. Essa é uma ferramenta bastante útil porque nos permite ir um passo além dos dados que de fato temos. Na Seção 1.4.11, apresentarei a ideia de que podemos prever qualquer valor utilizando a seguinte equação genérica:

$$\text{Saída}_i = (\text{Modelo}_i) + \text{Erro}_i \quad (5.1)$$



Isso significa apenas que a saída ou resposta que estamos tentando prever para uma determinada pessoa pode ser prevista por qualquer modelo que ajustarmos aos dados mais algum tipo de erro. Na regressão, o modelo que ajustamos é linear (“modelo linear” significa “modelo baseado em uma linha reta”) e você pode imaginar que está tentando resumir um conjunto de dados com uma linha reta (lembra a Figura 1.11). Assim, a palavra “modelo” na equação acima pode ser substituída por algo que defina a linha que ajustamos aos dados (veja a próxima seção).

Com qualquer conjunto de dados existem várias linhas que podem ser utilizadas para resumir a tendência geral e, desse modo, precisamos encontrar uma forma de decidir entre as muitas possibilidades. Para que seja possível tirar conclusões precisas, é necessário ajustar o *melhor* modelo que descreve os dados. Existem muitas maneiras de ajustar uma linha

reta aos dados que foram coletados. A mais simples é utilizar seus olhos para imaginar uma linha que pareça resumir bem os dados. Entretanto, esse método é bastante subjetivo e não assegura que o modelo é o melhor que se poderia ter escolhido. Portanto, vamos utilizar técnicas matemáticas para determinar a linha que melhor descreve os dados observados. Esse método é denominado *método dos mínimos quadrados*.

5.2.1 Algumas informações importantes sobre retas ①

Mencionei acima que a palavra modelo na equação geral seria substituída por “alguma coisa que define a linha que se ajusta aos dados”. De fato, qualquer linha reta pode ser definida por duas coisas: (1) a inclinação (ou gradiente) da linha (normalmente representado por b_1), e (2) o ponto em que a linha cruza o eixo vertical do gráfico (conhecido como *intercepto* da linha b_0). De fato, nosso modelo geral se torna a equação (5.2) na qual Y_i é a variável de saída que queremos prever e X_i é o escore do i -ésimo participante da variável previsoras.² O gradiente da linha reta ajustada aos dados é b_1 e b_0 é o intercepto da linha. Os parâmetros b_1 e b_0 são conhecidos como *coeficientes de regressão* e aparecerão de vez em quando no livro, onde você irá vê-los normalmente representados como b (sem qualquer subscrito) ou b_i (que significa b associado à variável i). Existe um termo resíduo, ε_i , que representa a diferença entre o valor previsto pela linha para o participante i e o escore que o participante i realmente obteve. A equação é muitas vezes apresentada sem esse termo resíduo (assim, ignore-o se ele estiver lhe incomo-

¹ Infelizmente, você encontrará pessoas (e o SPSS também) se referindo às variáveis na regressão como dependentes e independentes (como em experimentos controlados). No entanto, a pesquisa correlacional por sua natureza raramente controla as variáveis independentes para medir o efeito na variável dependente. Em vez disso, as variáveis são mensuradas simultaneamente e sem um controle restrito. Assim, não é correto falar das variáveis de regressão dessa forma. Por esse motivo, denomino as “variáveis independentes” como *previsores* e a “variável dependente” como *saída* ou *resultado*.

² Às vezes, você irá ver essa equação escrita como:

$$Y_i = (\beta_0 + \beta_1 X_i) + \varepsilon_i$$

A única diferença é que nessa equação aparece β em vez de b e, de fato, as duas versões são a mesma coisa – elas apenas utilizam letras diferentes para representar os coeficientes. Eu prefiro a notação- β (e a utilizei na edição anterior); no entanto, quando o SPSS estima os coeficientes nessa equação, ele as rotula como b . Para ser consistente com a saída do SPSS que você verá, utilizarei b em vez de β .

dando); no entanto, vale a pena saber que esse termo significa que o nosso modelo não irá se ajustar perfeitamente aos dados coletados:

$$Y_i = (b_0 + b_1X_i) + \varepsilon_i \quad (5.2)$$

Uma linha particular tem um gradiente e um intercepto específicos. A Figura 5.1 mostra um conjunto de linhas que apresentam o mesmo intercepto, mas diferentes gradientes e outro conjunto que tem o mesmo gradiente (inclinação), mas diferentes interceptos. A Figura 5.1 também ilustra outro ponto importante: o de que o gradiente da linha nos informa algo sobre a natureza do relacionamento sendo descrito. No Capítulo 4, vimos que uma relação pode ser tanto positiva quanto negativa (e com isso não quero dizer que é a diferença entre se dar bem com a sua namorada ou discutir o tempo todo!). Uma linha que apresenta um gradiente positivo descreve um relacionamento positivo, ao passo que uma com um gradiente negativo descreve um relacionamento negativo. Assim, se você olhar para os gráficos na Figura 5.1 em que os gradientes diferem, mas os interceptos são os mesmos, notará que a linha mais grossa descreve um relacionamento positivo e a mais fina, um negativo.

Se é possível descrever uma linha conhecendo somente o gradiente (inclinação) e o intercepto dessa linha, então podemos utilizar esses valores para descrever o nosso modelo (porque no modelo de regressão linear utilizamos uma linha reta). Assim, é possível

imaginar o modelo que ajustamos aos nossos dados na regressão linear como uma linha reta que pode ser descrita matematicamente pela equação (5.2). Com a regressão lutamos para encontrar a linha que melhor descreve os dados coletados, então estimamos o gradiente e o intercepto de tal linha. Tendo definido esses valores, podemos inserir diferentes valores na nossa variável previsora no modelo para estimar os valores da variável resposta (de saída).

5.2.2 O método dos mínimos quadrados ①

Já mencionei que o método dos mínimos quadrados é uma forma de encontrar a linha que melhor se ajusta aos dados (isto é, encontrar a linha que passe entre – ou o mais próximo possível – o maior número de pontos de dados). Essa “linha de melhor ajuste ou aderência” é encontrada ao verificarmos qual linha é a melhor de todas as possíveis que poderiam ser traçadas e resulta na menor diferença entre os pontos observados e a linha. A Figura 5.2 mostra que quando qualquer linha é ajustada a um conjunto de dados, existirá uma pequena diferença entre os valores previstos pela linha e os dados que foram efetivamente observados. Estamos interessados na diferença vertical entre a linha e os dados reais porque estamos utilizando a linha para prever valores de uma variável Y a partir de uma variável X . Embora alguns pontos de dados estejam exatamente sobre a linha, outros estão acima ou abaixo,

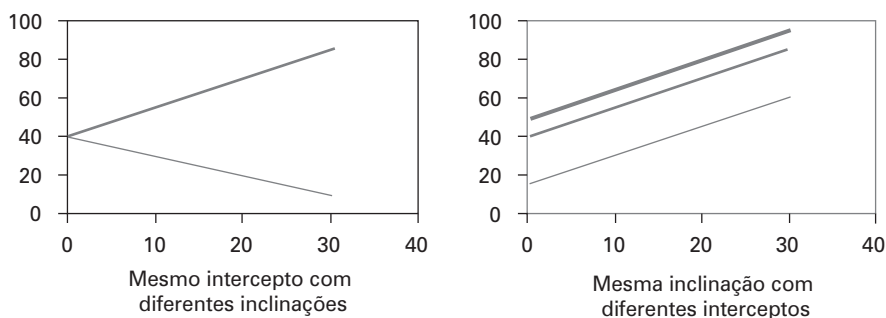


Figura 5.1 Linhas com os mesmos gradientes, mas com diferentes interceptos e linhas com o mesmo intercepto, mas com gradientes (inclinações) diferentes.

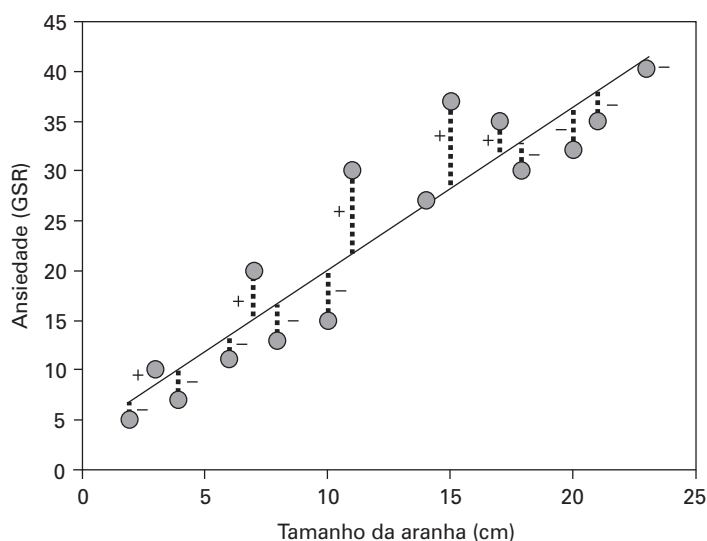


Figura 5.2 Esse gráfico mostra um diagrama de dispersão de alguns dados com uma linha de tendência entre eles. As linhas verticais (pontilhadas) representam as diferenças (resíduos) entre a tendência e os dados.

indicando que existe uma diferença entre o modelo que se ajusta aos dados e os dados coletados. Algumas dessas diferenças são positivas (os pontos estão acima da linha, indicando que o modelo está subestimando seus valores) e algumas são negativas (os pontos estão abaixo da linha, indicando que o modelo está superestimando seus valores). Essas diferenças são geralmente denominadas **resíduos**. Na discussão sobre variância na Seção 1.4.1, expliquei que se somarmos as diferenças positivas e negativas, elas tendem a se cancelar resultado numa soma zero. Para evitar esse problema, elevamos as diferenças ao quadrado antes de somá-las. As diferenças ao quadrado fornecem uma medida de quão bem uma determinada linha se ajusta aos dados: se as diferenças ao quadrado são grandes, a linha não é representativa dos dados; se as diferenças ao quadrado são pequenas, a linha é representativa. A soma das diferenças ao quadrado (ou **SS** abreviado) pode ser calculada para qualquer linha que é ajustada a dados; a qualidade de ajuste de cada linha aos dados pode ser avaliada pela soma das diferenças ao quadrado. O método dos mínimos quadrados seleciona a linha que

apresenta a menor soma das diferenças ao quadrado (assim, ele escolhe a linha que melhor representa os dados observados). Uma forma de selecionar essa linha ótima seria ajustar cada linha possível ao conjunto de dados, calcular a soma das diferenças ao quadrado para cada uma e então escolher a linha que apresentasse o menor valor para essa soma. Você certamente gastaria um bom tempo fazendo isso! Felizmente, existe uma técnica matemática para encontrar máximos e mínimos e essa técnica (cálculo) é utilizada para encontrar a linha que minimiza a soma das diferenças ao quadrado. O resultado final é que os valores do intercepto e da inclinação da “linha de melhor aderência” podem ser estimados. Cientistas sociais geralmente se referem a essa linha de melhor ajuste como linha de regressão.

5.2.3 Avaliando o ajuste: soma dos quadrados, R e R^2 ①

Uma vez que encontramos a linha de melhor ajuste, é importante que avaliemos o quão bem ela adere aos dados (ou seja, a **qualidade do ajuste** do modelo). Fazemos isso porque

Como posso
saber se o meu
modelo é bom?



embora essa linha seja a melhor disponível, ela ainda poderá ter um péssimo ajuste aos dados! Na Seção 1.4.1, vimos que uma medida de adequação de um modelo é a soma dos quadrados das diferenças (ou mais comumente ava-

liamos o modelo utilizando a equação (5.3)). Se quisermos avaliar a linha de melhor aderência, precisamos compará-la com algo e o que escolhemos é o modelo mais básico que podemos ter. Assim, vamos utilizar a equação (5.3) para calcular o ajuste do modelo mais básico e então o ajuste do melhor modelo (a linha de melhor aderência); se o melhor modelo for bom, ele deve aderir aos dados bem melhor de que o modelo básico. Isso é um tanto abstrato, portanto, vamos ver um exemplo:

$$\text{Desvios} = \sum (\text{observados} - \text{modelo})^2 \quad (5.3)$$

Imagine que eu quero prever a quantidade de discos vendidos (Y) a partir do investimento feito em publicidade (X). Um dia meu chefe vai até o meu escritório e diz “Andy, quantos discos iremos vender se gastarmos £100 000 em publicidade?”. Se eu não tivesse um bom modelo do relacionamento entre a venda de discos e publicidade, qual seria o meu melhor palpite? Bem, provavelmente a melhor resposta que eu poderia dar seria o número médio de discos vendidos (digamos, 200.000) porque na média é o quanto de discos esperamos vender. Essa resposta talvez satisfaça um executivo fonográfico sem cérebro. No entanto, o que você diria se ele tivesse perguntado: “Quantos discos seriam vendidos se gastássemos £1 em publicidade?”. Novamente, na ausência de uma informação mais precisa, o melhor palpite seria fornecer o número médio de discos vendidos (200.000). Aqui há um problema: não importa a quantidade investida em publicidade, sempre vou prever a mesma quantidade de vendas. Deve ficar bem claro, então, que a média é praticamente inútil como modelo de relacionamento entre duas variáveis, mas ele é o modelo mais simples disponível.

Assim, como estratégia básica para prever um resultado, podemos considerar a escolha da média, porque em média ela será um bom palpite para um determinado resultado. Utilizando a média como um modelo, podemos calcular diferenças entre os valores observados e os valores previstos pela média (equação (5.3)). Vimos na Seção 1.4.1 que elevamos todas essas diferenças ao quadrado para obtermos a soma das diferenças ao quadrado. Essa soma é conhecida como **soma total dos quadrados** (representada por SS_T) porque ela é a quantia total de diferenças presentes quando o modelo mais básico é aplicado aos dados. Esse valor representa o quão boa a média é como um modelo para os dados observados. Agora se ajustarmos um modelo mais sofisticado aos dados, como a linha de melhor aderência, podemos novamente obter as diferenças entre esse novo modelo e os dados observados, utilizando mais uma vez a equação (5.3). Na seção anterior vimos que o método dos mínimos quadrados encontra a melhor linha possível para descrever o conjunto de dados minimizando as diferenças entre o modelo ajustado aos dados e os próprios dados. Entretanto, mesmo com esse modelo ótimo ainda vai existir alguma falta de aderência que será representada pelas diferenças entre cada ponto de dados observado e o valor previsto pela linha de regressão. Como antes, essas diferenças são elevadas ao quadrado antes de serem somadas de modo que as direções das diferenças não se cancelem. O resultado é conhecido como **soma dos resíduos ao quadrado** ou **soma dos quadrados residual** (SS_R). Esse valor representa o grau de imprecisão quando o melhor modelo é ajustado aos dados. Podemos utilizar esses dois valores para comparar o quanto melhor é a linha de regressão (linha do melhor ajuste) que o valor que seria obtido utilizando a média como modelo (isto é, quão melhor é o nosso modelo ótimo quando confrontado com o pior modelo). A melhoria na previsão resultante da utilização da linha de regressão em vez da média é obtida calculando a diferença entre SS_T e SS_R . Essa diferença nos mostra a redução na imprecisão do modelo resultante do ajuste do

modelo da linha de regressão aos dados. Essa melhoria é a **soma dos quadrados do modelo** (SS_M). A Figura 5.3 mostra cada soma dos quadrados graficamente.

Se o valor do SS_M é alto, usar o modelo de regressão é bem diferente do que utilizar a média para prever o valor da variável resultante. Isso implica que o modelo de regressão apresenta uma grande melhoria em quão bem os valores resultantes podem ser previstos. No entanto, se o SS_M é pequeno, então utilizar o modelo de regressão é apenas um pouco

melhor do que usar a média (ou seja, o modelo de regressão não é muito melhor do que o nosso melhor palpite). Uma medida útil que se pode obter das somas dos quadrados é a proporção de melhoria debitada ao nosso modelo. Isso é facilmente calculado dividindo a soma dos quadrados para o modelo pela soma total dos quadrados. O valor obtido é denominado R^2 e ele é um percentual. Então, R^2 representa a quantidade de variância nas saídas explicadas pelo modelo (SS_M) relativa a quanta variação foi inicialmente explicada (SS_T).

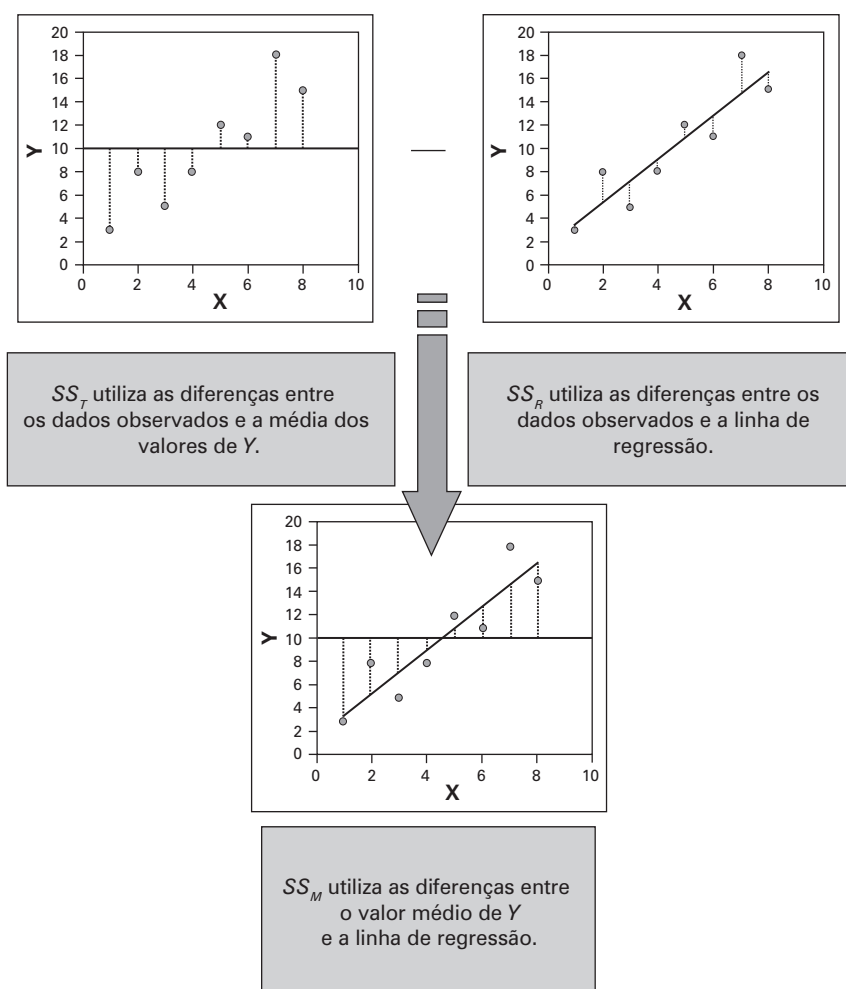


Figura 5.3 Diagrama mostrando de onde é derivada a soma dos quadrados da regressão.

Dessa forma, ele representa o percentual de variação nos valores previstos que podem ser explicados pelo modelo:

$$R^2 = \frac{SS_M}{SS_T} \quad (5.4)$$

Não por coincidência, esse valor é o mesmo que encontramos para R^2 no Capítulo 4 (Seção 4.5.3) e você notará que ele é interpretado da mesma forma. Portanto, na regressão simples podemos tomar a raiz quadrada do seu valor para obter o coeficiente de correlação de Pearson. Como tal, o coeficiente de correlação nos fornece uma boa estimativa da aderência do modelo de regressão e R^2 nos dá uma boa medida do valor desse relacionamento.

Uma segunda utilização das somas dos quadrados para avaliar o modelo é por meio do teste F. Abordaremos esse teste em profundidade no Capítulo 8, mas, brevemente, o teste F tem por base a razão de melhoria devida ao modelo (SS_M) e a diferença entre o modelo e os dados observados (SS_R). De fato, em vez de utilizar a soma dos quadrados, tomamos a média dessas somas (denominadas **quadrados médios** ou MS). Para obter a média da soma dos quadrados é necessário dividir pelos graus de liberdade (isso é semelhante a calcular a variância a partir da soma dos quadrados – veja a Seção 1.4.1). Para o SS_M , o grau de liberdade é simplesmente o número de variáveis no modelo, e para o SS_R , ele é o número de observações menos o número de parâmetros sendo estimados (isto é, o número de coeficientes beta incluindo a constante). O resultado é a média dos quadrados para o modelo (MS_M) e a média dos quadrados dos resíduos (MS_R). Nesse estágio não é essencial que você entenda como as médias dos quadrados são derivadas (isso será explicado no Capítulo 8). Entretanto, é importante que você entenda que a razão F (equação 5.5) é uma medida do quanto o modelo melhorou na previsão de valores comparado com o nível de não precisão do modelo:

$$F = \frac{MS_M}{MS_R} \quad (5.5)$$

Se o modelo é bom, então esperamos que a melhoria na previsão devida ao modelo seja grande (assim, MS_M será grande) e a diferença entre o modelo e os dados observados serão pequenas (assim, MS_R será pequeno). Resumindo, um bom modelo deverá ter uma razão F grande (maior do que 1 pelo menos) porque o numerador da equação (5.5) será maior do que o denominador. A magnitude exata da razão F pode ser calculado utilizando valores críticos para os graus de liberdade correspondentes (como exemplificado no Apêndice A.3).

5.2.4 Interpretando previsões individuais ①

Vimos que o preditor em um modelo de regressão tem um coeficiente (b_1), que na regressão simples representa o gradiente da linha de regressão. O valor de b_1 representa a mudança na saída resultante da mudança de uma unidade no preditor. Se o modelo for inútil para previsão, então se o valor do preditor mudar, o que se poderá esperar da mudança no valor de saída? Bem, se o modelo é muito ruim, esperamos que essa alteração não ocorra, isto é, seja zero. Relembre da Figura 5.3 (veja o quadro representando SS_T) na qual vimos que utilizar a média não era a melhor forma de prever um resultado. De fato, a linha representando a média é horizontal, o que significa que à medida que a variável preditora muda, o valor previsto *não* se altera (porque para cada nível da variável preditora, prevemos que o valor de saída será o valor da média). O ponto importante aqui é que modelos ruins (como a média) terão um coeficiente de regressão igual a zero. Um coeficiente de regressão igual a zero significa: (1) a mudança de unidade na variável preditora não resultará em uma alteração no valor previsto (o valor previsto não muda de qualquer forma) e (2) o gradiente da linha de regressão é zero, assim, a linha de regressão será horizontal. A consequência lógica é que se uma variável prevê significativamente um valor de saída, então ela deve ter um valor

b_1 significativamente diferente de zero. Essa hipótese é testada utilizando o teste t (veja o Capítulo 7). A estatística t testa a hipótese nula de que o valor de b é zero: dessa forma, se for significativo aceitamos a hipótese que o valor de b é significativamente diferente de zero e que a variável previsora contribui significativamente para a nossa habilidade de estimar o valor de saída.

Nosso problema em testar se o valor de b é diferente de zero é que ele depende da unidade de medida. Dessa forma, o teste t é calculado levando em conta o erro padrão. O erro padrão nos informa algo sobre como seria o comportamento dos valores b se fossem tomadas muitas amostras relativas a vendas de discos e investimento em publicidade e o valor de b fosse calculado para cada uma das amostras. Poderíamos fazer um diagrama da distribuição de frequências dos valores b dessas amostras para descobrir se eles seriam semelhantes ou bem diferentes (lembre a Seção 1.6). Podemos utilizar o desvio padrão dessa distribuição (conhecido como *erro padrão*) como uma medida da semelhança dos valores b entre as amostras. Se o erro padrão é pequeno, significa que muitas amostras provavelmente têm valores b similares ao que foi obtido na amostra coletada (porque a variação entre as amostras é pequena). O teste t nos informa se o valor b é diferente de zero relativo à variação dos valores b para amostras de mesmo tamanho. Quando o erro padrão é pequeno, mesmo pequenos desvios de zero podem refletir uma diferença significativa porque b é representativo da maioria das possíveis amostras.

A equação (5.6) mostra como o teste t é calculado e você irá encontrar uma versão geral dessa equação no Capítulo 7 (equação (7.1)). O b_{esperado} é simplesmente o valor de b que esperaríamos obter se a hipótese nula fosse verdadeira. Mencionei antes que a hipótese nula é de que b é zero, assim, esse valor pode ser substituído por zero.* A equação torna-se

o valor observado de b dividido pelo erro padrão com qual ele está associado:³

$$t = \frac{(b_{\text{observado}} - b_{\text{esperado}})}{SE_b}$$

$$= \frac{b_{\text{observado}}}{SE_b}$$

Os valores de t seguem uma distribuição especial que diferem de acordo com um parâmetro denominado grau de liberdade. Na regressão, o grau de liberdade é simplesmente $N - p - 1$, onde N é o tamanho da amostra e p é o número de previsores. Na regressão simples, quando existe apenas um predictor, o grau de liberdade passa a ser $N - 2$. Tendo determinado que a distribuição t precisa ser utilizada, o valor observado de t pode ser comparado com valores que se esperaria obter apenas por acaso: se t é muito grande, é improvável que ele tenha ocorrido por acaso (esses valores podem ser encontrados no Apêndice A.2). O SPSS fornece a probabilidade exata de que o valor de t tenha ocorrido por acaso se o valor de b fosse de fato zero. Como regra geral, se esse valor (significância) for menor do que 0,05 ou 5%, os cientistas sociais concordam que b é significativamente diferente de zero; ou seja, o predictor contribui de forma significativa para o valor da saída (veja o Capítulo 1).

* N. de T.: Em geral, a hipótese nula é de que o coeficiente angular é zero, mas não necessariamente.

³ Para comprovar isso você pode utilizar os valores da saída 5.3 do SPSS abaixo a fim de calcular t para a constante (intercepto) ($t = 134,140/7,537 = 17,79$) e investimento em publicidade (gradiente). Para o investimento em publicidade, você irá obter um valor de t diferente se utilizar os valores como os apresentados na saída do SPSS. Isso ocorre porque o SPSS arredonda os valores para 3 casas decimais nas tabelas, mas calcula o valor t sem arredondar esses valores (normalmente isso não acarreta diferenças substanciais, mas nesse caso sim!). Para obter os valores não-arredondados, clique duas vezes na tabela da saída do SPSS e então dê um duplo clique no valor que você quer ver sem arredondamento. Você deve encontrar que $t = 0,096124/0,009632 = 9,979$.

5.3 EXECUTANDO REGRESSÃO SIMPLES NO SPSS ①

Até agora, vimos um pouco da base teórica da regressão, embora restrita a casos em que existe apenas um previsor. Para auxiliar a esclarecer o que já aprendemos, iremos apresentar um exemplo de regressão simples utilizando o SPSS. Anteriormente, pedi que você imaginasse que eu trabalhava para uma gravadora e que meu chefe estava interessado em prever a venda de discos a partir da publicidade. Existem alguns dados para esse exemplo no arquivo **Record1.sav**. Esse arquivo de dados tem 200 linhas, cada uma representando um disco diferente. Existem também duas colunas, uma representando as vendas de cada disco na semana após o lançamento e outra representando a quantidade (em libras) gasta em promover o disco antes do seu lançamen-

to. Esse é o formato para a entrada de dados de regressão: a variável de saída e a previso- ra devem estar em colunas diferentes e cada linha deve representar valores independentes dessas duas variáveis. O padrão desses dados é mostrado na Figura 5.4 e deve ficar claro que existe um relacionamento positivo: assim, quanto mais dinheiro for gasto na promoção de um disco, mais ele deve vender. É claro que alguns discos vendem bem a despeito do que for gasto em publicidade (canto superior esquerdo do diagrama), mas não existe um que venda muito mal quando muito dinheiro é gasto em publicidade (canto inferior direito do diagrama). O diagrama também mostra a linha de melhor aderência para esses dados: tendo em mente que a média é representada por uma linha horizontal em torno do valor de vendas de 200 000, a linha de regressão é visivelmente diferente.

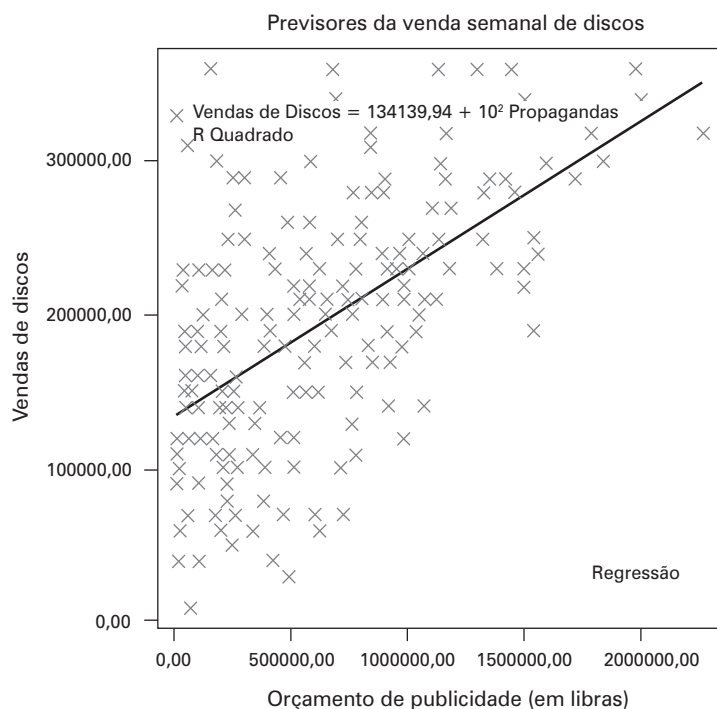


Figura 5.4 Diagrama de dispersão mostrando o relacionamento entre a venda de discos e a quantia gasta para promover o disco.

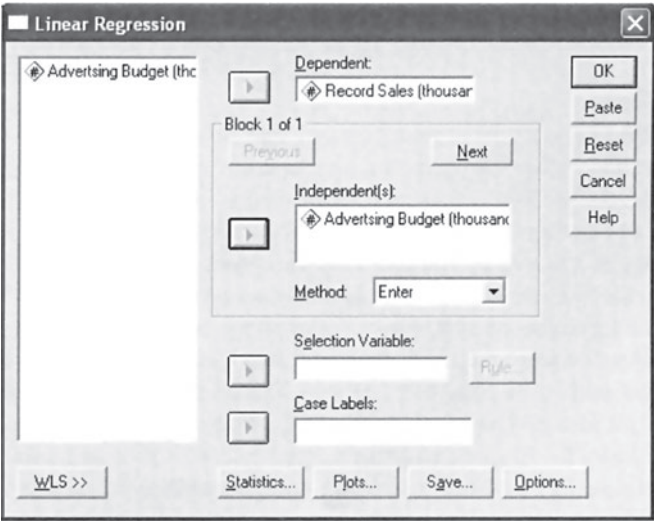


Figura 5.5 Caixa de diálogo principal para a regressão.

Para encontrar os parâmetros que descrevem a linha de regressão e para ver se essa linha é um modelo útil, precisamos executar uma análise de regressão. Para fazer a análise, você precisa acessar a caixa de diálogos principal utilizando o caminho **Analyze⇒Regression⇒Linear...** (Analisar⇒Regressão⇒Linear...). A Figura 5.5 mostra a caixa de diálogo resultante. Existe um espaço denominado *Dependent* (Dependente) no qual você deve colocar a variável de saída (nesse exemplo, **Sales** (Vendas)). Assim, selecione **sales** (vendas) da lista no lado esquerdo e transfira para o espaço clicando em . Existe outro espaço rotulado de *Independent(s)* (Independentes) no qual cada variável previsor deve ser colocada. Na regressão simples, utilizamos uma única variável previsor, (nesse exemplo, **adverts** [publicidade]) então você deve selecionar **adverts** da lista clicando em de modo a transferi-la para a lista dos previsores. Existe uma variedade de opções disponíveis, mas elas serão exploradas dentro do contexto da regressão múltipla (veja a Seção 5.5). Por enquanto, clique em **OK** para executar a análise.

5.4 INTERPRETANDO A REGRESSÃO SIMPLES ①

5.4.1 Ajuste global do modelo ①

A primeira tabela fornecida pelo SPSS é um resumo do modelo. Essa tabela de resumo (Saída do SPSS 5.1) fornece os valores do *R* e do *R*² para o modelo que foi calculado. Para esses dados, *R* tem um valor de 0,578, e porque existe apenas um previsor, esse valor representa a correlação simples entre o valor investido em publicidade e a venda de discos (você pode confirmar isso executando uma correlação utilizando o que foi visto no Capítulo 4). O valor de *R*² é 0,335, o que nos informa que o gasto em publicidade pode ex-

Saída 5.1 do SPSS

Model Summary (Resumo do modelo)

Model (Modelo)	<i>R</i>	<i>R</i> ²	Adjusted <i>R</i> Square (R quadrado ajustado)	Standard Error of the Estimate (Erro padrão da estimativa)
1	0.578 ^a	0.335	0.331	65.9914

a. (Previsores: (constante), Orçamento de publicidade (milhares de libras))

plicar 33,5% da variação das vendas de disco. Em outras palavras, se você está tentando explicar porque alguns discos vendem mais do que outros, podemos olhar a variação das vendas de discos diferentes. Devem existir muitos fatores que podem explicar essa variação, mas nosso modelo, que inclui somente a quantia gasta com publicidade, pode explicar 33% deles. Isso significa que 66% da variação das vendas não pode ser explicada apenas pela publicidade. Portanto, deve haver outras variáveis que também têm influência.

A próxima parte da saída relata uma análise de variância (ANOVA – veja o Capítulo 8). A tabela de resumo mostra as várias somas dos quadrados descritas na Figura 5.3 e os graus de liberdade associados com cada uma (Saída 5.2 do SPSS). A partir desses dois valores, a soma dos quadrados média pode ser calculada dividindo as somas dos quadrados pelo grau de liberdade associado. A parte mais importante da tabela é a razão F, que é calculada utilizando a equação (5.5) e a significância associada a esse valor F. Para esses dados, F é 99,59, que é significativo ao nível de $p < 0,001$ (porque o valor na coluna denominada *Sig.* é menor do que 0,001). Esse resultado nos informa que existe uma probabilidade menor do que 0,1% de que um valor F tão alto tenha ocorrido apenas por acaso. Desta forma, pode-se concluir que nosso modelo de regressão resulta em previsões melhores das vendas de

disco do que se utilizássemos o valor médio das vendas. Resumindo, o modelo de regressão prevê as vendas de disco bastante bem.

5.4.2 Parâmetros do modelo ①

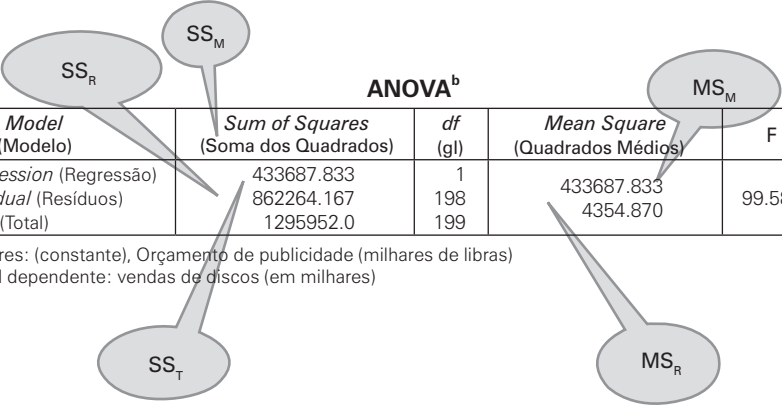
A ANOVA nos informa se o modelo, em geral, resulta em um grau de previsão significativamente bom dos valores da variável de saída. No entanto, a ANOVA não nos informa sobre a contribuição individual das variáveis no modelo (embora neste caso simples exista uma única variável no modelo e, assim, podemos inferir que esta variável é um bom preditor). A tabela na saída 5.3 do SPSS fornece detalhes dos parâmetros do modelo (os valores beta) e da significância desses valores. Vimos na equação (5.2) que b_0 é o intercepto y (ponto onde a linha corta o eixo Y) e esse é o valor B (na saída do SPSS) para a constante (intercepto). Assim, da tabela, podemos dizer que b_0 é 134,14 e isso significa que quando nenhum dinheiro for aplicado em publicidade (quando $X = 0$), o modelo prevê que 134 140 discos serão vendidos (lembre que a unidade de medida das vendas é em milhares de discos). Também podemos ler o valor de b_1 da tabela e esse valor representa a gradiente (inclinação)



Saída 5.2 do SPSS

ANOVA ^b					
Model (Modelo)	Sum of Squares (Soma dos Quadrados)	df (gl)	Mean Square (Quadrados Médios)	F	Sig.
1 Regression (Regressão)	433687.833	1	433687.833	99.587	0.000 ^a
Residual (Resíduos)	862264.167	198	4354.870		
Total (Total)	1295952.0	199			

a. Previsores: (constante), Orçamento de publicidade (milhares de libras)
b. Variável dependente: vendas de discos (em milhares)



Saída 5.3 do SPSS

Coefficients^a (Coeficientes)

Model (Modelo)	Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
	B	Std. Error (Erro Padrão)	Beta		
1 (Constant) Advertising Budget (thousands of pounds)	134.140	7.537		17.799	0.000
(Constante) (Milhares de libras) (Investimento em Publicidade)	9.612E-02	0.010	0.578	9.979	0.000

a.Variável dependente: vendas de discos (em milhares).

da linha de regressão. Ele é 9,612E-02, que é a expressão em notação científica do valor 0,09612.⁴ Embora esse valor seja a inclinação da linha de regressão, é mais útil pensar nele como representando *a mudança da variável de saída para cada alteração de uma unidade no previsor*. Dessa forma, se a nossa variável previsora é aumentada de uma unidade (se o orçamento de publicidade aumenta 1), nosso modelo prevê que 0,096 discos adicionais serão vendidos. Como a unidade de medida está em milhares de libras e milhares de discos vendidos, pode-se dizer que para um aumento de £1000 em publicidade, o modelo prevê 96 ($0,096 \times 1000 = 96$) discos extras vendidos. Como você pode imaginar, é um péssimo investimento para a gravadora: eles investem £1000 e vendem apenas 96 discos a mais! Felizmente, como já sabemos, a quantia investida em publicidade representa apenas um terço das vendas de discos!

⁴ Você pode ter percebido que esse valor é apresentado pelo SPSS como 9,612E-02 e muitos estudantes acham essa notação confusa. Bem, essa notação significa simplesmente $9,61 \times 10^{-2}$ (que poderá ser uma notação um pouco mais familiar). Ok, alguns de vocês continuam confusos. Então pense no E-02 como se significasse “mova o ponto decimal duas casas para a esquerda”, assim, 9,612E-02 torna-se 0,09612. Se o número fosse 9,612E-01, ele seria 0,9612 e se fosse 9,612E-03, seria 0,009612. Da mesma forma, leia o E+02 (note que o sinal de menos mudou) como “mova a vírgula duas casas para a direita”. Assim, 9,612E+02 é a representação científica de 961,2.

Vimos anteriormente que, em geral, o valor do coeficiente de regressão *b* representa a mudança na saída resultante da variação de uma unidade na variável de entrada. Se essa variável previsora está tendo um impacto significativo na nossa habilidade de prever um resultado, esse *b* deve ser diferente de 0 (e grande comparado ao erro padrão). Vimos ainda que o teste *t* nos informa se o valor *b* é diferente de 0. O SPSS fornece a probabilidade exata de que o valor de *t* ocorra se o valor de *b* é zero. Como regra geral, se essa significância observada é menor do que 0,05, os cientistas sociais concordam que o resultado reflete um efeito genuíno (veja o Capítulo 1). Para esses dois valores, as probabilidades são 0,000 (zero até a terceira décima) e, dessa forma, podemos dizer que a probabilidade de que esses valores *t* ocorram se os valores *b* são zero é menor do que 0,001. Assim, os *bs* são diferentes de 0 e podemos concluir que o investimento em publicidade tem uma contribuição significativa ($p < 0,001$) para prever o valor das vendas de disco.

5.4.3 Utilizando o modelo ①

Até agora descobrimos que temos um modelo útil, um que melhora significativamente nossa habilidade de prever a venda de discos. No entanto, o próximo estágio é usar o modelo para fazer algumas previsões. O primeiro estágio é definir o modelo substituindo os valores *b* na equação (5.2) com os valores da saída 5.3

do SPSS. Além disso, podemos substituir o X e o Y com os nomes das variáveis, assim, o modelo torna-se:

$$\begin{aligned} \text{Vendas de} &= b_0 + b_1 \text{Investimento em} \\ \text{discos}_i & \quad \text{Publicidade}_i \\ &= 134,14 + (0,09612 \times \\ & \quad \text{Investimento em} \\ & \quad \text{Publicidade}_i) \end{aligned} \quad (5.7)$$

Agora é possível fazer uma previsão sobre a venda de discos substituindo o valor da publicidade com o valor de interesse. Por exemplo, imagine que um executivo de uma gravadora queira gastar £100.000 em publicidade em um disco a ser lançado. Lembrando que as nossas unidades já estão em milhares de libras, você pode simplesmente substituir o valor da publicidade por 100. Ele irá, então, descobrir que a venda de discos deve ficar em torno de 144.000 para a primeira semana de vendas:

$$\begin{aligned} \text{Vendas de} &= 134,14 + (0,09612 \times \\ \text{discos}_i & \quad \text{Investimento em Publicidade}_i) \\ &= 134,14 + (0,09612 \times 100) \\ &= 143,75 \end{aligned} \quad (5.8)$$

5.5 REGRESSÃO MÚLTIPLA: O BÁSICO ②

Qual é a diferença entre a regressão simples e a múltipla?



Para resumir o que apreendemos até agora, na regressão linear simples a variável de saída Y é prevista utilizando uma equação de uma linha (equação (5.2)). Supondo que foram coletados vários valores das variáveis X

e Y , os parâmetros desconhecidos da equação podem então ser determinados. Eles são calculados pelo ajuste do modelo aos dados (nesse caso, uma linha reta) para o qual a soma das diferenças ao quadrado entre os valores sobre a linha e os dados originais é minimizada. Esse método é denominado método dos mínimos quadrados. A regressão múltipla é uma extensão lógica desses princípios para a situa-

ção em que existem vários previsores. Novamente, ainda utilizamos a equação básica:

$$\text{Saída}_i = (\text{Modelo}_i) + \text{Erro}_i$$

Desta vez o modelo é um pouco mais complicado. Ele é basicamente o mesmo da regressão simples, exceto que para cada predictor extra que for incluído, um coeficiente é adicionado; assim, cada variável predictor tem seu próprio coeficiente e a variável de saída é prevista a partir de uma combinação de todas as variáveis multiplicadas pelos seus respectivos coeficientes mais o (intercepto). (Veja a equação (5.9).) Os parênteses não são necessários, eles servem apenas para fazer uma analogia com a equação geral:

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n) + \varepsilon_i \quad (5.9)$$

Y é a variável de saída (resultado), b_1 é o coeficiente do primeiro predictor (X_1), b_2 é o coeficiente do segundo predictor (X_2), b_n é o coeficiente do n -ésimo predictor (X_n) e ε_i é a diferença entre o valor previsto e o observado de Y para o i -ésimo participante. Neste caso, o modelo ajustado é mais complicado, mas os princípios básicos são os mesmos da regressão simples. Isto é, procuramos encontrar uma combinação linear de previsores que se correlacionam de forma máxima com a variável de saída. Assim, quando nos referirmos ao modelo de regressão múltipla, estaremos falando sobre um modelo na forma da equação (5.9).

5.5.1 Um exemplo de um modelo de regressão múltipla ②

Imagine que o nosso executivo da gravadora está interessado em ampliar o modelo da venda de discos incorporando uma nova variável. Já sabemos que a publicidade é responsável por 33% da variação das vendas de discos, mas um valor bem maior (67%) permanece não explicado. O executivo pode avaliar um novo predictor para o modelo numa tentativa de justificar algumas das variações não explicadas na venda de discos. Ele decide medir o número de vezes que o disco é tocado na rádio 1 (Estação Nacional de Rádio Britânica) durante a semana

anterior ao lançamento. O modelo existente que foi derivado utilizando o SPSS (veja a equação (5.7)) pode agora ser entendido para incluir essa nova variável (**airplay** – execuções):

$$\text{Venda de discos}_i = b_0 + b_1 \text{Publicidade}_i + b_2 \text{Execuções}_i + \varepsilon_i \quad (5.10)$$

O novo modelo é baseado na equação (5.9) e inclui valores b para os dois previsores (e, é claro, a constante). Se calcularmos os valores b , podemos fazer previsões sobre a venda de discos não apenas utilizando a quantidade investida em publicidade, mas também em termos de execuções do disco na rádio. Existem somente dois previsores no modelo e, assim, ele pode ser apresentado graficamente em três dimensões (Figura 5.6).

A equação (5.9) descreve o trapézio escurecido no diagrama (conhecido como o *plano*

de regressão), e os pontos destacados representam os pontos de dados observados. Como na regressão simples, o plano ajustado aos dados pretende ser o melhor possível para esses dados. No entanto, existem invariavelmente algumas diferenças entre o modelo e os dados da vida real (esse fato é evidente porque alguns dos pontos não residem exatamente sobre a área escurecida do gráfico). O valor b para a publicidade descreve a inclinação de cima para baixo do plano de regressão enquanto que o valor b para a execução (*airplay*) fornece a inclinação da esquerda para a direita do plano de regressão. Conhecer esses dois valores permite que coloquemos o plano de regressão no espaço.

É bastante simples visualizar o modelo de regressão com dois previsores utilizando um diagrama de dispersão tridimensional. Entretanto, a regressão múltipla pode ser utilizada com três, quatro ou mesmo mais variáveis.

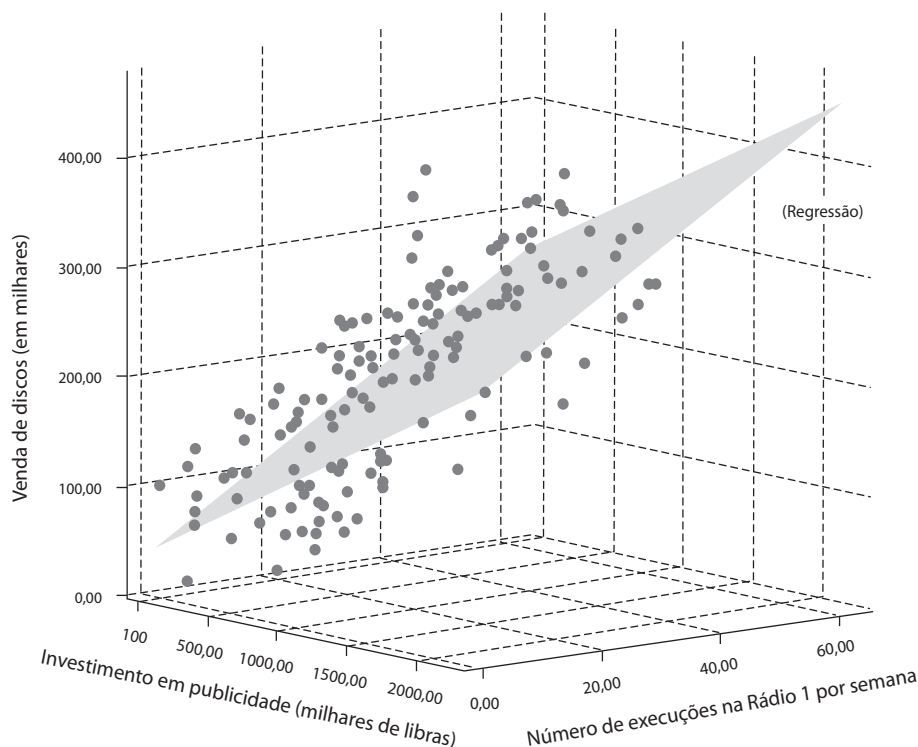


Figura 5.6 Diagrama de dispersão do relacionamento entre vendas de discos, investimento em publicidade e número de execuções do disco no rádio.

Embora você não possa imediatamente visualizar como tal modelo complexo se parece, ou o que os valores b representam, você deve ser capaz de aplicar os princípios desses modelos básicos a cenários mais complexos.

5.5.2 Soma dos quadrados, R e R^2 ②

Quando temos vários previsores, a partição da soma dos quadrados é a mesma que foi executada para uma única variável exceto que o modelo que estamos nos referindo toma a forma da equação (5.9) em vez de simplesmente ser uma linha reta bidimensional. Portanto, um SS_T pode ser calculado e representar a diferença entre os valores observados e a média dos valores da variável resultado. O SS_R ainda representa a diferença entre os valores previstos de Y pelo modelo e os valores observados de Y . Finalmente, SS_M pode ainda ser calculado e representa a diferença entre os valores de Y previstos pelo modelo e o valor da média. Embora o cálculo desses valores seja muito mais complicado que na regressão simples, conceitualmente esses valores são os mesmos.

Quando existem vários previsores, não faz sentido olhar para os coeficientes de correlação simples e, neste caso, o SPSS produz um coeficiente de correlação múltiplo (denominado **R Múltiplo**). O R Múltiplo é a correlação entre os valores observados de Y e os de Y previstos pelo modelo de regressão múltipla. Desta forma, valores grandes do R Múltiplo representam uma alta correlação entre os valores previstos e observados da variável de saída. Um R Múltiplo igual a 1 representa a situação na qual o modelo prediz com perfeição os valores observados, isto é, ele adere perfeitamente a todos os pontos. Como tal, o R Múltiplo é uma medida de quão bem o modelo prevê os dados observados. Segue que o R^2 resultante pode ser interpretado da mesma forma que na regressão simples: ele é a quantidade de variação na variável de saída que pode ser creditada ao modelo.

5.5.3 Métodos de regressão ②

Se você está interessado em construir um modelo complexo com vários previsores,

como decidimos que previsores utilizar? Muito cuidado deve ser tomado ao selecionarmos previsores para um modelo porque os valores dos coeficientes de regressão dependem das variáveis utilizadas no modelo. Desta forma, os previsores incluídos e a forma que eles são colocados podem ter um grande impacto. Num mundo ideal, previsores deveriam ser selecionados baseados em pesquisas anteriores.⁵ Se novos previsores estão sendo adicionados a modelos existentes, selecione aquelas novas variáveis com base na importância *teórica* substantiva de tais variáveis. Você não deve de forma alguma selecionar centenas de previsores ao acaso, juntá-los todos em uma análise de regressão e torcer pelo melhor. Além do problema da seleção de previsores, existem várias formas de entrada das variáveis no modelo. Quando os previsores são todos completamente não correlacionados, a ordem de entrada das variáveis tem pouco efeito nos parâmetros calculados; no entanto, nas pesquisas em ciências sociais raramente temos previsores não-correlacionados, assim, o método de seleção dos previsores é crucial.

5.5.3.1 Hierárquico (Entrada em blocos) ②

Na regressão hierárquica, os previsores são selecionados com base em trabalhos anteriores e o pesquisador decide em que ordem eles devem ser colocados no modelo. Como regra geral, previsores conhecidos (de outras análises) devem ser colocados no modelo primeiro pela sua ordem de importância em poder prever a variável de saída. Depois de entrar com todos os previsores conhecidos, o pesquisador pode adicionar os novos ao modelo. Novas variáveis podem ser colocadas de uma única vez, passo a passo ou hierarquica-

⁵ Eu, um tanto cinicamente, apresento essa sugestão propondo que os previsores sejam escolhidos com base em pesquisas anteriores que tenham utilizado uma boa metodologia. Se você basear tais decisões nas análises de regressão, selecione previsores baseados somente em pesquisas passadas que utilizaram a regressão de forma apropriada e com modelos confiáveis e generalizáveis!

mente (de modo que o novo preditor que se julga mais importante entre primeiro).

5.5.3.2 Entrada forçada ②

Entrada forçada (ou *Enter*, como é conhecida no SPSS) é um método em que todos os preditores são forçados no modelo ao mesmo tempo. Da mesma forma que a hierárquica, este método baseia-se em boas razões teóricas para incluir os preditores escolhidos, mas diferentemente da hierárquica, o pesquisador não toma decisões sobre a ordem que as variáveis são acrescentadas.

5.5.3.3 Métodos por passos (Stepwise) ②

Na regressão passo a passo, as decisões sobre a ordem em que os preditores são acrescentados ao modelo é baseada puramente em critérios matemáticos. No método *forward* (para frente), um modelo inicial contendo somente a constante (b_0) é definido. O computador então procura pelo preditor (entre as variáveis possíveis) que melhor prevê a variável de saída – ele faz isso selecionando o preditor que apresenta o coeficiente de correlação simples mais alto com a variável de saída. Se esse preditor aumenta significativamente a habilidade do modelo prever a saída, ele é mantido no modelo e o computador procura por um segundo preditor. O critério para selecionar esse segundo preditor é procurar pela variável que apresenta a maior correlação semiparcial com a variável de saída. Deixe-me explicar isso de outra forma. Imagine que o primeiro preditor pode explicar 40% da variação da variável de saída; portanto, existe ainda 60% não explicada. O computador procura pelo preditor que pode explicar a mais alta taxa dos 60% restantes (assim, ele não está interessado nos 40% que já foram explicados). Em termos estatísticos, você pode pensar isso como uma correlação parcial na qual o computador correlaciona cada um dos preditores com a saída enquanto controla o efeito do primeiro preditor. A razão para chamá-la de correlação *semiparcial* é porque o efeito do primeiro preditor é parcializado apenas entre os preditores restantes

e não são controlados na variável de saída. A correlação semiparcial fornece uma medida de quanto a “nova variância” na saída pode ser explicada por cada um dos preditores restantes (veja a Seção 4.6). O preditor que contribui mais para a nova variância é adicionado ao modelo, e se ele tiver uma contribuição significativa para o poder preditivo do modelo ele é mantido e outro preditor é procurado.

O método passo a passo (*stepwise*) do SPSS é o mesmo que o método *forward* (para frente), exceto que a cada vez que um preditor é adicionado a equação, um teste de remoção é feito sobre o preditor menos útil. Para tal a equação de regressão está constantemente sendo reaccessada para ver se algum preditor redundante pode ser removido. O método para trás (*backward*) é o oposto do método para frente (*forward*), já que o computador inicia colocando todos os preditores no modelo e então calcula a contribuição de cada um verificando a significância do teste *t* de cada preditor. O valor da significância é comparado com um critério de remoção (que pode ser tanto o valor absoluto de uma estatística teste quanto o valor da probabilidade da ocorrência dessa estatística teste). Se o preditor satisfaz o critério de remoção (isto é, se ele não está fazendo uma contribuição estatisticamente significativa para o quão bem o modelo prevê a variável resultado) ele é removido do modelo e o modelo é reestimado com os preditores restantes. A contribuição dos demais preditores é então reavaliada.

Se você decidir utilizar o método passo a passo (*stepwise*), o método para trás (*backward*) é preferível ao método para frente (*forward*). Isso ocorre em virtude dos efeitos supressores, que acontecem quando um preditor tem um efeito significativo, mas somente quando outra variável é mantida constante. A seleção para frente (*forward*) tem uma probabilidade maior do que a para trás (*backward*) de excluir preditores envolvidos em efeitos supressores. Assim, com o método para frente (*forward*) há um risco maior de cometermos erro do Tipo II (isto é, eliminar um preditor que de fato contribui para o modelo).

5.5.3.4 Escolhendo um método ②



O SPSS permite que você opte por qualquer um desses métodos e é importante selecionar o mais apropriado. Os métodos para frente, para trás e passo a passo (*forward*, *backward* e *stepwise*) aparecem todos

sob o título geral de métodos passo a passo (*stepwise methods*) porque em todos eles o computador escolhe as variáveis com base em um critério matemático. Muitos escritores argumentam que isso tira das mãos do pesquisador decisões metodológicas importantes. E mais, os modelos utilizados pelo computador frequentemente tiram vantagem da variação da amostragem aleatória e, assim, decisões sobre que variáveis devem ser incluídas serão baseadas em pequenas diferenças nas correlações semiparciais. Entretanto, essas pequenas diferenças estatísticas podem contrastar dramaticamente com a importância teórica do previsor para o modelo. Por essa razão, é melhor evitar os métodos passo a passo (*stepwise*) exceto em uma análise exploratória do modelo (ver Wright, 1997, p. 181). Quando existe uma forte literatura teórica disponível, tome por base o que as pesquisas anteriores apontaram. Inclua qualquer variável significativa no modelo pela sua ordem de importância. Após essa análise inicial, repita a regressão, mas exclua qualquer variável que for estatisticamente redundante na primeira vez. Existem considerações importantes na decisão de quais previsores incluir. É importante não incluir muitos previsores. Como regra geral, quanto menos, melhor e certamente inclua somente previsores para os quais você tenha razões teóricas sólidas. Assim, seja seletivo e lembre que você deve ter um tamanho de amostra decente (alguns sugerem pelo menos 15 participantes por previsor – veja a Seção 5.6.2.3)

5.6 QUÃO ACURADO É O MEU MODELO DE REGRESSÃO? ②

Quando produzimos um modelo baseado em uma amostra dos dados existem duas questões importantes a serem feitas: (1) o modelo representa bem os dados, ou ele é influen-



ciado por um pequeno número de casos e (2) o meu modelo pode ser generalizado para outras amostras? É fundamental fazer essas questões porque elas irão afetar como utilizaremos o modelo que está sendo construído. Essas questões são também, em certo sentido, hierárquicas, pois nós não queremos generalizar um modelo ruim. No entanto, é um erro pensar que se um modelo adere bem aos dados podemos tirar conclusões além da nossa amostra.

Generalização é um passo adicional crítico; se acharmos que o nosso modelo não é generalizável, devemos restringir qualquer conclusão baseada no modelo da amostra utilizada. Primeiro vamos analisar como determinamos se um modelo é uma representação acurada dos dados reais e depois, na seção 5.6.2, verificaremos como é possível avaliar se um modelo pode ser usado para fazer inferências além da amostra dos dados que foi coletada.

5.6.1 Interpretando o modelo de regressão I: diagnósticos ②

Para responder a questão de quão bem o modelo adere aos dados observados ou se ele é influenciado por um pequeno número de casos, podemos buscar valores atípicos (*outliers*) ou casos influenciadores (a diferença é explicada no quadro 5.1). Veremos isso em etapas.

5.6.1.1 Valores atípicos (*outliers*) e resíduos ②

Um valor atípico é um caso que difere substancialmente da maioria dos dados (veja o Quadro 3.1). A Figura 5.7 mostra um exemplo de tal caso na regressão. Valores atípicos

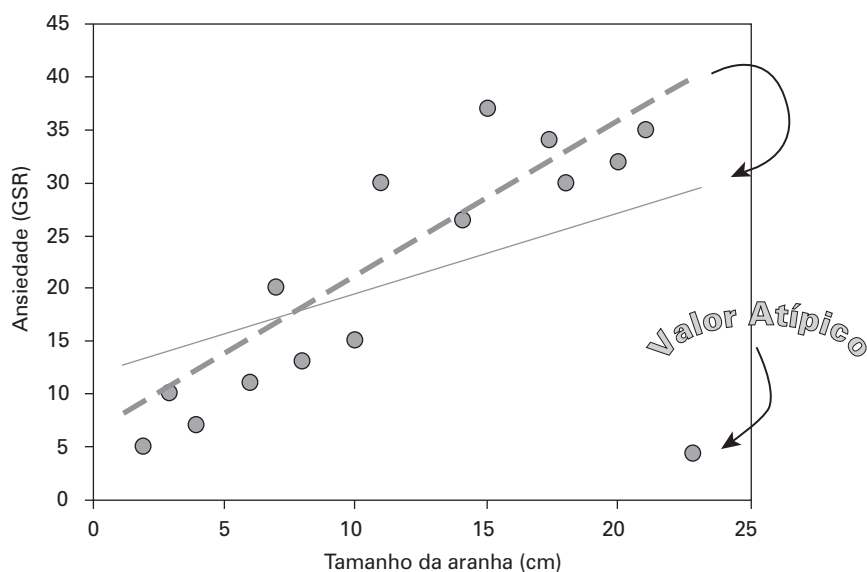


Figura 5.7 Gráfico demonstrando o efeito de um valor atípico. A linha pontilhada representa a linha de regressão original para esses dados (veja a Figura 5.2), e a linha cheia representa a linha de regressão quando um valor atípico está presente.

podem introduzir tendenciosidade no modelo, pois eles irão afetar os valores dos coeficientes de regressão estimados. Por exemplo, a Figura 5.7 utiliza os mesmos dados da Figura 5.2, exceto que o escore de um participante foi alterado para ser um valor atípico (nesse caso, uma pessoa que ficou extremamente calma na presença de uma aranha muito grande). A mudança desse único valor teve um efeito dramático no modelo de regressão escolhido. Com o valor atípico presente, o modelo de regressão muda: a inclinação (gradiente) é reduzida (a linha se torna mais horizontal) e o intercepto aumenta (a linha cortará o eixo Y num ponto mais alto). Deve ficar claro a partir desse diagrama que é importante tentar detectar os valores atípicos para ver se o modelo é tendencioso.

Como você acha que poderia detectar um valor atípico? Sabemos que um valor atípico, pela sua natureza, é muito diferente de todos os demais escores. Assim, você acha que o modelo irá prever os escores dessas pessoas de forma acurada? Obviamente, a resposta é *não*: olhando para a Figura 5.7 é evidente que mesmo que o valor atípico tenha distorcido o

modelo ele ainda prevê esse valor só que de uma forma muito ruim (a linha de regressão está bem longe dele). Portanto, se formos calcular as diferenças entre os valores dos dados coletados e os valores previstos pelo modelo, podemos detectar os valores atípicos buscando as diferenças grandes. Esse processo é o mesmo que procurar por casos em que o modelo não prevê de forma precisa. As diferenças entre os valores previstos pelo modelo e os valores observados na amostra são conhecidos como *resíduos*. Esses resíduos representam o erro que está presente no modelo. Se o modelo se ajusta bem aos dados da amostra, todos os resíduos devem ser pequenos (se o modelo aderir perfeitamente aos dados, todos os pontos estarão sobre a linha de regressão e todos os resíduos serão iguais a zero). Se o modelo não tiver uma boa aderência aos dados da amostra, os resíduos serão grandes. Além disso, se qualquer caso destacar-se por ter um grande resíduo, ele poderá ser atípico.

Os **resíduos normais** ou **não-padronizados** descritos acima são mensurados na mesma unidade da variável de saída e, portanto,

são difíceis de interpretar entre modelos. O que podemos fazer é procurar resíduos que se destacam por serem particularmente grandes. Para resolver esse problema, utilizamos **resíduos padronizados**, que são resíduos divididos por uma estimativa do seu desvio padrão. Já apresentamos o processo de padronização na Seção 1.4.1 como uma forma de converter variáveis em unidades padrões de medida (o desvio padrão); também lidamos com os escores-*z* (veja o Quadro 3.2), em que variáveis são convertidas em unidades de desvios padrão (isto é, elas são convertidas em escores que são distribuídos com média 0 e desvio padrão 1). Convertendo resíduos em escores-*z* (resíduos padronizados), podemos compará-los entre diferentes modelos e utilizarmos o que sabemos sobre os escores-*z* para derivar propriedades gerais sobre o que seria um valor aceitável (ou inaceitável). Por exemplo, vimos no Capítulo 3 que, numa amostra normalmente distribuída, 95% dos escores-*z* devem estar entre $-1,96$ e $+1,96$, 99% devem estar entre $-2,58$ e $+2,58$ e que 99,9% (isto é, praticamente todos) devem ficar entre $-3,29$ e $+3,29$. Algumas regras gerais para os resíduos padronizados são derivadas desses fatos: (1) resíduos padronizados com um valor absoluto maior do que 3,29 (de fato, normalmente utilizamos 3) são preocupantes porque geralmente, em uma amostra, valores grandes assim dificilmente ocorrem por acaso; (2) se mais do que 1% da amostra padronizada apresenta resíduos padronizados com valores absolutos maiores do que 2,58 (normalmente utilizamos 2,5), existem evidências de que o nível de erro dentro do nosso modelo é inaceitável (o modelo não se ajusta bem aos dados da amostra); e (3) se mais de 5% dos casos tem resíduos padronizados com um valor absoluto maior do que 1,96 (normalmente utilizamos 2 por conveniência), também há evidências de que o modelo é uma representação ruim dos dados.

Uma terceira forma de resíduo é o **resíduo estudentizado**, que é o valor não padronizado dividido por uma estimativa do desvio padrão entre eles que varia ponto a ponto. Esses resíduos apresentam as mesmas propriedades que

os padronizados, mas normalmente fornecem uma estimativa mais precisa da variância do erro para um caso específico.

5.6.1.2 Casos influentes ③

Além de procurar valores atípicos olhando para os erros do modelo, também é possível buscar certos casos que influenciam os parâmetros do modelo. Assim, se retirássemos determinados casos, obteríamos coeficientes da regressão diferentes? Esse tipo de análise pode ajudar a determinar se o modelo de regressão é estável por toda a amostra ou se ele pode estar sendo influenciado somente por poucos casos. Novamente, esse processo irá revelar os valores atípicos.



Existem várias estatísticas residuais que podem ser utilizadas para avaliar a influência de um caso em particular. Uma dessas estatísticas é o **valor previsto ajustado**, utilizada quando esse caso é excluído da análise. De fato, o computador calcula o novo modelo sem o caso em questão e então usa esse novo modelo para prever o valor que esse caso teria. Se o caso não exerce uma grande influência sobre o modelo, é esperado que o valor previsto e o valor previsto ajustado sejam semelhantes. Simplificando, se o modelo é estável, o valor previsto para o caso deve ser o mesmo não importa se ele tenha sido utilizado ou não para determinar o modelo. A diferença entre o valor previsto ajustado e o valor previsto original é conhecida como **DFFit** (veja adiante). Podemos ainda analisar o resíduo com base no valor previsto ajustado: isto é, a diferença entre o valor previsto ajustado e o valor original ajustado. Esse é o **resíduo excluído**. Esse resíduo pode ser dividido pelo erro padrão para fornecer um valor padronizado conhecido como **resíduo excluído estudentizado**. Esse resíduo pode ser comparado entre diferentes análises de regressão porque ele é medido em unidades padronizadas.

Os resíduos excluídos são muito úteis para verificar a influência de um caso na habilidade de o modelo prever esse caso. Contudo, eles não fornecem informação sobre como um

caso influencia o modelo como um todo (isto é, o impacto que um caso tem na habilidade do modelo prever *todos* os casos). Uma estatística que considera o efeito de um único caso no modelo como um todo é a **distância de Cook**. Ela é uma medida da influência global de um caso sobre o modelo e Cook e Weisberg (1982) sugeriram que valores maiores do que 1 merecem atenção.

Uma segunda medida desse tipo é a *influência* (*leverage*) – algumas vezes denominada *valores chapéu* (*hat values*) –, que mede o quanto um valor observado influencia o valor previsto da variável de saída. O valor médio da *influência* é definido como $(k + 1)/n$, onde k é número de previsores do modelo e n é o número de participantes.⁶ Os valores da *influência* podem variar entre 0 (indicando que o caso não tem influência alguma) e 1 (indicando que o caso tem total influência sobre a previsão). Se nenhum caso exerce excessiva influência sobre o modelo, espera-se que todos os valores da *influência* estejam próximos do valor médio ($(k + 1)/n$). Hoaglin e Welsch (1978) recomendam investigar casos com valores maiores do que 2 vezes o valor médio ($2(k + 1)/n$) e Stevens (1992) recomenda utilizar o triplo do valor médio ($3(k + 1)/n$) como uma base para identificar casos que exercem grande influência. Veremos como utilizar esses pontos de corte na Seção 5.8.6. Entretanto, casos com grandes valores de *influência* não terão necessariamente uma grande influência nos coeficientes da regressão porque eles estão mensurados na unidade da variável de saída em vez de nas dos previsores.

Relacionados aos valores de *influência* estão as **distâncias de Mahalanobis** (Figura 5.8), que medem os afastamentos dos valores a partir das médias das variáveis predictoras. Você precisa olhar para os casos com os maiores valores. Não é fácil estabelecer um ponto a partir do qual teremos motivos para ficarmos



Figura 5.8 Prasanta Chandra Mahalanobis olhando para o horizonte.

preocupados, embora Barnett e Lewis (1978) tenham produzido uma tabela de valores críticos que depende do número de previsores e do tamanho da amostra. A partir desse trabalho, fica claro que mesmo com grandes amostras ($N = 500$) e cinco previsores, valores acima de 25 são motivos de alerta. Em pequenas amostras ($N = 100$) com poucos previsores (normalmente três) valores maiores do que 15 são problemáticos, e em amostras bastante pequenas ($N = 30$) com somente dois previsores, valores maiores do que 11 devem ser examinados. Para mais detalhes, veja a tabela de Barnett e Lewis (1978).

É possível executar uma análise de regressão com o caso incluído e então executá-la novamente com o mesmo caso excluído. Se fizermos isso, sem dúvida existirão algumas diferenças entre os coeficientes b nas duas equações de regressão. Essas diferenças nos informarão quanta influência um determinado caso tem nos parâmetros do modelo de regressão. Como exemplo, imagine duas variáveis que apresentem um relacionamento negativo perfeito excepto por um único caso (caso 30). Se a análise de regressão for realizada com os 29 casos que estão linearmente relacionados de forma perfeita, obteremos um modelo no qual a variável previ-

⁶ Você pode encontrar a média da influência representada como p/n , onde p é o número de parâmetros sendo estimados. Na regressão múltipla, estimamos parâmetros para cada predictor e também para a constante; assim, p é equivalente ao número de previsores mais 1 ($k + 1$).

sora X prevê perfeitamente a variável de saída Y e não existirão erros. Se executarmos a análise novamente, mas desta vez incluindo o caso que não se ajusta (caso 30), o modelo resultante terá parâmetros diferentes. Alguns dados estão armazenados no arquivo **dfbeta.sav** que ilustra esta situação. Tente executar uma regressão simples primeiro com todos os casos incluídos e depois sem o caso 30. Os resultados estão resumidos na Tabela 5.1, que mostra (1) os parâmetros para o modelo de regressão quando o caso extremo está incluído ou excluído; (2) as equações de regressão resultantes; e (3) o valor de Y previsto para o escore do participante 30 na variável x (que é obtido substituindo o valor de X na equação de regressão pelo escore X do participante 30, que é 1).

Quando o caso 30 é excluído, esses dados apresentam um relacionamento negativo perfeito: o coeficiente para o previsor (b_1) é -1 (lembre que na regressão simples esse termo é o mesmo que o coeficiente de correlação de Pearson) e o coeficiente para a constante (o intercepto, b_0) é 31. Contudo, quando o caso 30 é incluído, os dois parâmetros são reduzidos⁷ e a diferença entre os parâmetros também é mostrada. A diferença entre um parâmetro estimado utilizando todos os casos e estimado quando um caso é excluído é conhecida como **DFBeta** no SPSS. O DFBeta é calculado para cada caso e para cada um dos parâmetros do modelo. Assim, no nosso exemplo, o DFBeta para a constante é -2 e o DFBeta para o previsor é $0,1$. Olhando para esses valores do DFBeta, é possível identificar casos que tem uma

grande influência nos parâmetros do modelo de regressão. Novamente, as unidades de medidas utilizadas afetarão esses valores, portanto, o SPSS produz um **DFBeta padronizado**. Esses valores padronizados são mais fáceis de usar porque pontos de corte universais podem ser aplicados. Nesse caso, valores absolutos acima de 1 indicam casos que substancialmente influenciam os parâmetros do modelo (embora Stevens, 1992, sugira olhar para os casos com valores absolutos maiores do que 2).

Uma estatística relacionada é a **DFFit**, que é a diferença entre o valor previsto para o caso quando o modelo é calculado incluindo o caso e quando o modelo é calculado excluindo o caso; neste exemplo, o valor é $-1,09$ (veja a Tabela 5.1). Se um caso não é influente, seu DFFit deve ser 0 – no entanto, é esperado que mesmo casos não-influente tenham pequenos valores DFFit. Contudo, há o problema de que essa estatística depende da unidade de medida da variável de saída e, assim, um DFFit de 0,5 será muito pequeno se a saída variar de 1 a 100, mas muito grande se as saídas variarem de 0 a 1. Portanto, o SPSS também apresenta uma versão padronizada dos valores DFFit (**DFFit padronizados**). Uma medida final é a razão de covariância (CVR), que é uma medida de quanto um caso influencia a variância dos parâmetros de regressão. Uma descrição do cálculo dessa estatística deixaria muitos leitores tontos e confusos, portanto, é suficiente dizer que se essa razão está próxima de 1, o caso tem pouca influência nas variâncias dos parâmetros do modelo. Belsey, Kuh e Welsch (1980) recomendam o seguinte:

- Se $CVR_i > 1 + [(3(k+1)/n)]$, então excluir o i -ésimo caso prejudicará a precisão de alguns dos parâmetros do modelo.

⁷ O valor de b_1 é reduzido porque os dados não apresentam mais um relação linear perfeita e existe agora uma variância que o modelo não pode explicar.

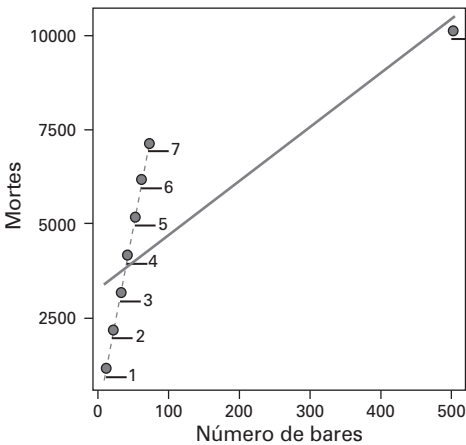
Tabela 5.1 A diferença entre os parâmetros do modelo de regressão quando um caso é excluído

Parâmetros b	Caso 30 incluído	Caso 30 excluído	Diferença
Constante (Intercepto)	+29,00	31,00	$-2,00$
Previsor (Inclinação)	$-0,90$	$-1,00$	$0,10$
Modelo (Linha de regressão)	$Y = (-0,9)X + 29$	$Y = (-1)X + 31$	
Y previsto	28,10	30,00	$-1,09$

Quadro 5.1

A diferença entre resíduos e estatísticas de influência ③

Nesta seção, descrevemos duas maneiras de procurar casos que podem prejudicar o modelo: resíduos e as estatísticas de influência. Para ilustrar como essas medidas diferem, imagine que o



prefeito de Londres do final do século passado estivesse interessado em como o ato de beber afetava a mortalidade. Londres é dividida em diferentes regiões denominadas municípios e, dessa forma, ele pode contar o número de bares e o número de mortes durante um período de tempo em oito dos seus municípios. Os dados estão em um arquivo denominado **pubs.sab**.

O diagrama de dispersão desses dados revela que sem o último caso, existe um relacionamento linear perfeito (a linha reta pontilhada). No entanto, a presença do último caso (caso 8) muda dramaticamente a linha de melhor ajuste (embora essa seja ainda um ajuste significativo dos dados – faça a análise de regressão e veja você mesmo).

É interessante olharmos para os resíduos e para as estatísticas de influência desses dados. O resíduo padronizado para o caso 8 é o segundo menor: esse valor atípico produz um resíduo bem pequeno (muitos dos dados típicos tem resíduos maiores) porque ele está bastante próximo da linha que foi ajustada aos dados. Como isso pode acontecer? Olhe para as estatísticas de influência e verá que elas são enormes para o caso 8, exercendo uma influência muito grande no modelo.

Como sempre, devemos fazer quando nos deparamos com estatísticas estranhas, é preciso perguntar o que está acontecendo no mundo real. O último ponto de dados representa a cidade de Londres, uma pequena área de apenas uma milha quadrada no centro de Londres onde poucas pessoas vivem, mas onde milhares de trabalhadores vem trabalhar e almoçam nos bares (*pubs*). Portanto, os bares não dependem de uma população que reside no local para fazer negócios e essa população não consumiu toda a sua cerveja! Dessa forma, havia um grande número de bares.

Isso ilustra que um caso que influencia muito pode produzir um pequeno resíduo – portanto, olhe ambos! (Agradeço a David Hitchin por esse exemplo, que ele obteve do Dr. Richard Roberts.)

Case summaries^a (Resumo dos casos)

	Standardized Residual (Resíduos padronizados)	Centred Leverage Value (Valor da Influência Centrado)	Standardized DFFit (DFFit padronizados)	Standardized DFBETA Intercept (Intercepto DFBETA padronizado)	Standardized DFBETA PUBS (DFBETA Bares padronizado)
1	-1.33839	0,04074	-0,74402	-0,74317	0,36886
2	-0.87895	0,03196	-0,40964	-0,40766	0,18484
3	-0.41950	0,02424	-0,17697	-0,17494	0,07132
4	0,3995	0,01759	0,01606	0,01572	-0,00564
5	0,49940	0,01200	0,20042	0,19337	-0,05933
6	0,95885	0,00748	0,40473	0,38333	-0,09618
7	1,41830	0,00402	0,68084	0,62996	-0,12023
8	-0,27966	0,86196	-460379232,7	92676016,019	-430238878,2
Total N	8	8	8	8	8

a Limitado aos primeiros 100 casos.

- Se $CVR_i < 1 - [(3(k+1)/n)]$, então excluir o i -ésimo caso irá melhorar a precisão de alguns dos parâmetros do modelo.



Nas duas expressões, k é o número de previsores do modelo, CVR_i é a razão da covariância para o i -ésimo participante e n é o tamanho da amostra.

5.6.1.3 Um comentário final sobre estatísticas diagnóstico ②

Existem muitas estatísticas diagnóstico que devem ser examinadas após a execução de uma análise de regressão e é difícil resumir essa riqueza de material em uma conclusão concisa. Contudo, eu gostaria de salientar um ponto discutido por Belsey e colaboradores (1980), que notou o perigo inerente nesses procedimentos. O problema é que os diagnósticos são recursos que permitem ver quão bom ou ruim é o seu modelo em termos da aderência aos dados da amostra. Eles são uma forma de avaliar o seu modelo. Eles não são, contudo, uma forma de justificar a remoção de dados que afetam algumas mudanças desejáveis nos parâmetros da regressão (por exemplo, eliminar um caso que muda um valor b não-significativo para um significativo). Stevens (1992), como sempre, oferece um excelente conselho:

Se um ponto é um significativo valor atípico em Y , mas a sua distância de Cook é menor do que um, não existe motivo real para eliminar esse ponto, uma vez que ele não apresenta um grande efeito na análise de regressão. Contudo, existe ainda o interesse de estudar tal ponto para tentar entender por que ele não está aderindo ao modelo (p. 118).

5.6.2 Interpretando o modelo de regressão II: generalização ②

Quando uma análise de regressão é feita, uma equação correta para os valores da amostra observada pode ser produzida. Entretanto, nas ciências sociais normalmente estamos interessados em generalizar nossas descobertas

para além da amostra observada. Assim, embora possa ser útil tirar conclusões sobre uma amostra particular de pessoas, geralmente é mais interessante se pudermos assumir que nossas conclusões são verdadeiras para uma população mais ampla. Para generalizar um modelo de regressão, devemos estar seguros de que as suposições foram satisfeitas, e para testar se o modelo de fato é generalizável, podemos fazer uma validação cruzada.

5.6.2.1 Avaliando hipóteses ②

Para tirar conclusões sobre uma população com base em um modelo de regressão realizado sobre uma amostra, várias suposições (hipóteses) devem ser verdadeiras (veja Berry, 1993).

- **Tipos de variáveis:** Todas as variáveis previsoras devem ser quantitativas ou categóricas (com duas categoriais), e a variável de saída deve ser quantitativa, contínua e não-limitada. Quantitativa significa que ela deve ser mensurada por intervalos e não-limitada quer dizer que não devem existir restrições na variabilidade da saída. Se a saída é uma medida que varia de 1 a 10 e os dados coletados variam entre 3 e 7, então esses dados são restritos.
- **Variância não-nula:** Os previsores devem ter alguma variação nos valores (isto é, eles não devem ter variância zero).
- **Multicolinearidade não deve ser perfeita:** Não deve existir relacionamento linear perfeito entre dois ou mais previsores. Assim, as variáveis previsoras não devem apresentar correlações muito altas (veja a Seção 5.6.2.4).
- **Previsores não-correlacionados com “variáveis externas”:** *Variáveis externas* são variáveis que não foram incluídas no modelo de regressão e que influenciam a variável de saída.⁸ Essas variáveis podem

⁸ Alguns autores se referem a essas variáveis externas como parte de um termo erro que inclui qualquer fator aleatório no modo pelo qual a saída varia. Contudo, para evitar confusão com os termos resíduos na equação de regressão, escolhi denominá-las “variáveis externas”. Embora esse termo implicitamente anule qualquer fator aleatório, eu reconheço esses fatores aqui!

ser pensadas como semelhantes a uma “terceira variável” que foi discutida na correlação. Essa suposição significa que não devem existir variáveis externas que se correlacionam com qualquer uma das variáveis incluídas no modelo de regressão. Obviamente, se as variáveis externas se correlacionam com os previsores, as conclusões feitas com o modelo não serão confiáveis (porque outras variáveis também podem prever a saída de forma precisa).

- **Homocedasticidade:** A cada nível das variáveis predictoras, a variância do termo residual deve ser constante. Isso significa que os resíduos a cada nível dos previsores devem ter a mesma variância (homocedasticidade): quando as variâncias são desiguais diz-se que existe heterocedasticidade (veja a Seção 3.6 também).
- **Erros independentes:** Para quaisquer duas observações os termos resíduos devem ser não-correlacionados (ou independentes). Essa eventualidade é algumas vezes descrita como falta de **autocorrelação**. Essa suposição pode ser verificada com o **teste de Durbin-Watson**, que testa a correlação serial entre erros. Especificamente, ele testa se resíduos adjacentes são correlacionados. A estatística teste pode variar entre 0 e 4, com 2 significando que os resíduos não são correlacionados. Um valor maior do que 2 indica correlação negativa entre resíduos adjacentes e um valor abaixo de 2 indica uma correlação positiva. O tamanho da estatística de Durbin-Watson depende do número de previsores do modelo e do número de observações. Para ser acurado, você deve olhar os valores aceitáveis exatos no artigo original de Durbin e Watson (1951). Uma regra bastante conservadora, afirma que valores menores do que 1 ou maiores do que 3 são definitivamente motivos de preocupação; contudo, valores próximos de 2 também podem ser problemáticos dependendo da sua amostra e modelo.
- **Erros normalmente distribuídos:** Assume-se que os resíduos em um modelo

são variáveis aleatórias, normalmente distribuídas com média zero. Essa hipótese significa que as diferenças entre o modelo e os dados observados são com mais frequência zero ou muito próximas a zero e que diferenças muito maiores do que zero acontecem apenas ocasionalmente. Algumas pessoas confundem essa hipótese com a ideia de que os previsores têm que ser normalmente distribuídos. Na verdade, os previsores não precisam ser normalmente distribuídos (veja a Seção 5.10).

- **Independência:** Assume-se que todos os valores da variável de saída são independentes (em outras palavras, cada valor da variável de saída provém de uma entidade separada).
- **Linearidade:** Os valores médios da variável de saída para cada incremento nos previsores devem estar sobre a linha. Em outras palavras, isso significa que é assumido que o relacionamento que estamos modelando é do tipo linear. Se modelarmos um relacionamento não-linear utilizando um modelo linear, isso obviamente limita a generalização do que encontrarmos.

Essa lista de hipóteses parece desestimulante e, de fato, muitos alunos de graduação (e alguns professores também) tendem a considerar as hipóteses como um assunto tedioso com o qual ninguém

deveria ser preocupar. Quando menciono as suposições estatísticas para psicólogos, eles tendem a me lançar um olhar do tipo “você é realmente pedante” e então me ignoram. Contudo, existem boas razões para levar as suposições a sério. Imagine que eu vou à casa de um amigo e as luzes estão acesas e é óbvio que alguém está em casa. Eu toco a campainha da porta e ninguém atende. A partir dessa experiência, concluo que o meu amigo me odeia e que sou uma pessoa terrível e mal-amada. Até que ponto essa conclusão é verdadeira? Bem, existe uma realidade que estou tentando ana-

Por que devo me preocupar com suposições?



lisar (isto é, se o meu amigo gosta ou não de mim) e eu coletei dados sobre essa realidade (eu fui à casa dele, vi que ele está em casa, toquei a campainha e não obtive resposta). Imagine que na verdade o meu amigo gosta de mim (ele nunca soube escolher bem os seus amigos!); neste cenário, minha conclusão é falsa. Por que os meus dados me levaram a uma conclusão errada? A resposta é simples: eu assumi que a campainha da porta do meu amigo estava funcionando e sob essa suposição a conclusão que fiz a partir dos meus fatos foi precisa (meu amigo ouviu a campainha, mas escolheu me ignorar porque ele me odeia). Contudo, essa suposição não era verdadeira – sua campainha não estava funcionando e foi por isso que ele não respondeu – e, como consequência, a conclusão que fiz sobre a realidade foi completamente falsa.

Chega de campainhas, amigos e minha vida social: o ponto a ser lembrado é que, quando as hipóteses são ignoradas, nós paramos de tirar conclusões válidas sobre a realidade. Em termos de regressão, quando as suposições são consideradas, o modelo que obtemos de uma amostra pode ser aplicado de forma precisa para a população de interesse (os coeficientes da equação de regressão não são *tendenciosos*). Algumas pessoas assumem que isso significa que quando as hipóteses são satisfeitas, o modelo de regressão de uma amostra é sempre idêntico ao modelo que seria obtido se fôssemos capaz de testar toda a população. Infelizmente, essa crença não é verdadeira. O que um modelo não tendencioso nos diz é que, em *média*, o modelo de regressão obtido a partir de uma amostra é o mesmo que o modelo populacional. Entretanto, deve ficar claro que mesmo quando as suposições são satisfeitas, é possível que um modelo obtido a partir de uma amostra não seja igual ao modelo populacional – mas a probabilidade de serem idênticos aumenta.

5.6.2.2 Validação cruzada do modelo ③

Mesmo se não pudermos ter certeza de que o modelo derivado da nossa amostra re-

presenta de forma precisa toda a população, existem maneiras de determinar quão bem nosso modelo pode prever a saída em uma amostra diferente. Determinar a precisão de um modelo entre diferentes amostras é conhecido como **validação cruzada**. Se um modelo pode ser generalizado, ele deve ser capaz de prever de modo preciso a mesma variável de saída a partir do mesmo conjunto de previsores em um grupo de pessoas diferentes. Se o modelo é aplicado a uma amostra distinta e existe uma grande diferença na sua capacidade de previsão, então o modelo claramente não é generalizável. Como uma primeira regra prática, nós devemos coletar dados suficientes para obter um modelo de regressão confiável (veja a próxima seção). Uma vez determinado o modelo de regressão, existem dois métodos principais de validação cruzada:

- **R^2 ajustado**: No SPSS, não apenas os valores de R e R^2 são calculados, mas também um R ajustado. Esse valor ajustado significa a perda do poder de previsão ou *encolhimento*. O R^2 informa quanto da variância de Y pode ser creditada ao modelo de regressão amostral, e o valor ajustado nos informa quanta variância de Y pode ser creditada ao modelo se ele tiver sido derivado da população de onde a amostra foi retirada. O SPSS determina o valor R^2 ajustado utilizando a equação de Wherry. Contudo, essa equação tem sido criticada porque nada informa sobre quão bem o modelo de regressão preveria um conjunto de dados inteiramente diferente (quão bem o modelo prevê escores de diferentes amostras retiradas da mesma população?). Uma versão do R^2 que nos informa quão bem o modelo valida de forma cruzada utiliza a fórmula de Stein, mostrada na equação (5.11) (veja Stevens, 1992):

$$R_2 \text{ ajustado} =$$

$$1 - \left[\left(\frac{n-1}{n-k-1} \right) \left(\frac{n-2}{n-k-2} \right) \left(\frac{n+1}{n} \right) \right] (1 - R^2) \quad (5.11)$$

Na equação de Stein, R^2 é o valor não ajustado, n é o número de casos e k é o número de previsores no modelo. Para os que se dão bem com a matemática, vale à pena utilizar essa equação para fazer a validação cruzada do modelo.

- **Divisão dos dados:** Essa abordagem envolve dividir ao acaso o conjunto de dados em dois, determinar a equação de regressão em cada uma das duas metades e depois comparar os modelos resultantes. Contudo, pesquisadores raramente conseguem conjuntos de dados grandes o suficiente para executar esse tipo de análise.

5.6.2.3 Tamanho da amostra na regressão ③



Na seção anterior, comentei que é importante coletar dados suficientes para obter um modelo de regressão confiável. Mas quanto é o suficiente? Há muitas regras práticas por aí; as duas mais comuns são que você deve ter 10 casos de dados para cada pre-

visor no modelo ou 15 casos de dados por previsor. Assim, se você tivesse cinco previsores, seria necessário 50 ou 75 casos, respectivamente (dependendo da regra que você escolhesse). Essas regras são bastante difundidas (eu mesmo utilizei a regra dos 15 casos por previsor na primeira edição do livro), mas elas simplificam demais o assunto. Na verdade, o tamanho da amostra necessário irá depender do tamanho de efeito que estamos tentando detectar (isto é, quão forte é o relacionamento que estamos tentando medir?) e de quanto poder queremos para detectar esses efeitos (veja o Capítulo 1). A regra mais simples é: quanto maior a amostra melhor! A razão é que a estimativa de R que obtemos da regressão é dependente do número de previsores, k , e do tamanho da amostra, N . O R esperado para dados aleatórios é $k/(n - 1)$ e, dessa forma, com amostras pequenas os dados aleatórios podem parecer mostrar um efeito forte; por exemplo, com seis previsores e 21 casos de dados, $R = 6/(21 - 1) = 0,3$ (um ta-

manho de efeito médio segundo o critério de Cohen descrito na Seção 4.2.2). Obviamente, para dados aleatórios esperaríamos que R fosse zero (nenhum efeito), e para isso ser verdadeiro, precisamos de grandes amostras (tomando o exemplo anterior, se tivéssemos 100 casos em vez de 21, o R esperado seria 0,06, um valor bem mais aceitável).

É ótimo saber que quanto mais, melhor, mas os pesquisadores em geral precisam de orientações mais objetivas (adoraríamos coletar 1000 casos de dados, porém, nem sempre isso é possível!). Green (1991) apresenta duas regras práticas para o tamanho mínimo aceitável de uma amostra, o primeiro tomando por base o teste do modelo como um todo (isto é, testando o R^2) e o segundo tomando por base os testes individualizados dos previsores do modelo (isto é, testando os valores b do modelo). Se você quer testar o modelo como um todo, ele recomenda um tamanho mínimo de amostra de $50 + 8k$, onde k é o número de previsores. Assim, se você tiver cinco previsores, irá precisar de um tamanho de amostra de $50 + 40 = 90$. Se você quiser testar os previsores individualmente, ele sugere um tamanho mínimo de amostra de $104 + k$; assim, novamente, se tivermos cinco previsores, iremos precisar de uma amostra mínima de $104 + 5 = 109$ elementos. É claro que o na maioria dos casos estamos interessados tanto em testar o modelo quanto a contribuição dos previsores (valores b), nesse caso, Green recomenda que você calcule o tamanho da amostra mínima (que acabei de descrever) pelas duas formas que foram apresentadas e adote aquela que fornecer o valor mais alto (portanto, no caso de cinco previsores, usaríamos 109 em vez de 90).

Essas recomendações são um bom guia, mas elas continuam simplificando o problema. Como já foi mencionado, o tamanho necessário da amostra depende do tamanho de efeito (isto é, quão bem nossos previsores prevêm a saída) e de com quanto poder estatístico queremos detectar esses efeitos. Miles e Shevin (2001) fornecem alguns gráficos úteis que ilustram o tamanho da amostra necessário para obter diferentes níveis de poder para diferentes tamanhos

de efeito quando o número de previsores varia. Para uma estimativa precisa do tamanho da amostra que você deve utilizar, recomendo o uso desses gráficos. Resumi alguns conselhos gerais de Miles e Shevin na Figura 5.9. Esse diagrama mostra o tamanho da amostra necessário para obter altos níveis de poder (tomei por base o valor de 0,8, proposto por Cohen, 1998) dependendo do número de previsores e do tamanho de efeito esperado. Algumas das recomendações de Miles e Shevin são: (1) se você espera encontrar um grande efeito, então uma amostra de tamanho 80 será suficiente (com até 20 previsores), e se existem menos previsores, então você poderá se dar o luxo de ter amostras de tamanhos ainda menores; (2) se você estiver esperando um efeito médio, um tamanho de amostra de 200 será suficiente para até 20 previsores, mas você deve sempre ter um tamanho de amostra acima de 60, e com seis ou menos previsores, uma amostra de tamanho 100 será suficiente; e (3) se você estiver esperando um tamanho de efeito pequeno, vai precisar ter tempo e recursos para coletar pelo menos 600 casos de dados (e muito mais se você tiver seis ou mais previsores!).

5.6.2.4 Multicolinearidade ②

A multicolinearidade existe quando observamos uma forte correlação entre dois ou mais previsores em um modelo de regressão. A multicolinearidade cria um problema apenas para a regressão múltipla porque (sem querer dizer o óbvio) a regressão simples tem apenas um preditor. Uma **colinearidade perfeita** existe quando pelo menos um preditor é uma combinação linear perfeita de outros (o exemplo mais simples é quando dois preditores estão perfeitamente correlacionados – eles têm um coeficiente de correlação igual a 1). Se existe uma colinearidade perfeita entre preditores, torna-se impossível obter estimativas únicas dos coeficientes de regressão porque existe um número infinito de combinações de coeficientes que funcionarão igualmente bem. Simplificando, se tivermos dois preditores que são perfeitamente correlacionados, os valores de b de cada variável podem ser trocados. A boa notícia é que a colinearidade perfeita é rara com dados reais. A má notícia é que uma colinearidade não-perfeita é prati-

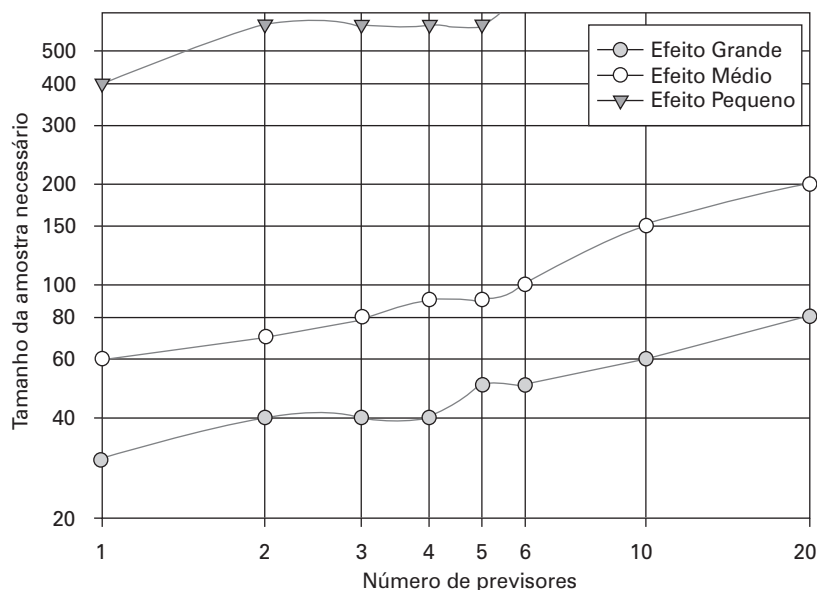


Figura 5.9 Diagrama mostrando o tamanho da amostra necessário na regressão em função do número de previsores e do tamanho de efeito esperado.

camente inevitável. Níveis baixos de colinearidade não representam uma ameaça para os modelos gerados pelo SPSS, mas à medida que a colinearidade aumenta, também aumentam os erros padrões dos coeficientes b , que por sua vez afetam a significância estatística desses coeficientes. Resumindo, altos níveis de colinearidade aumentam a probabilidade de que um bom preditor da variável de saída seja declarado não-significativo e excluído do modelo (Erro do Tipo II). Existem três outras razões por que a presença de multicolinearidade representa uma ameaça à validade de uma análise de regressão múltipla:

- **Ela limita o tamanho do R:** Lembre que R é uma medida de correlação múltipla entre os preditores e a saída e que R^2 indica a variância da saída que é de responsabilidade dos preditores. Imagine uma situação em que uma única variável prevê a variável de saída com um sucesso razoável (por exemplo, $R = 0,80$) e, depois, uma segunda variável preditora é adicionada ao modelo. Essa segunda variável deve ser responsável por boa parte da variância dos resultados (que é o motivo de ela ter sido acrescentada ao modelo), mas a variância que pode ser debitada a ela é a mesma da primeira variável. Em outras palavras, uma vez que a variância debitada à primeira variável tenha sido removida, o segundo preditor torna-se responsável por muito pouco da variância restante (a segunda variável apresenta uma variância própria muito pequena). Assim, a variável total dos resultados que pode ser debitada aos dois preditores é um pouco maior apenas do que quando um único preditor estava no modelo (assim, R pode aumentar de 0,80 para 0,82). Essa ideia está conectada à noção de correlação parcial que foi explicada no Capítulo 3. Se, contudo, os dois preditores são completamente não-correlacionados, o segundo preditor tem uma probabilidade maior de ser responsável por uma variância dos resultados do que quando existia um único preditor. Dessa forma, embora por si só o segundo

preditor seja responsável apenas por uma pequena parte da variância dos resultados, a variância debitada aos dois é diferente do que quando tinha apenas um preditor (portanto, quando ambos estão no modelo, R aumenta substancialmente, por exemplo, para 0,95). Desse modo, preditores não-correlacionados são desejáveis.

- **Importância dos preditores:** A multicolinearidade entre preditores dificulta a avaliação da importância individual de um preditor. Se os preditores são altamente correlacionados e cada um é responsável por uma variância similar dos resultados, como podemos saber qual entre duas variáveis é mais importante? Simples: não podemos dizer qual é a mais importante, o modelo pode incluir qualquer uma delas de forma intercambiável.
- **Equações com preditores instáveis:** Descrevi como a multicolinearidade aumenta as variâncias dos coeficientes de regressão, resultando em equações com preditores instáveis. Isso significa que um valor estimado dos coeficientes de regressão (os valores b) será instável de amostra para amostra.

Uma forma de identificar multicolinearidade é examinar a matriz de correlações de todas as variáveis predictoras e ver se alguma se correlaciona de forma bastante alta (por alta queremos dizer acima de 0,80 ou 0,90). Esse é um método prático, mas deixa escapar formas mais sutis de multicolinearidade. Felizmente, o SPSS produz vários diagnósticos de colinearidade, entre eles o FIV (**Fator de Inflação da Variância**). O FIV indica se um preditor tem um relacionamento linear forte com outro(s) preditor(es). Embora não existam regras simples sobre quais valores do FIV devem ser motivo de alerta, Meyers (1990) sugere que 10 é um bom valor a partir do qual podemos nos preocupar. Bowerman e O'Connell (1990) sugerem que se na média o FIV é maior do que 1, a multicolinearidade pode tornar o modelo de regressão tendencioso. Relacionado ao valor FIV existe a estatística **tolerância**, que é o valor inverso de FIV ($1/\text{FIV}$). Assim, valores

abaixo de 0,10 indicam problemas sérios, embora Menard (1995) sugira que valores abaixo de 0,20 já sejam motivo de preocupação.

Outra medida útil na descoberta de valores de previsores dependentes são os *autovalores da matriz dos produtos cruzados não reduzida e não centrada*, o *índice de condição* e as *proporções da variância*. Essas estatísticas são extremamente complexas e serão abordadas na parte de interpretação das saídas do SPSS (veja a Seção 5.8.5). Se nada disso faz sentido para você, dê uma olhada em Hutcheson e Sofroniou (1999, p. 78-85), que fornece uma explicação bastante clara da multicolinearidade.

5.7 COMO EXECUTAR UMA REGRESSÃO MÚLTIPLA NO SPSS ②

5.7.1 Principais opções ②

Imagine que um executivo de uma gravadora está interessado em estender o modelo da venda de discos para incorporar outras variáveis. Ele decide mensurar duas novas variáveis: (1) o número de vezes que o disco toca na Radio 1 (a maior estação de rádio da Grã-Bretanha) durante a semana anterior ao lançamento (**airplay**) e (2) a atratividade da banda (**attract**). Antes de um disco ser lançado, o executivo anota a quantia gasta

em publicidade, o número de vezes que o disco é tocado na rádio na semana anterior ao seu lançamento e a atratividade da banda. Ele faz isso para 200 tipos diferentes de discos (cada um de uma banda diferente). A atratividade é mensurada solicitando a uma amostra aleatória da audiência pretendida que atribua uma nota variando de 0 (insossos) a 10 (maravilhosos) para cada banda. A moda da atratividade atribuída para cada banda foi utilizada na regressão (porque ele estava interessado no que a maioria das pessoas pensava, em vez da opinião média das pessoas).


Esses dados estão no arquivo **Record2.sav** e você deve perceber que cada variável tem sua própria coluna (o mesmo método utilizado para a correlação) e que cada linha representa um disco diferente. Assim, o primeiro disco apresenta £10 269 de gastos em publicidade, vendeu 330 000 cópias, foi executado 43 vezes na Radio 1 na semana anterior ao seu lançamento e a maioria das pessoas julgou a banda maravilhosa (Figura 5.10).



O executivo tem pesquisas anteriores indicando que a quantia investida em publicidade é um preditor significativo da venda de discos e, assim, ele deve incluir essa variável primeiro no modelo. Suas novas variáveis (**airplay** e **attract**) devem, dessa forma, entrar no modelo

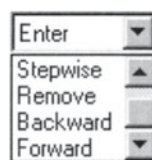
	adverts	sales	airplay	attract	v
1	10.26	330.00	43.00	10.00	
2	985.69	120.00	28.00	7.00	
3	1445.56	360.00	35.00	7.00	
4	1188.19	270.00	33.00	7.00	
5	574.51	220.00	44.00	5.00	
6	568.95	170.00	19.00	5.00	
7	471.81	70.00	20.00	1.00	

Figura 5.10 Apresentação dos dados para a regressão múltipla.

após a quantia investida em publicidade. Esse modelo é hierárquico (o pesquisador decide a ordem na qual as variáveis devem ser agregadas ao modelo com base em pesquisas anteriores). Para executar o modelo hierárquico no SPSS, temos que entrar as variáveis em blocos (cada bloco representando um passo na hierarquia). Para chegar à caixa de diálogo principal da regressão (*regression*), você deve clicar no menu *Analyze* (Analisar) e selecionar *Regression* (Regressão) e depois *Linear* (Linear) (**A**nalyze⇒**R**egression⇒**L**inear). A caixa de diálogo principal é mostrada na Figura 5.11 – ela é a mesma que foi utilizada na regressão simples.

A caixa de diálogo principal é bastante autoexplicativa: existe um espaço para especificar a variável dependente (saída ou resultado) e um espaço para colocar uma ou mais variáveis independentes de entrada (variáveis previsoras). As variáveis no editor de dados são listadas num painel à esquerda. Destaque ou marque a variável *record sales* (vendas de discos) na lista clicando sobre ela e a transfira para a caixa denominada *Dependent* (Dependente) clicando em . Também precisamos especificar a variável previsora para o primei-

ro bloco. Foi decidido que a quantia aplicada em publicidade deve ser a primeira variável a entrar no modelo (porque pesquisas anteriores indicam que ela é um previsor importante), assim, marque-a na lista e a transfira para o quadro à direita rotulado de *Independent(s)* (Independente(s)) clicando em . Na parte inferior do quadro das variáveis independentes (*Independent(s)*) existe um menu tipo lista suspensa onde é possível especificar o método (*Method*) da regressão (veja a Seção 5.5.3). Você pode selecionar um método diferente de entrada de variáveis para cada bloco clicando em , próximo de onde diz *Method* (Método). A opção por omissão (*default*) é a entrada forçada e essa é a opção que queremos, mas se você estiver executando um trabalho mais exploratório, pode escolher um dos métodos passo a passo (para a frente – *forward*, para trás – *backward*, passo a passo – *stepwise* ou remover – *remove*).



Uma vez especificado o primeiro bloco da hierarquia, precisamos nos mover para o segundo. Para informar ao computador que queremos especificar um novo bloco de va-

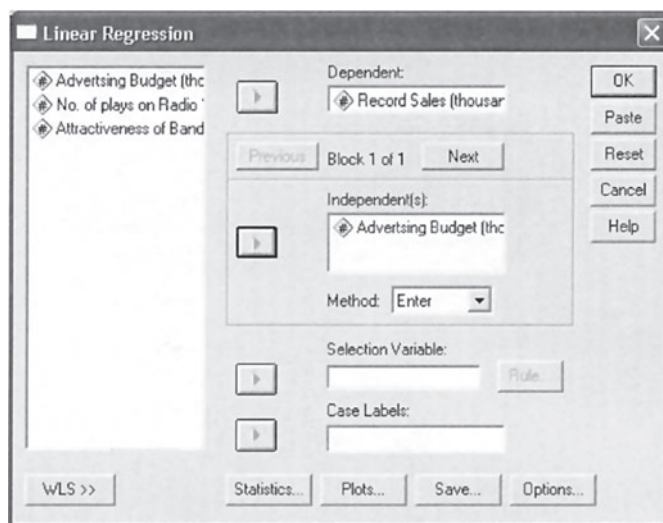


Figura 5.11 Caixa de diálogo principal para o primeiro bloco da regressão múltipla.

riáveis previsoras você deve clicar em **Next** (próximo). Esse processo limpa o quadro *Independent(s)* (Independente) de maneira que você pode entrar os novos previsores (você deve notar, também, que acima desse bloco está escrito agora *Block 2 of 2* (Bloco 2 de 2), indicando que você está no segundo bloco dos dois que foram especificados). Decidimos que o segundo bloco deveria conter os dois novos previsores, portanto, você deve clicar em **airplay** e **attract** na lista de variáveis e transferi-las, uma a uma, para o quadro rotulado de *Independent(s)* (Independente(s)) clicando em **►**. A caixa de diálogo fica semelhante a da Figura 5.12. Para se movimentar entre os blocos, utilize os botões **Previous** (prévio) e **Next** (próximo) (por exemplo, para voltar para o bloco 1, clique em **Previous**).

É possível selecionar métodos diferentes de entrada de uma variável para blocos diferentes em uma hierarquia. Assim, embora tivéssemos especificado a entrada forçada para o primeiro bloco, podemos agora especificar um método passo a passo para o segundo. Dado que não temos pesquisas anteriores em relação às variáveis *attractiveness* (atratividade) e *airplay* (número de reproduções na Rádio 1), em relação à vendas podemos justificar

a escolha pelo método passo a passo para esse bloco. Contudo, em virtude de problemas com o método passo a passo, optarei pelo de entrada forçada para os dois blocos neste exemplo.

5.7.2 Estatísticas ②

Na caixa de diálogo principal da regressão (*regression*), clique em **Statistics...** (estatísticas) para abrir a caixa de diálogo e selecionar várias opções importantes relacionadas ao modelo (Figura 5.13). Muitas dessas opções estão relacionadas aos parâmetros do modelo; contudo, existem procedimentos disponíveis para verificar a hipótese da não-existência de multicolinearidade (diagnóstico de colinearidade) e independência serial dos erros (Durbin-Watson). Quando você tiver selecionado as estatísticas necessárias (recomendo selecionar todas exceto a matriz de covariâncias como regra geral), clique em **Continue** (continue) para retornar à caixa de diálogo principal.

- **Estimates** (Estimativas): Essa opção é selecionada por omissão porque ela nos fornece os coeficientes estimados do modelo de regressão (isto é, estimativas dos valores b). A estatística teste e sua significância são fornecidas para cada coeficiente.

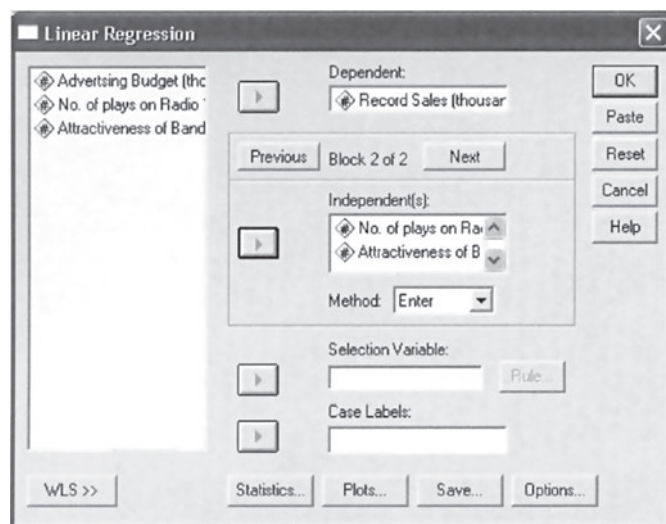


Figura 5.12 Caixa de diálogo principal para o segundo bloco da regressão múltipla.

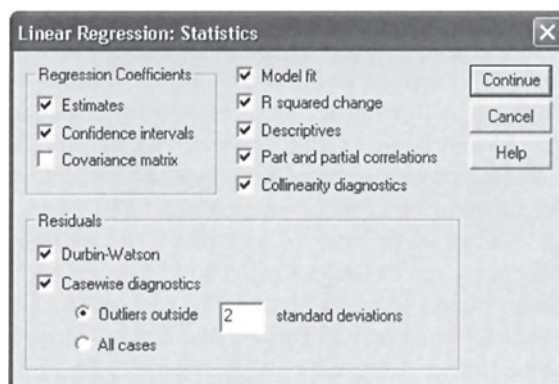


Figura 5.13 Caixa de diálogo *Statistics* (Estatísticas) para a análise de regressão.

te de regressão: um teste t é utilizado para verificar se cada b difere significativamente de zero (veja a Seção 5.2.4).

- **Confidence intervals** (Intervalos de confiança): Essa opção, se selecionada, produz intervalos de confiança para um dos coeficientes de regressão não-padronizados. Intervalos de confiança podem ser recursos úteis para avaliar a probabilidade dos coeficientes de regressão na população – irei descrever as interpretações exatas mais tarde.
- **Covariance matrix** (Matriz de covariâncias): Se selecionada, essa opção irá apresentar a matriz de covariâncias, os coeficientes de correlação e as variâncias entre os coeficientes de regressão para cada variável no modelo. Uma matriz de variâncias-covariâncias é produzida com as variâncias apresentadas na diagonal e as covariâncias apresentadas como elementos fora da diagonal principal. As correlações são apresentadas em uma matriz separada.
- **Model fit** (Aderência do modelo): Essa opção é vital e, portanto, ela é selecionada por omissão. Ela fornece não somente um teste estatístico da habilidade de o modelo prever a variável de saída (o teste F – veja a Seção 5.2.3), mas também o valor de R (ou R múltiplo), o R^2 correspondente e o R^2 ajustado.
- **R squared change** (Alterações no R ao quadrado): Essa opção apresenta as alte-

rações que ocorrem no R^2 resultantes da inclusão de um novo preditor (ou bloco de preditores). Essa medida é uma maneira útil de avaliar a contribuição do novo preditor (ou conjunto de preditores) na explicação da variância das saídas.

- **Descriptives** (Descritivas): Se selecionada, essa opção mostrará uma tabela com a média, o desvio padrão e o número de observações de todas as variáveis incluídas na análise. A matriz de correlações é também apresentada mostrando as correlações entre todas as variáveis e a probabilidade unilateral para cada coeficiente de correlação. Essa opção é extremamente útil porque a matriz de correlação pode ser utilizada para avaliar se os preditores estão inter-relacionados (o que pode ser utilizado para verificar se existe multicolinearidade).
- **Part and partial correlations** (Correlação parcial e por partes): Essa opção produz a correlação de ordem-zero (a correlação de Pearson) entre cada preditor e a saída, controlada para todos os outros preditores do modelo. Ele produz a correlação por partes (ou correlação semiparcial) entre cada preditor e a saída. Essa correlação representa o relacionamento entre cada preditor e a parte da saída que não é explicada pelos outros preditores no modelo. Como tal, ela mede o relacionamento único entre um preditor e a saída (veja a Seção 4.6).

- **Collinearity diagnostics** (Diagnóstico de colinearidade): Essa opção serve para obter as estatísticas de colinearidade, como VIF, tolerância, autovalores da matriz reduzida e não-centrada dos produtos cruzados e índice de condição e proporção de variâncias (veja a Seção 5.6.2.4).
- **Durbin-Watson**: Essa opção produz a estatística teste de Durbin-Watson, que testa a suposição de independência dos erros. Infelizmente, o SPSS não fornece o valor da significância desse teste, assim, você mesmo deve decidir se o valor é suficientemente diferente de 2 para ser motivo de preocupação (veja a Seção 5.6.2.1).
- **Casewise diagnostics** (Diagnósticos por casos): Essa opção, se selecionada, lista os valores observados da saída, os valores de saída previstos, a diferença entre esses valores (os resíduos) e essa diferença padronizada. Além disso, esses valores podem ser listados para todos os casos ou apenas para casos onde o resíduo padronizado for maior do que 3 (quando o sinal de \pm for ignorado). O critério de valor igual a 3 poderá ser alterado e eu recomendo alterá-lo para 2 por razões que logo ficarão evidentes. Uma tabela resumo de estatísticas residuais indicando o mínimo, o máximo, a média e o desvio padrão tanto dos valores previstos pelo modelo quanto dos resíduos (veja a Seção 5.8.6) também é produzida.

5.7.3 Diagramas da regressão ②

Uma vez que você esteja de volta à caixa de diálogo principal, clique em **Plots...** (Gráficos) para ativar a caixa de diálogo para os diagramas (*plots*) da regressão, mostrada na Figura 5.14. Essa caixa de diálogo fornece os meios para especificar vários gráficos, que podem auxiliar no estabelecimento da validade de algumas hipóteses da regressão. A maior parte desses gráficos envolvem vários valores residuais (*residual*), que serão descritos com mais detalhes na Seção 5.7.4.

No lado esquerdo da caixa há uma lista de diversas variáveis.

- **DEPENDNT** (a variável de saída).
- ***ZPRED** (os valores previstos padronizados da variável dependente com base no modelo). Esses valores são formas padronizadas dos valores previstos pelo modelo.
- ***ZRESID** (os resíduos padronizados, ou erros). Esses valores são as diferenças padronizadas entre os dados observados e os valores que o modelo prevê.
- ***DRESID** (os resíduos excluídos). Veja a Seção 5.6.1.1 para detalhes.
- ***ADJPRED** (os valores previstos ajustados). Veja a Seção 5.6.1.2 para detalhes.
- ***SRESID** (os resíduos estudentizados).
- ***SDRESID** (os resíduos estudentizados excluídos). Esse valor é o resíduo excluído dividido pelo erro padrão.

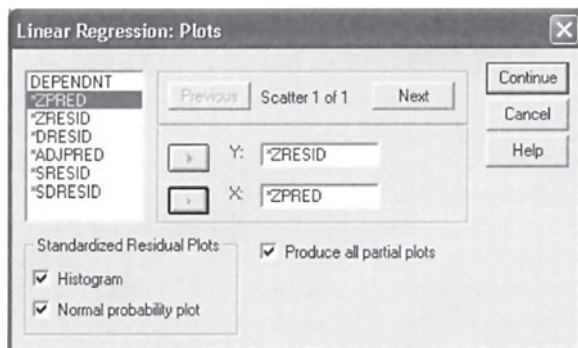







Figura 5.14 Regressão linear: caixa de diálogo gráfica.


As variáveis listadas nessa caixa de diálogo são todas apresentadas sob um título geral de resíduos e foram discutidas em detalhe na Seção 5.6.1.1. Para uma análise básica vale à pena obter os diagramas do *ZRESID (eixo y) contra *ZPRED (eixo x), porque esse gráfico é útil para determinar se as hipóteses de erros aleatórios e homocedasticidade foram satisfeitas. Um diagrama do *SRESID (eixo y) contra *ZPRED (eixo x) mostrará qualquer heterocedasticidade também. Embora muitas vezes esses dois diagramas sejam praticamente idênticos, o segundo é mais sensível em uma situação caso a caso. Para criar esses diagramas, basta selecionar uma variável da lista e transferi-la para o espaço denominado tanto *X* quanto *Y* (referentes aos eixos) clicando em . Quando você tiver selecionado as variáveis para o primeiro diagrama (neste caso, semelhante ao da Figura 5.14) você pode especificar um novo diagrama clicando em  (próximo). Esse procedimento limpa os espaços em que as variáveis foram especificadas. Se você clicar em  e quiser retornar ao último diagrama especificado, clique em  (anterior). Você pode especificar até nove diagramas.

É possível selecionar a caixa de marcar denominada *Produce all partial plots* (Produza todos os diagramas parciais), que irá determinar diagramas de dispersão dos resíduos da variável de saída e cada um dos previsores quando ambas as variáveis são analisadas separadamente com os previsores restantes. Mesmo que você não tenha entendido a frase anterior, esses diagramas apresentam várias características importantes que faz valer à pena inspecioná-los. Primeiro, o gradiente da linha de regressão entre duas variáveis residuais é equivalente ao coeficiente do preditor na equação de regressão. Assim, qualquer valor atípico óbvio em um diagrama parcial representa um caso que pode ter excessiva influência no coeficiente de regressão de um preditor. Segundo, relacionamentos não-lineares entre um preditor e a variável de saída são mais detectáveis utilizando esses gráficos. Finalmente, eles são formas úteis de detectar colinearidade. Por essas razões, recomendo que eles sejam utilizados.

Existem várias opções para apresentar graficamente os resíduos. Primeiro, você pode selecionar um histograma dos resíduos padronizados (isto é extremamente útil para verificar a hipótese da normalidade dos erros). Segundo, você pode solicitar um diagrama de probabilidade normal, que também informa se os resíduos do modelo são normalmente distribuídos. Quando você tiver selecionado as opções necessárias clique em  para voltar para a caixa de diálogo principal da regressão.

5.7.4 Salvando os diagnósticos da regressão ②

Na Seção 5.6, vimos dois tipos de diagnóstico da regressão: aqueles que nos ajudam a avaliar quão bem o modelo se ajusta a nossa amostra e aqueles que nos auxiliam a detectar casos que apresentam uma grande influência no modelo gerado. No SPSS, podemos escolher salvar essas variáveis diagnósticas no editor de dados (assim, o SPSS irá calculá-las e criar novas colunas no editor de dados em que os valores serão apresentados).

Para salvar os diagnósticos da regressão, você precisa clicar em  (salvar) na caixa de diálogo principal da regressão. Esse procedimento irá ativar a caixa de diálogo para salvar as variáveis (veja a Figura 5.15). Uma vez que a janela esteja ativa, basta marcar as opções desejadas nos quadradinhos à direita de cada uma. A maioria das opções disponíveis foram explicadas na Seção 5.6, e a Figura 5.15 mostra o que eu considero um conjunto básico de diagnóstico. Versões padronizadas (e estudentizadas) desses diagnósticos são geralmente mais fáceis de interpretar, assim, sugiro selecioná-las em vez das versões não-padronizadas. Uma vez que a regressão tenha sido executada, o SPSS irá criar uma coluna no editor de dados para cada uma das estatísticas solicitadas, e ele tem um conjunto padronizado de nomes dessas variáveis para nomear cada uma. Depois do nome existirá um número que se refere à análise que foi executada. Assim, na primeira vez que a regressão é executada, o conjunto de nomes das variáveis será seguido do valor 1; se você executar uma segunda

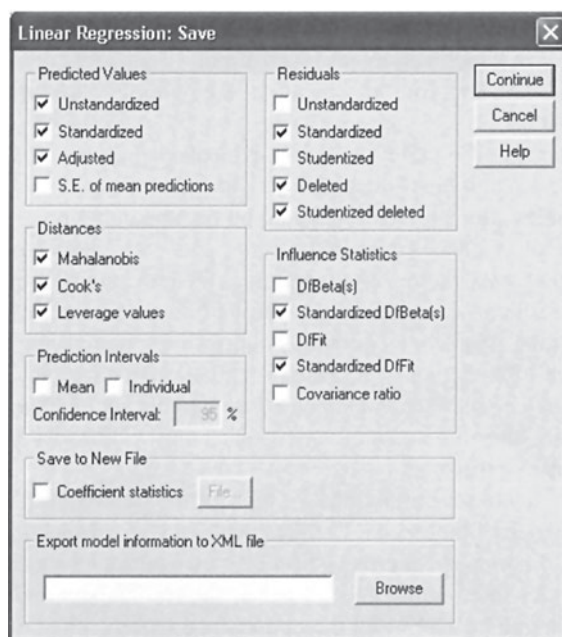


Figura 5.15 Caixa de diálogo para os diagnósticos da regressão.

regressão, ele irá criar um segundo conjunto de variáveis com o nome seguido pelo valor 2, e assim por diante. Os nomes das variáveis no editor de dados são apresentados a seguir. Depois que você selecionar os diagnósticos desejados (clique nos lugares adequados), clique em **Continue** para retornar à caixa de diálogo principal da regressão.

- **pre_1:** (valor previsto não-padronizado).
- **zpr_1:** (valor previsto padronizado).
- **adj_1:** (valor previsto ajustado).
- **sep_1:** (erro padrão do valor previsto).
- **res_1:** (resíduo não-padronizado).
- **zre_1:** (resíduo padronizado).
- **sre_1:** (resíduo estudentizado).
- **dre_1:** (resíduo excluído)
- **sdr_1:** (resíduo estudentizado excluído).
- **mah_1:** (distância de Mahalanobis).
- **coo_1:** (distância de Cook).
- **lev_1:** (valor centrado da influência).
- **sdb0_1** (DFBETA padronizado (intercepto)).
- **sdb1_1** (DFBETA padronizado (previsor 1)).

- **sdb2_1:** (DFBETA padronizado (previsor 2)).
- **sdf_1:** DFFIT padronizado.
- **cov_1:** (razão de covariância).

5.7.5 Opções adicionais ②

Como uma etapa final da análise você pode clicar em **Options...** para levá-lo à caixa de diálogo Opções (Figura 5.16). O primeiro conjunto de opções permite que você mude os critérios utilizados para entrar as variáveis na regressão passo a passo (*stepwise*). Se você insistir em realizar a regressão passo a passo, provavelmente é melhor manter o critério padrão de uma probabilidade de 0,05. Contudo, você pode tornar esse critério mais restrito (0,01). Existe outra opção para construir um modelo que não inclui a constante (isto é, não apresenta interseção com o eixo Y). Essa opção também deve ser ignorada! Finalmente, você pode selecionar um método para lidar com os pontos de dados omitidos (*missing*). Por padrão, o SPSS exclui casos por linhas (*listwise*), o que significa que se uma pessoa

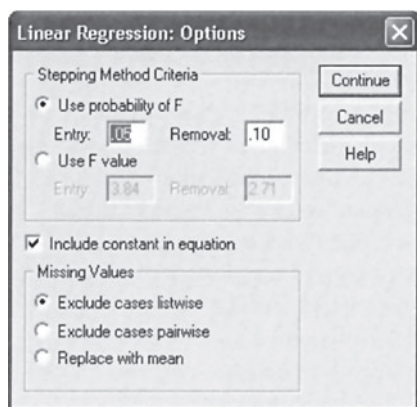


Figura 5.16 Opções para a regressão linear.

tem um valor não conhecido em uma variável, ela será excluída da análise. Assim, por exemplo, se o nosso executivo da gravadora não tem um escore para a atratividade de uma de suas bandas, os dados desse disco não serão utilizados no modelo de regressão. Outra opção é excluir casos em uma base de pares, o que significa que se o participante tem um escore desconhecido para uma variável em particular, seus dados serão excluídos da análise somente em cálculos envolvendo a variável para a qual ele não apresenta valor. Assim, o dado para a banda que não apresenta escore de atratividade seria utilizado para calcular o relacionamento entre valor investido em publicidade, número de vezes que o disco foi executado na Radio 1 e discos vendidos. Contudo, se você fizer isso, muitas das suas variáveis podem não fazer sentido, e talvez você acabe gerando absurdos como um R^2 tanto negativo quanto maior do que 1,0. Portanto, essa não é uma boa opção.

Outra possibilidade é substituir o valor que falta com um valor médio dessa variável e depois incluir o caso na análise (assim, no nosso exemplo a banda teria um valor de atratividade igual à medida da atratividade de todas as bandas). O problema dessa última escolha é que provavelmente o verdadeiro valor do desvio padrão será reduzido (e, ainda mais importante, o do erro padrão). O desvio padrão será diminuído porque para cada caso que for

substituído não existirá diferença entre a média e o escore a ele atribuído, enquanto que se o dado tivesse sido coletado para o tal caso ele quase certamente iria apresentar alguma diferença em relação à média. Obviamente, se a amostra é grande e o número de valores desconhecidos é pequeno, esse não é um problema sério. Contudo, se existirem muitos valores desconhecidos, essa escolha é potencialmente perigosa porque erros padrão pequenos terão maior probabilidade de resultar em valores significativos que serão o produto da substituição dos dados e não um efeito verdadeiro. A opção final é utilizar a rotina do SPSS “Análise de valores desconhecidos”, destinada para especialistas. Ela faz uso do fato de que se as suas variáveis estão presentes e correlacionadas para muitos casos no arquivo e um valor ocasional estiver faltando, você poderá substituí-lo por estimativas melhores do que a média (Tabachnick e Fidell, 2001, Capítulo 4, descrevem alguns desses procedimentos).

5.8 INTERPRETANDO A REGRESSÃO MÚLTIPLA ②

Uma boa estratégia para adotar com a regressão é medir as variáveis previsoras para as quais existem razões teóricas para esperar que prevejam bem o resultado. Execute a análise de regressão em que todos os previsores são colocados no modelo e examine a saída para ver quais previsores contribuem substancialmente para o modelo prever um resultado. Uma vez que você tenha determinado quais são as variáveis importantes, execute novamente a análise incluindo somente esses previsores e utilize as estimativas dos parâmetros resultantes para definir o modelo de regressão. Se a análise inicial revelar que existem dois ou mais previsores significativos, você pode considerar a execução de uma análise passo a passo para frente (*forward stepwise*) (em vez de uma entrada forçada) a fim de encontrar a contribuição individual de cada predictor.

Gastei um monte de tempo explicando a teoria de base da regressão e de algumas das ferramentas de diagnóstico necessárias para

medir a precisão de um modelo de regressão. É importante lembrar que o SPSS pode parecer bastante esperto, porém, de fato, ele não é. Ele pode, sim, realizar muitos cálculos complexos em questão de segundos, mas ele não consegue controlar a qualidade do modelo gerado – para tanto, é imprescindível um cérebro humano (e de preferência um que tenha sido treinado). O SPSS pode gerar facilmente saídas com base em qualquer tipo de lixo fornecido no editor de dados e ele não julgará julgar os resultados ou indicar se o modelo é generalizável ou mesmo válido. Contudo, o SPSS fornece as estatísticas necessárias para julgar essas coisas e a partir desse ponto nós devemos fazer o trabalho – o que é levemente preocupante (especialmente se o seu cérebro é tão pequeno quanto o meu!).

Tendo selecionado todas as opções relevantes e retornado à caixa de diálogo principal, precisamos clicar em **OK** para executar a análise. O SPSS exibirá inúmeras quantidades de saídas na janela visualizadora e nós precisamos descobrir como fazer com que essas informações façam sentido.

5.8.1 Descritivas ②

A saída descrita nesta seção é produzida utilizando a opção na caixa de diálogo *statistics* (estatísticas) da regressão linear (veja a Figura 5.13). Para começar, se você selecionou a opção *Descriptives* (Descritivas), o SPSS irá produzir a tabela visualizada na saída do SPSS 5.4. Essa tabela nos fornece a média e o desvio padrão de cada variável do nosso conjunto de dados, assim, sabemos que o número médio de discos vendidos foi de 193.200. Essa tabela não é necessária para interpretar o modelo de regressão, mas é útil como resumo dos dados. Além das estatísticas descritivas, a seleção

dessa opção fornece também uma matriz de correlações. A tabela mostra três coisas. Primeiro, é apresentado o valor do coeficiente de correlação de Pearson entre cada par de variáveis (por exemplo, podemos ver que o gasto em publicidade tem uma grande correlação positiva com a venda de discos, $R = 0,578$). Segundo, é mostrada a significância unilateral de cada correlação (por exemplo, a correlação acima é significativa, $p < 0,001$). Finalmente, é mostrado o número de casos que contribui com cada correlação ($N = 200$).

Você deve ter notado que ao longo da diagonal da matriz os valores para os coeficientes de correlação são todos iguais a 1,00 (isto é, uma correlação positiva perfeita). O motivo disso é que esses valores representam a correlação de cada uma das variáveis com ela mesma, assim, obviamente os valores resultantes são iguais a 1. A matriz de correlações é extremamente útil para fornecer uma ideia aproximada do relacionamento entre os previsores e a variável de saída e para um primeiro exame da multicolinearidade. Se não existir multicolinearidade nos dados, não deve existir valores de correlação substanciais ($R > 0,90$) entre os previsores.

Se olharmos apenas para os previsores (ignorando as vendas de discos), a correlação mais alta é entre a atratividade da banda e a quantidade de execução significativa ao nível de 0,01 ($R = 0,182$ e $p = 0,005$). A despeito da significância dessa correlação, o coeficiente é pequeno e, assim, parece que os previsores estão medindo coisas diferentes (não existe colinearidade). Podemos ver, também, que de todos os previsores, o número de execuções na Radio 1 é o que melhor se correlaciona com a saída ($R = 0,599$, $p < 0,001$); portanto, é provável que essa variável seja a melhor para prever a venda de discos.



Dica da Samanta Ferrinho

Use as estatísticas descritivas para verificar na matriz de correlações a multicolinearidade; isto é, os previsores altamente correlacionados entre si, $R > 0,90$.

Saída 5.4 do SPSS Estatísticas descritivas para a análise de regressão

Descriptive Statistics (Estatística Descritiva)

	Mean (Média)	Std. Deviation (Desvio padrão)	N
Record Sales (thousands) (Vendas de discos) (milhares)	193.2000	80.6990	200
Advertising Budget (Thousands of pounds) Verba de publicidade (Milhares de libras)	614.4123	485.6552	200
No. Of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	27.5000	12.2696	200
Attractiveness of Band (Atratividade da Banda)	6.7700	1.3953	200

Correlations (Correlações)

		Record Sales (thousands) (Vendas de discos) (milhares)	Advertising Budget (Verba de publicidade)	No. Of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	Attractiveness of Band (Atratividade da Banda)
Pearson Correlation (Correlação de Pearson)	Record Sales (thousands) (Vendas de discos) (milhares)	1.000	0.578	0.599	0.326
	Advertising Budget (Thousands of pounds) Verba de publicidade (Milhares de libras)	0.578	1.000	0.102	0.081
	No. of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	0.599	0.102	1.000	0.182
	Attractiveness of Band (Atratividade da Banda)	0.326	0.081	0.182	1.000
Sig. (1-tailed) (Sig.) (Unilateral)	Record Sales (thousands) (Vendas de discos) (milhares)	.	0.000	0.000	0.000
	Advertising Budget (Thousands of pounds) Verba de publicidade (Milhares de libras)	0.000	.	0.076	0.128
	No. of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	0.000	0.076	.	0.005
	Attractiveness of Band (Atratividade da Banda)	0.000	0.128	0.005	.
N	Record Sales (thousands) (Vendas de discos) (milhares)	200	200	200	200
	Advertising Budget (Thousands of pounds) Verba de publicidade (Milhares de libras)	200	200	200	200
	No. of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	200	200	200	200
	Attractiveness of Band (Atratividade da Banda)	200	200	200	200

5.8.2 Resumo do modelo ②

A próxima seção da saída descreve o modelo globalmente (ela informa se o mo-

delo é eficaz em prever a venda de discos). Lembre que escolhemos o método hierárquico e, portanto, cada conjunto de estatísticas-resumo é repetido para cada estágio na

Saída 5.5 do SPSS Resumo do modelo de regressão

Model Summary^c (Resumo do modelo)

Model (Modelo)	R	R Square (R Quadrado)	Adjusted R Square (R Quadrado Ajustado)	Std. Error of the Estimate (Erro Padrão da Estimativa)	Change Statistics (Estatísticas de Mudanças)					Durbin-Watson
					R Square Change (Mudança no R Quadrado)	F Change (Mudança no F)	df1 (gl1)	df2 (gl2)	Sig. F Change (Mudança na Sig. do F)	
1	0.578 ^a	0.335	0.331	65.9914	0.335	99.587	1	198	0.000	1.950
2	0.815 ^b	0.665	0.660	47.0873	0.330	96.447	2	196	0.000	

a. Previsores: (constante), Verba publicitária (em milhares de libras).

b. Previsores: (constante), Verba publicitária (em milhares de libras), Atratividade da Banda, Número de execuções por semana na Radio 1.

c. Variável dependente: Vendas de discos (em milhares).

hierarquia. Na Saída do SPSS 5.5, você deve notar que existem dois modelos. O Modelo 1 se refere ao primeiro estágio da hierarquia quando apenas a verba de publicidade é utilizada como um preditor. O Modelo 2 se refere a quando todos os três preditores são utilizados. A Saída do SPSS 5.5 é o *resumo do modelo* (model summary) e essa tabela foi gerada marcando-se a opção *Model fit* (Ajustar modelo). Essa opção é selecionada por omissão no SPSS porque ela nos fornece algumas informações importantes sobre o modelo: os valores de R , R^2 e o R^2 ajustado. Se as opções *R squared change* (Mudanças no R quadrado) e *Durbin Watson* forem selecionadas, esses valores também serão incluídos (se eles não forem selecionados, você terá uma tabela menor).

A tabela resumo é mostrada na Saída do SPSS 5.5, e você notará que ela informa qual era a variável dependente (outcome) e quais eram os preditores em cada um dos dois modelos. Na coluna denominada R estão os valores dos coeficientes de correlação múltipla entre os preditores e a saída. Quando somente a verba de publicidade é utilizada como um preditor, existe uma correlação simples entre os valores investidos em publicidade e a venda de discos (0,578). Na verdade todas as estatísticas do modelo 1 são as mesmas do modelo de regressão simples visto anteriormente (Seção 5.4). A próxima coluna fornece o valor de R^2 , que nós já conhecemos como uma medida

de quanta variabilidade da saída pode ser debitada aos preditores. Para o primeiro modelo esse valor é 0,335, o que significa que a verba de publicidade é responsável por 33,5% da variação nas vendas de discos. Contudo, quando os outros dois preditores são também incluídos (modelo 2), esse valor aumenta para 0,665 ou 66,5% da variância na venda de discos. Dessa forma, se a publicidade é responsável por 33,5%, podemos verificar que a atratividade e o número de execuções são responsáveis por mais 33%.⁹ Assim, a inclusão dos dois novos preditores explicou boa parte da variação das vendas de discos.

O R^2 ajustado fornece uma noção de quão bem nosso modelo generaliza, e idealmente nós gostaríamos que esse valor fosse igual ao, ou muito próximo do, valor do R^2 . Nesse exemplo, a diferença para o modelo final é pequena (de fato, a diferença entre os valores é $0,665 - 0,660 = 0,005$ ou aproximadamente 0,5%). Isso significa que se o nosso modelo fosse derivado da população em vez de uma amostra, ele explicaria aproximadamente 0,5% menos da variância da saída. Estudantes avançados podem querer aplicar a fórmula de Stein do R^2 para ter uma ideia de quão provável seria esse valor entre diferentes amostras. A fórmula de Stein foi apresentada na equação

⁹ Isto é, $33\% = 66,5\% - 33,5\%$ (esse valor é o *R Square Change* (alteração no R quadrado) apresentado na tabela).

(5.11) e pode ser aplicada pela substituição de n pelo tamanho da amostra (200) e k com o número de previsores (3):

R^2 ajustado = $1 -$

$$\left[\left(\frac{200-1}{200-3-1} \right) \left(\frac{200-2}{200-3-2} \right) \left(\frac{200+1}{200} \right) \right] \\ (1-0,665) \\ = 1 - [(1,015)(1,015)(1,005)] \\ (0,335) \\ = 1 - 0,347 \\ = 0,653$$

Esse valor é muito semelhante ao valor do R^2 observado, de 0,665, indicando que a validação cruzada do modelo é muito boa.

As estatísticas de mudança são fornecidas somente se solicitadas e elas nos informam se a mudança no R^2 é significativa. A significância do R^2 pode ser testada utilizando a razão F , e esse F é calculado a partir da seguinte equação (na qual N é o número de casos ou participantes e k é o número de previsores no modelo):

$$F = \frac{(N - k - 1)R^2}{k(1 - R^2)}$$

Na saída do SPSS 5.5, a mudança nesse F é relatada para cada um dos blocos da hierarquia. Assim, o modelo 1 causa uma mudança no R^2 de zero para 0,335, e essa mudança na quantidade de variância explicada dá origem a uma razão F de 99,587, que é significativa com uma probabilidade menor do que 0,001. Tenha em mente que para esse primeiro modelo temos apenas um preditor (assim, $k = 1$) e 200 casos ($N = 200$). Esse F é obtido da equação anterior:¹⁰

$$F_{\text{Model 1}} = \frac{(200 - 1 - 1)0,334648}{1(1 - 0,334648)} = 99,587$$

A adição de novos previsores (modelo 2) causa um aumento do R^2 por 0,330 (veja acima). Podemos calcular a razão F para essa mudança utilizando a mesma equação, mas porque estamos avaliando as mudanças no modelo, utilizaremos essas mudanças, R^2_{Change} , e o R^2 para o novo modelo (modelo 2; nesse caso, vamos chamá-lo de R^2_2). Também usamos a mudança no número de previsores, k_{Change} (o modelo 1 tem um preditor e o modelo 2 tem três previsores, assim, a mudança no número de previsores é $3 - 1 = 2$), e o número de previsores no novo modelo, k_2 (porque estamos interessados no modelo 2). Novamente, se utilizarmos algumas casas decimais a mais do que as que a Tabela do SPSS apresenta, iremos obter aproximadamente a mesma resposta obtida pelo SPSS:

$$F_{\text{Change}} = \frac{(N - k_2 - 1)R^2_{\text{Change}}}{k_{\text{Change}}(1 - R^2_2)} \\ = \frac{(200 - 3 - 1) \times 0,330}{2(1 - 0,664668)} \\ = 96,44$$

Desse modo, a mudança na variância que pode ser explicada fornece uma razão F de 96,44, que é novamente significativo ($p < 0,001$). As estatísticas de mudança, portanto, nos informam sobre as diferenças que ocorrem quando adicionamos novos previsores ao modelo.

Finalmente, se você solicitou a estatística de *Durbin-Watson*, ela será encontrada na última coluna da tabela da saída 5.5 do SPSS. Essa estatística nos informa se a hipótese de independência dos erros é satisfeita (veja a Seção 5.6.2.1). Como uma regra conservadora, sugiro que valores menores do que 1 ou maiores do que 3 devam, definitivamente, ser motivos de preocupação (embora eu o aconselhe a determinar valores precisos para a situação de interesse). Quanto mais próximo de 2 o valor estiver, melhor; para esses dados o valor

¹⁰ Para obter o mesmo valor que o SPSS, precisamos utilizar o valor exato do R^2 , que é, 0,3346480676231 (se você não acredita dê um clique duplo na tabela de saída do SPSS que apresenta esse valor, depois dê um clique duplo na célula da tabela contendo o valor do R^2 e você verá que 0,335 torna-se o valor que foi digitado acima!).

é 1,950, bem perto de 2, ou seja, a hipótese quase com certeza está satisfeita.

A saída 5.6 do SPSS mostra a próxima parte da saída, que contém uma análise de variância (ANOVA) que testa se o modelo é significativamente melhor para prever a saída do que utilizar a média como um “bom palpite”. Especificamente, F representa a razão de melhoria na previsão que resulta do ajuste do modelo em comparação com a imprecisão que ainda existe no modelo. Somos informados do valor da soma dos quadrados do modelo (esse valor é o SS_M da Seção 5.2.3 e representa a melhoria na previsão resultante do ajuste de uma linha aos dados em vez de utilizar a mé-

dia como uma estimativa das saídas). Recebemos ainda o valor da soma dos quadrados dos resíduos (esse valor é o SS_R da Seção 5.2.3 e representa a diferença total entre o modelo e os dados observados). Somos apresentados ainda aos graus de liberdade (gl) para cada termo. No caso de melhoria ocasionada pelo modelo, esse valor é igual ao número de previsores (um para o primeiro modelo e três para o segundo), e para o SS_R ele é o número de observações (200) menos o número de coeficientes no modelo de regressão. O primeiro modelo tem dois coeficientes (um para a variável previsora e outro para a constante), enquanto que o segundo tem quatro (um para cada variável

Saída 5.6 do SPSS

ANOVA^c (Análise de Variância)

	<i>Model</i> (Modelo)	<i>Sum of squares</i> (Soma dos quadrados)	<i>df</i> (gl)	<i>Mean Square</i> (Média dos quadrados)	<i>F</i>	<i>Sig.</i>
1	<i>Regression</i> (Regressão)	433687.833	1	433687.833	99.587	0.00 ^a
	<i>Residual</i> (Resíduo)	862264.167	198	4354.870		
	<i>Total</i> (total)	1295952.0	199			
2	<i>Regression</i> (Regressão)	861377.418	3	287125.806	129.498	0.00 ^b
	<i>Residual</i> (Resíduo)	434574.582	196	2217.217		
	<i>Total</i> (total)	1295952.0	199			

a. Predictors: (Constant), Advertising Budget (thousands of pounds) (Previsores: (constante), Verba publicitária (em milhares de libras))

b. Predictors: (Constant), Advertising Budget (thousands of pounds), Attractiveness of Band, No. of plays on Radio 1 per week (Previsores: (constante), Verba publicitária (em milhares de libras)), Atratividade da Banda, Número de execuções por semana na Radio 1.

c. Dependent variable: Record Sales (thousands) (Variável dependente: Vendas de discos (em milhares))

Dica da Samanta Ferrinho



O ajuste do modelo de regressão pode ser avaliado utilizando o **resumo do modelo (Model Summary)** e as tabelas da ANOVA do SPSS. Um exame do R^2 informará a proporção da variância que é explicada pelo modelo. Se você tiver feito uma regressão hierárquica, poderá avaliar a melhoria do modelo em cada estágio da análise analisando as mudanças no R^2 e se essa mudança é significativa (procure por valores menores do que 0,05 na coluna denominada *Sig. F Change* (Mudança na significância do F)). A ANOVA nos informa também se o modelo adere de forma significativa aos dados (procure por valores menores do que 0,05 na coluna denominada *Sig.*). Finalmente, existe uma suposição de que os erros na regressão são independentes; essa suposição é provavelmente satisfeita se a estatística de Durbin-Watson estiver próxima de 2 (e entre 1 e 3).

previsora e um para a constante). Portanto, o modelo 1 tem 198 graus de liberdade, ao passo que o modelo 2 tem 196. A média da soma dos quadrados (MS) é então calculada para cada termo pela divisão do SS pelo gl. A razão F é calculada dividindo-se a melhoria no aumento médio de previsão do modelo (MS_M) pela diferença média entre o modelo e os dados observados (MS_R). Se a melhoria devido ao ajuste do modelo de regressão é muito maior do que a variação no interior do modelo, então o valor de F será maior do que 1 e o SPSS calcula a probabilidade exata de obter o valor de F por acaso. Para o modelo inicial, a razão F é 99,587, o que é absolutamente improvável de ter acontecido por acaso ($p < 0,001$). Para o segundo modelo, o valor de F é ainda maior (129,498), o que é também muito significativo ($p < 0,001$). Esses resultados significam que o modelo inicial melhorou de forma significativa nossa habilidade de prever um valor da variável de saída, mas o novo modelo (com os previsores adicionais) foi ainda melhor (porque o valor de F é ainda mais significativo).

5.8.3 Parâmetros do modelo ②

Até agora verificamos várias estatísticas resumo que nos informam se o modelo melhorou ou não a nossa habilidade de prever valores da variável de saída. A próxima parte da saída está relacionada com os parâmetros do modelo. A saída 5.7 do SPSS mostra os parâmetros do modelo para os dois passos da hierarquia. O primeiro passo da hierarquia foi incluir a verba publicitária (como foi feito na regressão simples anteriormente neste capítulo) e, assim, os parâmetros do modelo são idênticos aos parâmetros obtidos na Saída 5.3 do SPSS. Portanto, estamos interessados somente nos parâmetros do modelo final (no qual todos os previsores foram incluídos). O formato das tabelas dos coeficientes dependerá das opções selecionadas. Os intervalos de confiança para os valores b , o diagnóstico de colinearidade e as correlações por parte e parciais serão apresentadas somente se forem selecionadas na caixa de diálogo conforme apresentado na Figura 5.3.

Lembre que de regressão múltipla o modelo toma a forma da equação (5.9) e nessa equação existem várias quantidades desconhecidas (os valores b). A primeira parte da tabela apresenta estimativas dos valores b e esses valores indicam a contribuição individual de cada previsor para o modelo. Se substituirmos os valores b na equação (5.9) veremos que é possível definir o modelo como na equação (5.12):

$$\begin{aligned} \text{Vendas}_i &= b_0 + b_1\text{publicidade}_i + b_2\text{execuções}_i \\ &\quad + b_3\text{atratividade}_i \\ &= -26,61 + (0,08\text{publicidade}_i) \\ &\quad + (3,7\text{execuções}_i) \\ &\quad + (11,09\text{atratividade}_i) \end{aligned} \quad (5.12)$$

Os valores b nos informam sobre o relacionamento entre a venda de discos e cada previsor. Se o valor é positivo, podemos dizer que existe um relacionamento positivo entre o previsor e a saída enquanto que um coeficiente negativo representa um relacionamento negativo. Para esses dados, todos os três previsores apresentam valores b positivos, indicando relacionamentos positivos. Assim, à medida que a verba publicitária aumenta, a venda de discos aumenta; se o número de execuções no rádio aumenta, a venda aumenta, e finalmente se a atratividade da banda aumenta, mais discos serão vendidos. Os valores b nos informam mais do que isso; eles nos informam em que grau cada previsor afeta a saída *se todos os demais previsores forem mantidos constantes*.

- **Verba publicitária (Advertising budget)** ($b = 0,085$): Esse valor indica que se a verba publicitária aumentar em uma unidade, a venda de discos irá aumentar 0,085 unidades. Ambas as variáveis foram mensuradas em milhares, portanto, para cada £1000,00 a mais gastas em publicidade, 0,085 milhares de discos (85 discos) serão vendidos. Essa interpretação é verdadeira somente se os efeitos da atratividade da banda e o número de execuções forem mantidos constantes.
- **Número de execuções (airplay)** ($b = 3,367$): Esse valor indica que se o número de vezes que o disco toca na rádio na

Saída 5.7 do SPSS Coeficientes do modelo de regressão¹¹

Coeficients^a (Coeficientes)

<i>Model (Modelo)</i>	<i>Unstandardized Coefficients</i> (Coeficientes não-padronizados)		<i>Standardized Coefficients</i> (Coeficientes padronizados)	<i>t</i>	<i>Sig.</i>	<i>95% Confidence Interval for B</i> (Intervalo de Confiança de 95% para B)	
	<i>B</i>	<i>Std. Error</i> (Erro Padrão)	<i>Beta</i>			<i>Lower Bond</i> (Limite Inferior)	<i>Upper bond</i> (Limite Superior)
1 <i>(Constant)</i> (Constante)	134.140	7.537		17.799	0.000	119.278	149.002
<i>Advertising Budget (thousands of pounds)</i> (Investimento em Publicidade) (em milhares de libras)	9.612E-02	0.10	0.578	9.979	0.000	0.077	0.115
2 <i>(Constant)</i> (Constante)	-26.613	17.350		-1.534	0.127	-60.830	7.604
<i>Advertising Budget (thousands of pounds)</i> (Investimento em Publicidade) (em milhares de libras)	8.488E-02	0.007	0.511	12.261	0.000	0.071	0.099
<i>No. Of plays on Radio 1 per week</i> (Número de execuções na Radio 1 por semana)	3.367	0.278	0.512	12.123	0.000	2.820	3.915
<i>Attractiveness of Band</i> (Atratividade da Banda)	11.086	2.438	0.192	4.548	0.000	6.279	15.894

a. Variável dependente: vendas de discos (em milhares).

Coeficients^a (Coeficientes)

<i>Model (Modelo)</i>	<i>Correlations</i> (Correlações)			<i>Collinearity Statistics</i> (Estatísticas de Colinearidade)	
	<i>Zero-order</i> (De ordem zero)	<i>Partial</i> (Parcial)	<i>Part</i> (Parte)	<i>Tolerance</i> (Tolerância)	<i>VIF</i>
1 <i>Advertising Budget (thousands of pounds)</i> (Investimento em Publicidade) (em milhares de libras)	0.578	0.578	0.578	1.000	1.000
2 <i>Advertising Budget (thousands of pounds)</i> (Investimento em Publicidade) (em milhares de libras)	0.578	0.659	0.507	0.986	1.015
<i>No. Of plays on Radio 1 per week</i> (Número de execuções na Radio 1 por semana)	0.599	0.655	0.501	0.959	1.043
<i>Attractiveness of Band</i> (Atratividade da Banda)	0.326	0.309	0.188	0.963	1.038

a. Variável dependente: vendas de discos (em milhares).

¹¹ Para poupar sua visão, dividi esta parte da saída em duas tabelas; contudo, ela deverá aparecer como uma única e longa tabela no visualizador de saídas do SPSS.

semana anterior ao lançamento aumenta em uma unidade, as vendas aumentarão 3,367 unidades. Dessa forma, cada execução adicional de uma música no rádio (na semana anterior ao lançamento) é associada com uma venda extra de 3,367 milhares de discos (3367 discos). Essa interpretação é verdadeira somente se a verba publicitária e a atratividade da banda permanecerem constantes.

- **Atratividade (Attractiveness)** ($b = 11,086$): Esse valor indica que se a atratividade da banda aumenta um grau na escala, então 11,086 unidades adicionais serão vendidas. Portanto, cada unidade de atratividade da banda está associada com uma venda extra de 11086 discos. Essa interpretação somente será verdadeira se a verba publicitária e o número de execuções permanecerem constantes.

A cada um desses valores β está associado um erro padrão indicando até que ponto esses valores podem variar entre amostras, e esses erros são utilizados para determinar se os valores b diferem significativamente de zero. Como visto na Seção 5.2.4, uma estatística t pode ser associada para testar se um valor b é significativamente diferente de zero. Na regressão simples, um valor significativo de t indica que a inclinação da linha de regressão é significativamente diferente de uma linha horizontal, mas na regressão múltipla, não é fácil vislumbrar o que o valor nos informa. É fácil conceituar o teste t como uma medida que informa se o preditor está contribuindo de forma significativa para o modelo. Portanto, se o teste t associado com um valor b é significativo (se o valor rotulado como *Sig.* é menor do que 0,05), então o preditor está fazendo uma contribuição significativa para o modelo. Quanto menor for o valor da *Sig.* (e maior o valor de t), maior será a probabilidade de essa contribuição ocorrer. Para esse modelo, a verba publicitária ($t(196) = 12,26, p < 0,001$), a quantidade de execuções no rádio antes do lançamento ($t(196) = 12,12, p < 0,001$) e a atratividade da banda ($t(196) = 4,55, p < 0,001$) são todos preditores significativos para

a venda de discos.¹² A partir da magnitude da estatística t podemos ver que a verba publicitária e o número de execuções têm impactos semelhantes, enquanto que a atratividade da banda tem um impacto menor.

Os valores b e suas significâncias são estatísticas importantes; contudo, as versões padronizadas dos valores b são mais fáceis de interpretar (porque elas não são dependentes das unidades de medida das variáveis). Os valores padronizados de beta são também fornecidos pelo SPSS (denominados Beta, β) e eles nos apresentam o número de desvios padrão que a saída irá mudar como resultado de uma alteração de um desvio padrão no respectivo preditor. Os valores β padronizados são todos mensurados em termos de unidades desvios padrão e são, dessa forma, comparáveis diretamente; portanto, eles fornecem uma ideia melhor da “importância” de um preditor para o modelo. Os valores β padronizados para as execuções e a verba publicitária são praticamente idênticos (0,512 e 0,511, respectivamente), indicando que as duas variáveis apresentam um grau de importância comparável para o modelo (isso concorda com a magnitude da estatística t)¹³. Para interpretar literalmente esses valores, precisamos saber os desvios padrão de todas as variáveis e esses valores podem ser encontrados na Saída 5.4 do SPSS.

- **Verba publicitária (Advertising budget)** (β padronizado = 0,511): Esse valor indica que se a verba publicitária aumentar em um desvio padrão (£485655), a

¹² Para todos os preditores escrevi $t(196)$. O número nos parênteses é o grau de liberdade. Vimos na Seção 5.2.4 que na regressão os graus de liberdade são $N - p - 1$, onde N é o tamanho da amostra (nesse caso, 200) e p é o número de preditores (nesse caso, três). Para esses dados, obtemos $200 - 3 - 1 = 196$.

¹³ A estatística t é comparável ao valor beta padronizado porque ela é obtida por meio da divisão pelo erro padrão e , dessa forma, representa a versão estudentizada dos valores beta. Os valores beta padronizados são calculados dividindo-se pelo desvio padrão. No Capítulo 1, vimos que o desvio padrão e o erro padrão estão intimamente relacionados: portanto, estatísticas padronizadas e estudentizadas são, de certa maneira, comparáveis.

venda de discos irá aumentar 0,511 desvios padrão. O desvio padrão da venda de discos é 80699 e, assim, teremos uma mudança de 41240 vendas (0,511.80699). Portanto, para cada £485655 a mais gasto em publicidade, uma venda extra de discos de 41240 unidades será obtida. Essa interpretação é verdadeira somente se os efeitos da atratividade da banda e o número de execuções no rádio permanecerem constantes.

- **Número de execuções (airplay)** (β padronizado = 0,512): Esse valor indica que se o número de vezes que o disco toca na rádio na semana anterior ao lançamento aumenta em um desvio padrão (12,27), as vendas aumentarão em 0,512 desvios padrão. O desvio padrão da venda de discos é 80699 e, desse modo, uma mudança de 41320 nas vendas poderá ocorrer (0,512.80699). Portanto, se a Rádio 1 toca a música um número extra de 12,27 vezes na semana anterior ao lançamento, uma venda extra de 41320 discos é esperada. Essa interpretação é verdadeira somente se a verba publicitária e a atratividade da banda permanecerem constantes.
- **Atratividade (Attractiveness)** (β padronizado = 0,192): Esse valor indica que se a atratividade da banda aumenta um desvio padrão (1,40 unidades), pode-se esperar 0,192 desvios padrão de unidades adicionais vendidas. Isso representa uma mudança de 15490 discos (0,192.80699). Portanto, uma banda com uma atratividade de 1,40 superior a outra banda pode esperar uma venda adicional de 15490 discos. Essa interpretação somente será verdadeira se a verba publicitária e o número de execuções permanecerem constantes.

Imagine que coletamos 100 amostras de dados e mensuramos as mesmas variáveis do nosso modelo atual. Para cada amostra, podemos criar um modelo de regressão para representar os dados. Se o modelo é fidedigno, esperamos encontrar parâmetros bastante semelhantes em ambos. Portanto, cada amostra deve produzir aproximadamente os mesmos

valores b . O intervalo de confiança para os valores β não-padronizados são limites construídos tais que em 95% dessas amostras esses limites irão conter os verdadeiros valores de b (veja a Seção 1.6.2). Portanto, se tivermos coletado 100 amostras e calculado os intervalos de confiança para b , diremos que temos uma confiança de 95% de que esses intervalos conterão os verdadeiros valores dos coeficientes b . Assim, podemos garantir que o intervalo de confiança que construímos para essas amostras conterá o verdadeiro valor de b populacional. Isso acontecendo, um bom modelo apresentará intervalos de confiança pequenos, indicando que os valores de b nessa amostra estão próximos do verdadeiro valor de b na população. O sinal (positivo ou negativo) dos valores b nos informa sobre a direção do relacionamento entre o previsor e o resultado. Desse modo, podemos esperar que um modelo muito ruim apresente intervalos contendo o zero, indicando que em algumas amostras o previsor tem um relacionamento positivo com o resultado, ao passo que em outras ele tem um relacionamento positivo. Nesse modelo, os dois melhores previsores (verba publicitária e número de execuções) apresentam intervalos bastante estreitos, indicando que as estimativas para o modelo atual são provavelmente representativas dos valores populacionais. O intervalo para a atratividade é mais largo (mas ainda não contém o zero), indicando que o parâmetro para essa variável é menos representativo, mas ainda assim significativo.

Se você solicitou as correlações por parte (ou parciais), elas irão aparecer na saída em colunas separadas da tabela. As correlações de ordem zero são os coeficientes de correlação simples de Pearson (e correspondem aos valores na Saída 5.4 do SPSS). As correlações parciais representam o relacionamento entre cada previsor e a variável de saída, controlados os efeitos dos demais previsores. A correlação por partes representa o relacionamento entre cada previsor e a saída, controlado o efeito que as outras duas variáveis apresentam na variável de saída. De fato, essas correlações por partes representam o relacionamento

Dica da Samanta Ferrinho



A contribuição individual de cada variável para o modelo de regressão pode ser encontrada na tabela dos **coeficientes** do SPSS. Se você fez uma regressão hierárquica, olhe para os valores do modelo final. Você pode ver se cada variável previsor fez uma contribuição significativa para prever o resultado olhando para a coluna rotulada como *Sig.* (valores menores do que 0,05 são significativos). Você deve também observar os valores β padronizados porque eles informam sobre a importância de cada previsor (quanto maior o valor absoluto, melhor). Os valores da tolerância e do FIV também serão úteis mais tarde, portanto, tome nota deles!

único que cada previsor tem com a variável de saída. Se você optar por executar uma regressão passo a passo (*stepwise*), verá que a variável da entrada é baseada inicialmente na variável com a maior correlação de ordem zero e depois com a correlação por parte com as variáveis restantes. Portanto, o número de execuções será adicionado primeiro (em virtude da maior correlação de ordem zero), depois a verba publicitária (porque sua correlação por partes é maior do que a atratividade) e finalmente a atratividade. Tente executar uma regressão passo a passo para frente (*forward stepwise*) nesses dados para ver se eu estou certo! Finalmente, iremos fornecer detalhes sobre as estatísticas de colinearidade, mas elas serão discutidas na Seção 5.8.5.

5.8.4 Variáveis excluídas ②

A cada estágio da análise de regressão, o SPSS fornece um resumo de qualquer variável que ainda não foi adicionada ao modelo. No modelo hierárquico, esse resumo apresenta detalhes das variáveis que foram especificadas para entrar no modelo em passos subsequentes, e na regressão passo a passo essa tabela (Saída 5.8 do SPSS) contém resumos das variáveis que o SPSS determinou que devem entrar no modelo. Para esse exemplo, existe um resumo das variáveis excluídas no primeiro estágio da hierarquia (não existe resumo para o segundo estágio porque todos os previsores estão no modelo). O resumo fornece uma estimativa de cada valor beta do previsor se ele entrar na equação nesse ponto e fornece um

Saída 5.8 do SPSS

Excluded Variables^b (Variáveis Excluídas)

Model (Modelo)	Beta In	t	Sig.	Partial Correlations (Correlações)	Collinearity Statistics (Estatísticas de Colinearidade)		
					Tolerance (Tolerância)	FIV	Minimum Tolerance (Tolerância Mínima)
1 No. Of plays on Radio 1 per week (Número de execuções na Radio 1 por semana)	0.546 ^a	12.513	0.000	0.665	0.990	1.010	0.990
Attractiveness of Band (Atratividade da Banda)	0.281 ^a	5.136	0.000	0.344	0.993	1.007	0.993

a. Previsores no modelo: (constante), Verba publicitária (em milhares de libras).
b. Variável dependente: Venda de discos (em milhares).

teste *t* para esse valor. Na regressão passo a passo (*stepwise*), o SPSS entra o previsor com a estatística *t* mais alta e continua entrando previsores até que não reste nenhum com estatística *t* que tenha valor significativo inferior a 0,05. A correlação parcial também fornece algumas indicações sobre qual contribuição (se existir alguma) um previsor excluído faria se ele tivesse entrado no modelo.

5.8.5 Avaliando as hipóteses de não-multicolinearidade ②

A Saída 5.7 do SPSS fornece algumas medidas para verificar a colinearidade nos dados. Especificamente, ela fornece as estatísticas FIV e tolerância (com a tolerância sendo 1 dividido pela FIV). Existem algumas observações da Seção 5.6.2.3 que podem ser aplicadas aqui:

- Se a maior FIV for maior do que 10, então existe motivo para preocupação (Myers, 1990; Bowerman e O’Connell, 1990).
- Se a FIV média é substancialmente maior do que 1, então a regressão pode ser tendenciosa (Bowerman e O’Connell, 1990).
- Tolerância abaixo de 0,10 indica problemas sérios.
- Tolerância acima de 0,20 indica um problema em potencial (Menard, 1995).

Para o modelo atual, os valores FIV estão todos abaixo de 10 e as tolerâncias todas acima de 0,20; portanto, podemos concluir com segurança que não existe colinearidade dentro dos nossos dados. Para calcular a FIV média, simplesmente adicionamos os valores para cada previsor e dividimos pelo número de previsores (*k*):

$$\overline{FIV} = \frac{\sum_{i=1}^k FIV_i}{k} = \frac{1,015 + 1,043 + 1,038}{3} = 1,032$$

A FIV média está muito próxima de 1 e confirma que a colinearidade não é um problema para esse modelo. O SPSS também produz uma tabela de autovalores da matriz dos produtos cruzados, índices de condição e proporção de variâncias. Existe uma longa discussão, e um exemplo, sobre colinearidade na Seção 6.8 e é apresentada a forma como detectá-la utilizando proporções de variâncias. Portanto, vou me limitar agora em dizer que procuraremos por grandes proporções de variâncias em um mesmo *pequeno* autovalor. Assim, na saída 5.9 do SPSS olhamos para as últimas linhas inferiores da tabela (esses são os pequenos autovalores) e procuramos por variáveis que tenham proporções de alta variância para esse autovalor. As

Saída 5.9 do SPSS

Collinearity Diagnostics^a (Diagnósticos de Colinearidade)

Model (Modelo)	Dimension (Dimensão)	Eigenvalue (Autovalor)	Condition Index (Índice de Condição)	Variance Proportions (Proporções da Variância)			
				(Constant) (Constante)	Advertising Budget (thousands of pounds) (Verba publicitária (em milhares de libras))	No. Of plays on Radio 1 per week (Execuções na Radio 1 por semana)	Attractiveness of Band (Atratividade da Banda)
1	1	1.785	1.000	0.11	0.11		
	2	0.215	2.883	0.89	0.89		
2	1	3.562	1.000	0.00	0.02	0.01	0.00
	2	0.308	3.401	0.01	0.96	0.05	0.01
	3	0.109	5.704	0.05	0.02	0.93	0.07
	4	2.039E-02	13.219	0.94	0.00	0.00	0.92

a. Dependent Variable: Record Sales (thousands) (Variável dependente: Venda de discos (em milhares))

proporções de variâncias variam entre 0 e 1 e para cada previsor deve estar distribuída entre as diferentes dimensões (ou autovalores). Para esse modelo, você poderá ver que cada previsor tem a maioria de suas variâncias em dimensões diferentes (a verba publicitária tem 96% da variância na dimensão 2, o número de execuções

tem 93% da variância na dimensão 3 e a atratividade da banda tem 92% da variância na dimensão 4). Esses dados representam o exemplo clássico de multicolinearidade. Para um exemplo de quando a colinearidade existe nos dados e algumas sugestões sobre o que pode ser feito, veja os Capítulos 5 e 11 (Seção 6.8).

Dica da Samanta Ferrinho



Para checar as hipóteses de inexistência de multicolinearidade, use os valores FIV da tabela denominada **Coefficientes** (Coefficients) na saída do SPSS. Se esses valores forem menores do que 10, provavelmente não há motivo de preocupação. Se o valor médio da VIF e não for substancialmente maior do que 1, isso também indica que não há motivos para alarme.

Quadro 5.2

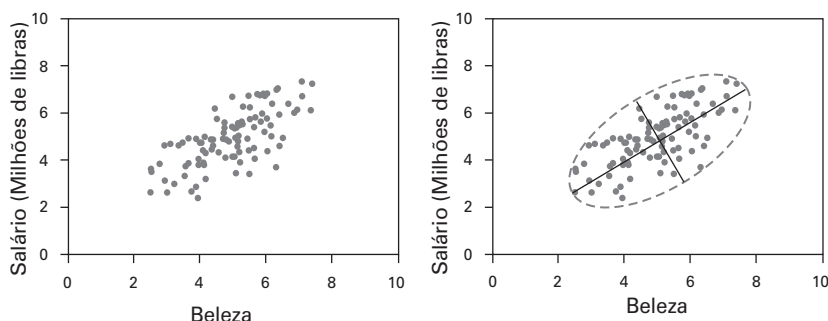
O que são autovalores e autovetores ④



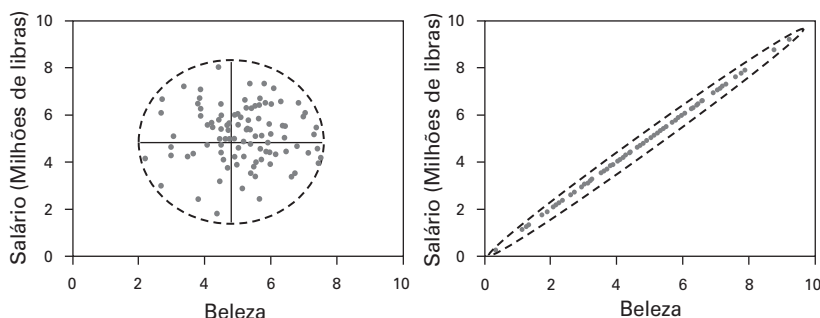
As definições e a matemática dos autovalores e autovetores são bastante complicadas e geralmente não precisamos nos preocupar com elas (embora eles surjam novamente nos Capítulos 14 e 15). Contudo, embora a matemática envolvida seja difícil, ela é fácil de visualizar! Imagine que temos duas variáveis: o salário anual de uma supermodelo e quão bela ela é. Imagine, ainda, que essas duas variáveis são normalmente distribuídas e, portanto, podem ser considerados em conjunto como uma distribuição normal bivariada. Se essas variáveis estiverem correlacionadas, o diagrama de dispersão forma uma elipse. Isso é mostrado nos diagramas de dispersão abaixo: se traçarmos uma linha pontilhada ao redor dos valores superiores do diagrama, obteremos algo de forma oval. Agora, podemos traçar duas linhas para medir o comprimento e a largura da elipse. Essas linhas são os autovetores da matriz de correlação original para essas variáveis (um vetor é apenas um conjunto de números que nos informa a localização de uma linha no espaço geométrico). Observe que as duas linhas traçadas (uma para a largura e outra para o comprimento da oval) são perpendiculares, isto é, elas formam um ângulo de 90 graus, o que significa que elas são independentes uma da outra. Assim, com duas variáveis, autovetores são apenas linhas medindo o comprimento e a largura da elipse ao redor do diagrama de dispersão dos dados para essas variáveis. Se adicionarmos uma terceira variável (por exemplo, a experiência da supermodelo), o diagrama de dispersão obtém uma terceira dimensão, a elipse torna-se algo com a forma de uma bola de rúgbi (ou de futebol americano), e porque adicionamos a terceira dimensão (comprimento, largura e altura), ganhamos um terceiro autovetor que mede a dimensão extra.

Se adicionarmos uma quarta variável, uma lógica semelhante é aplicada (embora seja mais difícil visualizá-la): obtemos uma dimensão extra e um autovetor para medir essa dimensão. Agora, cada autovetor tem um autovalor que nos informa seu comprimento (isto é, a distância de um ponto (início) a outro (final) do autovetor). Assim, examinando todos os autovetores de um conjunto de dados, sabemos as dimensões da elipse ou bola de rúgbi; ou seja, nós sabemos as dimensões dos dados. Portanto, os autovetores mostram quão equitativamente as variâncias da matriz estão distribuídas.

(Continua)

Quadro 5.2 (Continuação)

No caso de duas variáveis, a *condição* dos dados é relacionada à razão entre o tamanho do maior autovalor pelo do menor. Vamos olhar para os dois extremos: quando não existe relacionamento entre as variáveis e quando existe um relacionamento perfeito. Quando não existe relacionamento, o diagrama de dispersão irá estar mais ou menos contido em uma circunferência (ou em uma esfera, se tivermos três variáveis). Se novamente traçarmos linhas que medem o comprimento e a largura do círculo, veremos que elas apresentam o mesmo valor. Os autovalores medem o comprimento; portanto, os autovalores serão os mesmos. Assim, quando dividirmos o comprimento do maior autovalor pelo menor, obteremos 1 (porque os autovalores são os mesmos). Quando as variáveis são perfeitamente correlacionadas (isto é, existe uma colinearidade perfeita), o diagrama de dispersão forma uma linha reta e a elipse ao redor será muito estreita (com largura próxima de zero). Consequentemente, quando dividirmos o maior autovalor pelo menor, iremos obter um valor que tende para o infinito (porque o menor autovalor está próximo de zero.) Dessa forma, um índice de condição infinito é sinal de um grande problema.



5.8.6 Diagnósticos por casos (Casewise) ②

O SPSS produz uma tabela resumo de estatísticas residuais e elas devem ser examinadas para casos extremos. A Saída 5.10 do SPSS mostra casos com um resíduo padronizado menor que -2 ou maior do que 2 (lembre que mu-

damos o critério padrão de 3 para 2 na Figura 5.13). Mencionei na Seção 5.6.1.1 que em uma amostra comum esperamos que 95% dos casos tenham resíduos padronizados dentro de ± 2 . Nós temos uma amostra de 200, assim, é razoável esperar que aproximadamente 10 casos (5%) tenham resíduos padronizados fora desses limi-

Saída 5.10 do SPSS


Casewise Diagnostics^a (Diagnósticos caso a caso)

<i>Case number</i> (Caso número)	<i>Std. Residual</i> (Resíduo Padronizado)	<i>Record Sales (thousands)</i> (Vendas de discos (milhares))	<i>Predicted value</i> (Valor previsto)	<i>Residual</i> (Resíduo)
1	2.125	330.00	229.9203	100.0797
2	-2.314	120.00	228.9490	-108.9490
10	2.114	300.00	200.4662	99.5338
47	-2.442	40.00	154.9698	-114.9698
52	2.069	190.00	92.5973	97.4027
55	-2.424	190.00	304.1231	-114.1231
61	2.098	300.00	201.1897	98.8103
68	-2.345	70.00	180.4156	-110.4156
100	2.066	250.00	152.7133	97.2867
164	-2.577	120.00	241.3240	-121.3240
169	3.061	360.00	215.8675	144.1325
200	-2.064	110.00	207.2061	-97.2061

a. Dependent variable: Record Sales (thousands) (Variável dependente: vendas de discos (em milhares))

tes. Da Saída 5.10 do SPSS podemos ver que temos 12 casos (6%) que estão fora dos limites: portanto, nossa amostra está 1% além do que esperávamos. Além disso, 99% dos casos devem estar entre $\pm 2,5$ e, desse modo, esperamos apenas 1% dos casos fora desses limites. A partir dos casos listados aqui ficou claro que dois casos (1%) estão fora dos limites (casos 164 e 169). Portanto, nossa amostra parece se ajustar ao que esperaríamos de um modelo bastante preciso. Esses diagnósticos nos informam que não existe um motivo real de preocupação exceto pelo fato do caso 169 ter resíduo padronizado acima de 3, provavelmente grande o suficiente para merecer uma investigação adicional.

Lembre que na Seção 5.7.4 solicitamos que o SPSS salvasse várias estatísticas de diagnóstico. Você deve verificar que o editor de dados contém agora colunas para essas variáveis. É perfeitamente aceitável checar esses valores no editor de dados, mas você também pode fazer com que o SPSS liste os valores na janela de visualização. Para listar essas variáveis, você precisa utilizar o comando *Case Summaries* (Resumo de casos), que pode ser encontrado pela utilização do caminho **Analyze**⇒**Reports**⇒**Case Summaries...** (Analisar⇒Relatório⇒Casos Resumos...).¹⁴ A

Figura 5.17 mostra a caixa de diálogo para essa função. Selecione as variáveis que você quer listar e transfira-as para o painel denominado *Variables* (Variáveis) clicando em . Por omissão, o SPSS irá limitar a saída aos primeiros 100 casos, mas se você quiser listar todos os seus casos, basta retirar a seleção dessa opção. É também muito importante selecionar a opção *Show case numbers* (Mostrar número dos casos) porque de outro modo talvez você não seja capaz de identificar um caso problemático.

Uma estratégia útil é utilizar o diagnóstico caso a caso para identificar situações que merecem uma análise mais detalhada. Assim, para economizar espaço, criei uma variável código (1 = inclui, 0 = exclui), de modo que pude identificar os 12 casos listados na Saída 5.11 do SPSS em um grupo e todos os demais em outro. Utilizando essa variável código e especificando-a como uma variável de agrupamento na caixa de diálogo *summarize cases* (resumo de casos), pude olhar os 12 casos em conjunto e descartar todos os outros.

A Saída 5.11 do SPSS mostra as estatísticas de influência para os 12 casos que selecionei. Nenhuma delas tem a distância de Cook maior do que 1 (mesmo o caso 169 está bem abaixo desse critério) e, assim, nenhum desses casos irá influenciar indevidamente o modelo. A influência média pode ser calculada como $0,02 = (k + 1/n) = 4/200$ e, desse modo, iremos procurar por casos com valo-

¹⁴ **Statistics**⇒**Summarize**⇒**Case Summaries...** (Estatística⇒Resumir⇒Resumo de Casos) na versão 8.0 e anteriores.

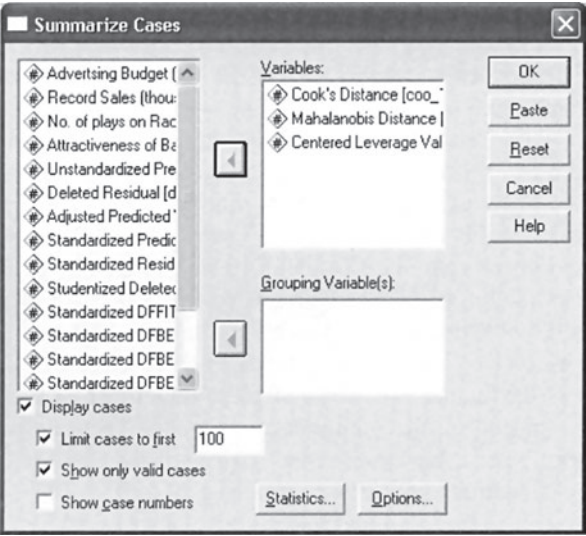


Figura 5.17

res duas (0,04) ou três vezes (0,06) maiores dependendo em qual estatístico você acredita mais (veja a Seção 5.6.1.2)! Todos os casos estão dentro dos limites de três vezes a média e somente o caso 1 está próximo de duas vezes a média. Finalmente, das recomendações para a distância de Mahalanobis, vimos que

para uma amostra de 100 com três previso- res, valores maiores do que 15 seriam proble- máticos. Temos três previsores e uma grande amostra, assim, esse valor será um ponto de corte bem conservador, já que nenhum dos nossos casos está próximo de exceder esse critério. A evidência sugere que não existem

Saída 5.11 do SPSS

Case Summaries (Resumo de Casos)

	Case number (Caso número)	Standardized DFBETA Intercept (Intercepto do DFBETA padronizado)	Standardized DFBETA ADVERTS (PUBLICIDADE DFBETA padronizada)	Standardized DFBETA AIRPLAY (EXECUÇÕES DFBETA padronizadas)	Standardized DFBETA ATTRACT (ATRATIVIDADE DFBETA padronizada)	Standardized DFFIT (DFFIT padronizado)	COVRATIO
1	1	− 0.31554	− 0.24235	0.15774	0.35329	0.48929	0.97127
2	2	0.01259	− 0.12637	0.00942	− 0.01868	− 0.21110	− 0.92018
3	10	− 0.01256	− 0.15612	0.16772	0.00672	0.26896	0.94392
4	47	0.06645	0.19602	0.04829	− 0.17857	− 0.31469	0.91458
5	52	0.35291	0.02881	− 0.13667	− 0.26965	0.36742	0.95995
6	55	0.17427	0.32649	− 0.02307	− 0.12435	− 0.40736	0.92486
7	61	0.00082	− 0.01539	0.02793	0.02054	0.15562	0.93654
8	68	− 0.00281	0.21146	− 0.14766	− 0.01760	− 0.30216	0.92370
9	100	0.06113	0.14523	− 0.29984	0.06766	0.35732	0.95888
10	164	0.17983	0.28988	− 0.40088	− 0.11706	− 0.54029	0.92037
11	169	− 0.16819	− 0.25765	0.25739	0.16968	0.46132	0.85325
12	200	0.16633	− 0.04639	0.14213	− 0.25907	− 0.31985	0.95435
Total	N	12	12	12	12	12	12

(Continua)

Case Summaries (Resumo de Casos)

	<i>Case number</i> (Caso número)	<i>Cook's Distance</i> (Distância de Cook)	<i>Mahalanobis Distance</i> (Distância de Mahalanobis)	<i>Centered Leverage Value</i> (Valor da Influência Centrado)
1	1	0.05870	8.39591	0.04219
2	2	0.01089	0.59830	0.00301
3	10	0.01776	2.07154	0.01041
4	47	0.02412	2.12475	0.01068
5	52	0.03316	4.81841	0.02421
6	55	0.04042	4.19960	0.02110
7	61	0.00595	0.06880	0.00035
8	68	0.02229	2.13106	0.01071
9	100	0.03136	4.53310	0.02278
10	164	0.07077	6.83538	0.03435
11	169	0.05087	3.14841	0.01582
12	200	0.02513	3.49043	0.01754
Total	N	12	12	12

casos que podem influenciar nos nossos dados (embora todos os casos tenham que ser examinados para confirmar esse fato).

Podemos também examinar a estatística DFBeta para ver se algum caso tem uma grande influência nos parâmetros de regressão. Um valor absoluto maior do que 1 será um problema e em todos os casos os valores estão entre ± 1 , o que mostra que esses casos não tem

influência indevida sobre os parâmetros de regressão. Existe também uma coluna para a razão de covariâncias. Vimos na Seção 5.6.1.2 que precisamos utilizar o seguinte critério:

- $CVR_i > 1 + [3(k + 1)/n] = 1 + [3(3 + 1)/200] = 1,06$.
- $CVR_i < 1 - [3(k + 1)/n] = 1 - [3(3 + 1)/200] = 0,94$.

Dica da Samanta Ferrinho

Você precisa procurar casos que possam influenciar o modelo de regressão.

- Verifique os resíduos padronizados e confirme que não mais do que 5% dos casos apresentam valores absolutos acima de 2 e não mais do que 1% apresentam valores absolutos acima de 2,5. Qualquer caso com valor acima de 3 será um valor atípico.
- Olhe no editor de dados as distâncias de Cook: qualquer valor acima de 1 indica um caso que poderá influenciar o modelo.
- Calcule a influência média (o número de previsores mais 1 dividido pelo tamanho da amostra) e então observe valores maiores do que duas ou três vezes a influência média.
- Para a distância de Mahalanobis, uma verificação rápida é procurar valores acima de 25 em amostras grandes (500) e valores acima de 15 em amostras pequenas (100). Contudo, Barnett e Lewis (1978) devem ser consultados para uma análise mais detalhada.
- Procure valores com DFBeta maiores do que 1.
- Calcule os limites superior e inferior dos valores aceitáveis para a razão de covariâncias, CVR. O limite superior é 1 mais três vezes a influência média, enquanto que o limite inferior é 1 menos três vezes a influência média. Casos que apresentam um CVR fora desses limites podem ser problemáticos.

Portanto, estamos procurando por casos que se desviem substancialmente desses limites. Muitos dos nossos 12 valores atípicos potenciais apresentam valores CVR dentro ou próximo desses limites. O único caso que pode preocupar é o 169 (novamente!), cujo CVR está de alguma forma abaixo do limite inferior. Contudo, verificando a distância de Cook para esse caso, provavelmente não há motivo para preocupação.

Se você tivesse solicitado outras estatísticas de diagnóstico e com o que você já sabe de discussões anteriores sobre elas, você saberia que deve analisá-las na situação de casos não usuais nos dados. Contudo, a partir desse conjunto mínimo de diagnóstico parece que temos um modelo razoavelmente confiável que não tem sido influenciado por qualquer subconjunto de casos.

5.8.7 Conferindo as hipóteses ②

Para finalizar a análise, você deve verificar as hipóteses do modelo. Já checamos a colinearidade dentro dos dados e utilizamos a estatística de Durbin-Watson para verificar se os resíduos no modelo são independentes. Na Seção 5.7.3, solicitamos um gráfico dos *ZRESID contra *ZPRED e um histograma e um diagrama das probabilidades normais dos resíduos. O gráfico do *ZRESID e *ZPRED deve ser semelhante a um conjunto aleatório de pontos dispersos em torno de zero. Se esse gráfico afunila, é provável que os dados sejam heterocedásticos. Se existir qualquer tipo de curva nesse gráfico, as chances são de que esses dados não satisfazem a hipótese de linearidade. A Figura 5.18 mostra vários exemplos de gráficos dos resíduos padronizados

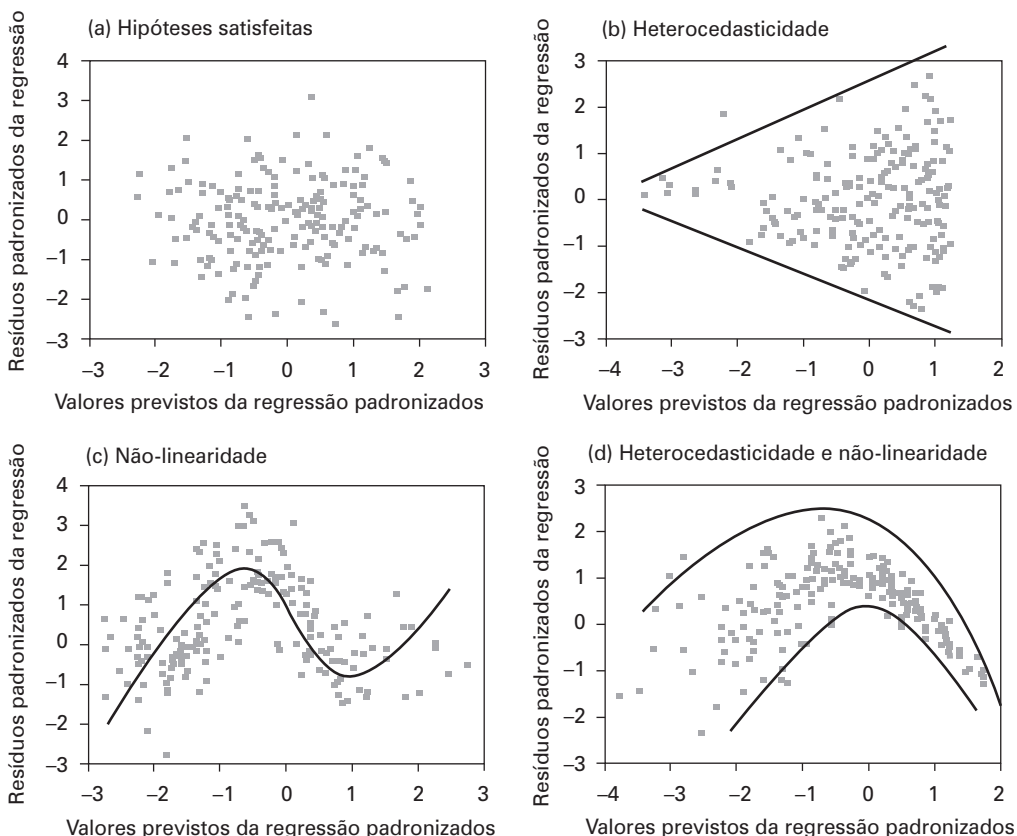


Figura 5.18 Diagramas dos *ZREDID contra *ZPRED.

dos contra os valores previstos padronizados. O painel (a) mostra o gráfico do nosso exemplo da venda de discos. Note como os pontos estão aleatoriamente dispersos por todo o quadrante. Esse padrão é um indicativo da situação em que as suposições de linearidade e homocedasticidade foram satisfeitas. O painel (b) mostra um diagrama semelhante para um conjunto de dados que viola a hipótese de homocedasticidade. Note que os pontos se espalham como uma forma de funil ficando cada vez mais distantes entre si à medida que os dados avançam para a direita. Essa é a forma típica da heterocedasticidade e indica um crescimento da variância nos resíduos. O painel (c) mostra um diagrama de dados que não tem um relacionamento linear entre os previsores e a variável de saída. Esse padrão é mostrado pelos resíduos. Uma linha

ilustrando a relação não-linear foi desenhada sobre o gráfico destacando a tendência dos dados. Finalmente, o painel (d) representa uma situação na qual os dados não apenas apresentam um relacionamento não-linear, mas também heterocedasticidade. Note inicialmente que num canto do diagrama os pontos estão bem próximos, mas na outro eles estão bem separados. Quando essas suposições não forem satisfeitas você não irá ver exatamente esses padrões, mas espero que esses diagramas o ajude a entender os tipos de anomalias que você deve procurar.

Para testar a normalidade dos resíduos, devemos olhar para os histogramas e o diagrama de probabilidades normais selecionados na Figura 5.14. A Figura 5.19 mostra o histograma e o diagrama das probabilidades normais dos dados para o exemplo atual (lado esquerdo). O

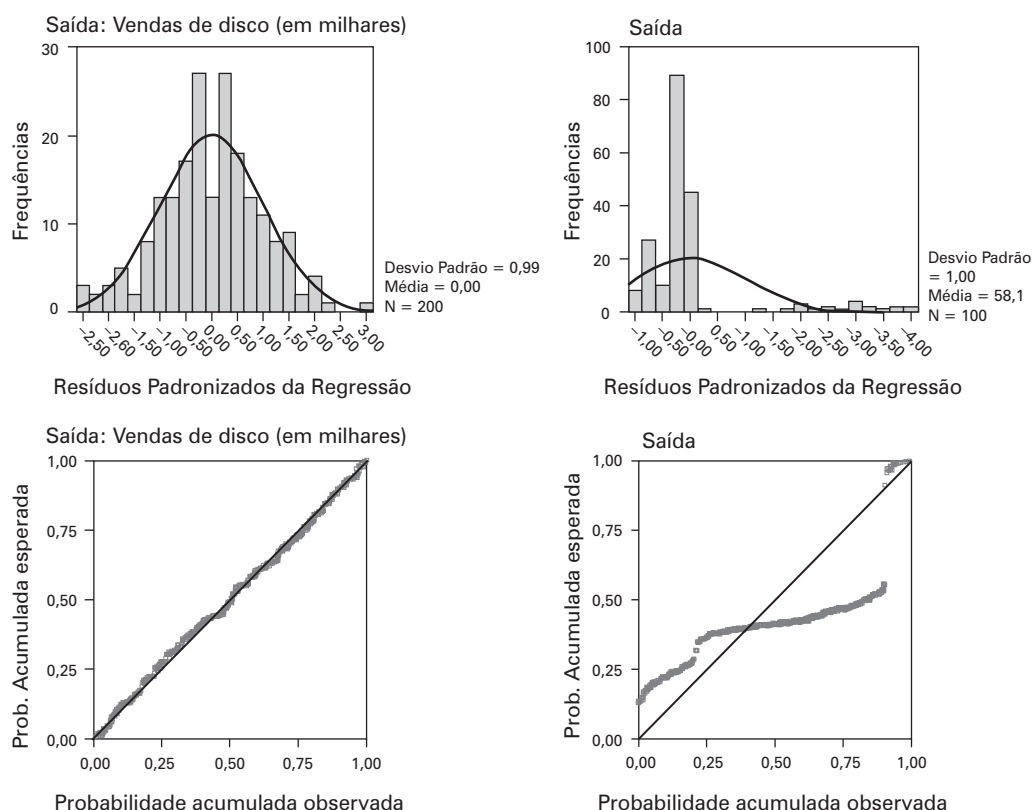
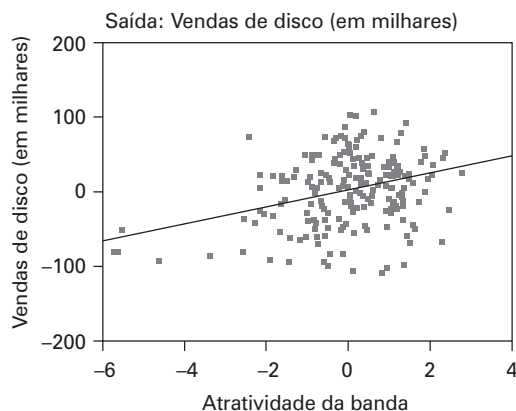
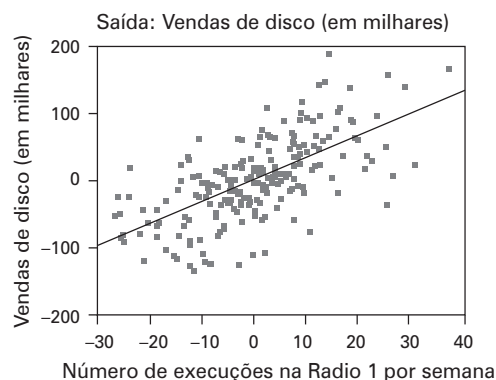
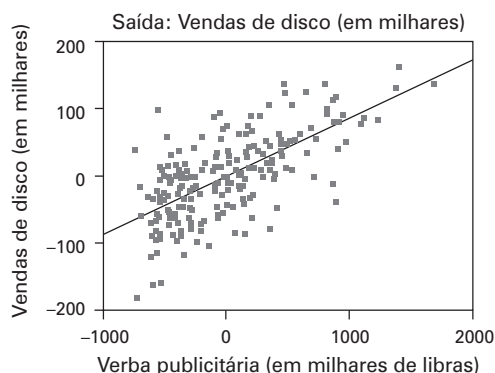


Figura 5.19 Histogramas e diagramas P-P dos resíduos normalmente distribuídos (lado esquerdo) e resíduos não normalmente distribuídos (lado direito).

histograma deve ser parecido ao de uma distribuição normal. O SPSS desenha a curva sobre o histograma para mostrar a forma da distribuição. Para os dados da gravadora, a distribuição é aproximadamente normal (embora exista uma leve deficiência dos resíduos exatamente em zero). Compare esse histograma com o que está ao lado dele, extremamente não-normal, e deve ficar claro que a distribuição não-normal apresenta uma assimetria acentuada. Assim, você deve procurar curvas que tenham a mesma forma que essa dos dados da venda de discos: qualquer desvio dessa curva é um sinal de não-normalidade e quanto maior a diferença, mais os resíduos estarão afastados da normalidade. O diagrama das probabilidades normais mostra também os desvios da normalidade (veja o Capítulo 2). A linha reta neste diagrama

representa a distribuição normal e os pontos, os resíduos observados. Portanto, em um conjunto de dados perfeitamente normal; todos os pontos estarão bem próximos da linha. Isso é o que enxergamos para os dados da venda de discos. Contudo, próximo ao diagrama dos dados normalmente distribuídos da venda de discos está um exemplo de um afastamento extremo da normalidade. Neste diagrama, os pontos estão bem distantes da linha, o que indica um grande afastamento da normalidade. Para os dois diagramas, os dados não-normais são casos extremos e você deve ficar atento para desvios da normalidade mais sutis. É claro que você pode utilizar o que aprendeu no Capítulo 2 para realizar um teste K-S nos resíduos padronizados a fim de ver se eles se desviam ou não de forma significativa da normalidade.



Um conjunto final de diagramas, os diagramas parciais, foi especificado na Figura 5.14. Esses diagramas são dispersões dos resíduos da variável de saída e de cada um dos previsores quando as duas variáveis são modeladas separadamente dos demais previsores. Mencionei antes que valores atípicos óbvios em um gráfico parcial representam casos que podem ter influência nos coeficientes dos previsores da regressão e que relacionamentos não-lineares e heterocedásticos podem ser detectados, da mesma forma, utilizando esses diagramas.

Para os dados da verba publicitária, os diagramas parciais mostram um relacionamento positivo forte com as vendas de disco. O gradiente da linha é b para a verba publicitária no modelo (essa linha não é mostrada por omissão). Não existem valores atípicos óbvios nesse diagrama e a nuvem de pontos é igualmente espaçada em torno da linha, indicando homocedasticidade.

Para o número de execuções (*airplay*), os diagramas parciais mostram um relacionamento positivo forte com a venda de discos. O gradiente e o padrão dos dados são semelhantes aos da verba publicitária (o que já era esperado devido à semelhança dos beta padroni-

zados desses previsores). Não existem valores atípicos óbvios nesse diagrama e a nuvem de pontos está igualmente espaçada em torno da linha indicando homocedasticidade.

Para a atratividade (*attractiveness*), os diagramas parciais mostram um relacionamento positivo forte com as vendas de discos. O relacionamento parece menos linear do que os demais previsores e os pontos parecem afunilados indicando um aumento da variância em níveis mais altos da atratividade. Não se percebem valores atípicos óbvios nesses dados, mas a nuvem de pontos em forma de funil pode indicar uma violação da hipótese de homocedasticidade. Seria adequado coletar alguns dados adicionais sobre bandas com baixa atratividade como forma de verificar o modelo atual.

Podemos resumir dizendo que o modelo parece, às vezes, em muitos sentidos, ser tanto preciso para a amostra e generalizável para a população. O único senão é se a atratividade da banda violou a hipótese de homocedasticidade. Portanto, podemos concluir que em nossa amostra a verba publicitária e o número de execuções são igualmente importantes em prever a venda de discos. A atratividade da banda é um preditor significativo da ven-

Dica da Samanta Ferrinho



Você precisa certificar-se de que algumas hipóteses da regressão foram satisfeitas a fim de ficar seguro de que ele pode ser generalizado para além da amostra.

- Verifique o gráfico do *ZRESID contra o *ZPRED. Se ele parecer com um conjunto aleatório de pontos, isso é bom. Se o gráfico parecer abrir como um funil de um lado ou outro, provavelmente teremos uma violação da hipótese de homogeneidade da variância. Se os pontos tiverem um padrão (forma curva), provavelmente teremos uma violação da linearidade. Se os dados parecem ter um padrão e estão mais espalhados em um ponto do que em outros, é provável que isso reflita uma violação tanto da homogeneidade da variância quanto da linearidade. Qualquer um desses cenários colocará a validade do nosso modelo em cheque. Repita a análise para os diagramas parciais também.
- Examine os histogramas e os diagramas P-P. Se os histogramas parecem com uma distribuição normal (e o diagrama P-P se assemelha a uma linha diagonal), tudo está bem. Se o histograma parece não-normal e o diagrama P-P se assemelha a uma cobra enrolada em torno da linha diagonal, as coisas não estão tão boas! Fique atento: distribuições normais podem parecer não-normais quando trabalhamos com pequenas amostras!

da de discos, mas é menos importante que os demais previsores (e provavelmente precisa de uma confirmação de uma possível heterocedasticidade). As hipóteses parecem ter sido satisfeitas e podemos provavelmente assumir que o modelo poderá ser generalizado para os novos discos sendo lançados.

5.9 COMO RELATAR A REGRESSÃO MÚLTIPLA ②

Se você seguir as recomendações da APA (American Psychological Association) para relatar a regressão múltipla, os resultados tabelados devem ser como os apresentados adiante. A APA também exige que sejam apresentados os betas padronizados, seus valores da significância e algumas estatísticas gerais sobre o modelo (tal como o R^2). Se você decidir fazer uma tabela, os valores beta e seus erros padrão também serão bastante úteis assim como a constante, para que os leitores do seu trabalho possam reproduzir o modelo completo se necessário. Ainda, se você executou uma regressão hierárquica, deve relatar esses valores em cada estágio da hierarquia. Desse modo, basicamente, você vai reproduzir a tabela denominada de coeficientes (*coefficients*) da saída do SPSS omitindo as informações não-essenciais. Para o exemplo deste capítulo podemos produzir uma tabela como a 5.2.

Assim, você poderá retomar a algumas das saídas do SPSS neste capítulo e descobrir de

onde esses valores vieram. As coisas a serem destacadas são: (1) arredondamos todos os valores para a segunda casa decimal; (2) para os betas padronizados não existe o zero antes da vírgula (porque esses valores não podem exceder 1),* mas para todos os demais valores menores do que um o zero está presente; (3) a significância de uma variável é representada por um asterisco com uma nota de rodapé a fim de indicar o nível de significância sendo utilizado (se existir mais do que um nível de significância, você pode representar isso com múltiplos asteriscos, tal como, $*p < 0,05$, $**p < 0,01$ e $***p < 0,001$) e (4) o R^2 para o modelo inicial e a mudança no R^2 (representada por ΔR^2) para cada passo subsequente do modelo são apresentados embaixo da tabela.

5.10 PREVISORES CATEGÓRICOS E A REGRESSÃO MÚLTIPLA ③

Muitas vezes na análise de regressão, coletamos dados sobre grupos de pessoas (alguns exemplos das ciências sociais podem ser o grupo étnico, o gênero, o status socioeconômico e a categoria de diagnóstico). Você pode querer esses grupos como previsores no modelo de regressão; contudo, vimos que entre as hipóteses do modelo está a necessidade de as variáveis serem contínuas ou pelo menos categóricas com apenas duas categorias. Vimos na Seção 4.5.6 que a correlação ponto-bisserial é o r de Pearson entre duas variáveis quando uma é contínua e a outra tem duas categorias codificadas como 0 e 1. Também aprendemos que a regressão simples é baseada no r de Pearson, assim, não será necessário muita imaginação para ver que, como a correlação ponto-bisserial, podemos construir um modelo de regressão com um predictor que tem duas categoriais (por exemplo, gênero). Da mesma forma, não deve ser tão inconcebível perceber que podemos estender esse modelo para incorporar vários previsores que tem duas

Tabela 5.2 Como relatar a regressão múltipla

	<i>B</i>	<i>SE B</i>	β
Passo 1			
Constante	134,14	7,54	
Verba publicitária	0,10	0,01	0,58*
Passo 2			
Constante	-26,61	17,35	
Verba publicitária	0,09	0,01	0,51*
Execuções na Rádio 1 da BBC	3,37	0,28	0,51*
Atratividade	11,09	2,44	0,19*

Nota: $R^2 = 0,34$ para o Passo 1; $\Delta R^2 = 0,33$ para o passo 2 ($ps < 0,001$). * $p < 0,001$.

* N. de T.: Essa observação não cabe aqui, pois os valores no Brasil sempre são escritos com o zero antes da vírgula, diferentemente do sistema anglo-saxão que costuma representar números entre -1 e 1 sem o zero antes do ponto decimal (vírgula para nós).

categorias. O que importa é que as categorias sejam codificadas com os valores 0 e 1. Por que é importante que só existam duas categorias e que elas sejam codificadas como 0 e 1? Não quero entrar nesse assunto porque este capítulo já está muito longo e os editores irão me xingar se ele aumentar mais, e eu irei abordá-lo, de alguma forma, mais tarde no livro (Seções 7.8 e 8.2.2). Assim, por enquanto, apenas acredite em mim!

5.10.1 Variáveis *dummy* ③

O problema óbvio com a utilização de variáveis categóricas como precursores é que muitas vezes temos mais do que duas categorias. Por exemplo, se a religiosidade for mensurada você poderá ter categorias como: muçulmanos, judeus, hindus, católicos, budistas, protestantes, Jedi (para os não-britânicos: foi feito um censo no país a alguns anos e uma quantidade significativa de pessoas declararam Jedi como sua religião). Claramente esses grupos não podem ser distinguidos utilizando uma única variável codificada como zero e um. Nesses casos, teremos que utilizar o que foi chamado de **variável *dummy*** (fictícia). A codificação *dummy* é uma forma de representar grupos de pessoas com somente zeros e uns. Para tanto, precisamos criar diversas variáveis; de fato, o número de variáveis necessárias é uma a menos do que o número de categorias que estamos codificando. Existem oito passos básicos:

1. Conte o número de grupos que você quer codificar e subtraia 1.
2. Crie tantas variáveis quanto o valor que foi determinado no passo 1. Essas são as variáveis *dummy*.
3. Escolha um dos seus grupos como base (isto é, o grupo contra o qual todos os demais serão comparados). Ele deve ser utilizado como grupo-controle, ou, se você não tem uma hipótese específica, ele deve ser o grupo que representa a maioria das pessoas (porque será importante comparar outros grupos contra a maioria).
4. Tendo escolhido o grupo-base, atribua a esse grupo o valor 0 para todas as variáveis *dummy*.

5. Para a primeira variável *dummy*, atribua o valor 1 ao primeiro grupo que você quer comparar contra o grupo-base. Atribua em todos os demais grupos o valor 0 para essa variável.
6. Para a segunda variável *dummy* atribua o valor 1 ao segundo grupo que você quer comparar ao grupo-base. Atribua a todos os demais grupos o valor 0 para essa variável.
7. Repita isso até que você não tenha mais variáveis *dummy*.
8. Coloque todas as variáveis *dummy* na análise de regressão!

Vamos testar isso utilizando um exemplo. No Capítulo 3, vimos um exemplo no qual uma bióloga estava preocupada sobre os efeitos potenciais dos festivais de música na saúde. Ela foi ao festival de música de Glastonbury e mensurou a higiene dos frequentadores do concerto durante os três dias de realização do mesmo, utilizando uma técnica que resulta em escores variando de 0 (a pessoa cheira como se tivesse tomado banho no esgoto) e 5 (a pessoa cheira como pão recém-saído do forno). No Capítulo 3, nos preocupamos apenas com a distribuição dos escores dos três dias do festival; mas imagine se a bióloga quisesse verificar o decréscimo da higiene ao longo do três dias do festival. O arquivo de dados **GlastonburyFestival-Regression.sav** pode ser encontrado no *site* www.artmed.com.br. Ele contém os escores da higiene para cada um dos três dias do festival e uma variável denominada **mudança** (*Change*), que é a mudança na higiene durante os três dias do festival (a alteração ocorrida do dia 1 ao dia 3).¹⁶ A bióloga categorizou as pessoas de acordo com a sua afiliação musical: as que preferiam música alternativa foram chamadas de “pagodeiros”, as que gostavam mais de *heavy metal* foram denominadas “metaleiros”, e as que gostavam de algum tipo de *hippie/folk* ambiente receberam o nome de *punks*. Qualquer uma que não se encaixasse nessas catego-

¹⁶ Lembre-se do Capítulo 3 que nem todo mundo pode ser mensurado no terceiro dia, assim, os escores de mudança são para um subconjunto da amostra original.

riais foi rotulada de “eccléticos”. No arquivo de dados, ela codificou esses grupos como 1, 2, 3 e 4, respectivamente.

A primeira coisa que voc  deve fazer   calcular o n mero de vari veis *dummy*. Se temos quatro grupos, dever o existir tr s vari veis *dummy* (uma a menos que o n mero de grupos ou categorias). Depois precisamos definir o grupo-base. Queremos comparar as pessoas que t m afilia  o musical com as que n o t m, assim, a nossa categoria base ser  “eccl ticos”. Daremos a esse grupo o c digo 0 para todas as nossas vari veis *dummy*. Para a primeira vari vel *dummy*, podemos olhar para grupo *punk* e para fazer isso daremos a todos os *punks* o c digo 1, e aos demais o c digo 0. Para a nossa segunda vari vel *dummy*, usaremos o grupo dos “metaleiros”, e daremos a qualquer pessoa que pertenc a a esse grupo o c digo 1, e aos outros o 0. Temos ainda uma vari vel *dummy* sobrando, destinada ao grupo “pagodeiros”: vamos dar a cada pagodeiros o c digo 1 e aos demais, o c digo 0. O esquema final   mostrado na Tabela 5.3. Voc  deve observar que em cada grupo o c digo 1 foi utilizado em apenas uma vari vel *dummy* (exceto na categoria base que   sempre codificada como 0).

Como j  disse, iremos prestar aten  o em como essa codifica  o funciona na Se  o 7.8 e 8.2.2; por enquanto, verifique como codificar nossa vari vel de agrupamento nessas vari veis *dummy* utilizando o SPSS. Para codificar vari veis, voc  precisa utilizar a fun  o *Recode* (recodificar), Use **Transform⇒Recode⇒Into Different Variables...** (Transformar⇒Recodificar⇒Em diferentes vari veis) para acessar a caixa de di logo vista na Figura 5.20. A primeira caixa de di logo lista

todas as vari veis no editor de dados, e voc  precisa selecionar as que quiser recodificar (nesse caso, **music** – m sica) e transferi-las para o painel denominado *Numeric Variable → Output Variable* (Vari vel num rica → Vari vel de Sa da) clicando em . Voc  precisa ent o dar um nome para a nova vari vel (a vari vel de sa da, de acordo com a denomina  o do SPSS). Para tanto, v  para a parte denominada *Output Variable* (Vari vel de Sa da) e no espa o onde diz *Name* (Nome), escreva o nome para a sua primeira vari vel *dummy* (voc  poder  cham -la de **music1**). Voc  deve dar a essa vari vel um nome mais descritivo digitando algo no espa o denominado *Label* (R tulo) (denominamos essa primeira vari vel *dummy* de eccl ticos × *punks*). Quando voc  tiver feito isso, clique em a fim de transferir essa nova vari vel para o painel denominado *Numeric Variable → Output Variable* (Vari vel num rica → Vari vel de Sa da); esse painel deve apresentar agora a seguinte legenda *music → music1*.

Uma vez definida a primeira vari vel *dummy*, precisamos informar ao SPSS como recodificar os valores da vari vel **music** nos novos valores que queremos para a nova vari vel **music1**. Para tanto, clique em a fim de abrir a segunda caixa de di logo como a da Figura 5.20. Essa caixa de di logo   utilizada a fim de mudar os valores da vari vel original para valores diferentes da nova vari vel. Para a primeira vari vel *dummy*, queremos que qualquer um que seja *punk* tenha um c digo de 1 e os demais, um c digo de 0. *Punks* foram codificados com o valor 3 na vari vel original, assim,   preciso digitar 3 na se  o denominada *Old Value* (Valor Antigo) no espa o chamado de *Value* (Valor). O novo valor que queremos   1, dessa forma, ser  necess rio digitar o valor 1 na se  o denominada *New Value* (Valor novo) no espa o chamado de *Value* (Valor). Depois, clique em para adicionar isso a lista de altera  es (a lista   mostrada no painel denominado *Old → New* (Velho → Novo), em que deve agora estar escrito 3 → 1, como no diagrama). O pr ximo passo   alterar os grupos remanescentes que apresen-

Tabela 5.3 Codifica  o *dummy* para os dados do festival de Glastonbury

	Vari�vel <i>dummy</i> 1	Vari�vel <i>dummy</i> 2	Vari�vel <i>dummy</i> 3
Eccl�ticos	0	0	0
Pagodeiros	0	0	1
Metaleiros	0	1	0
<i>Punks</i>	1	0	0

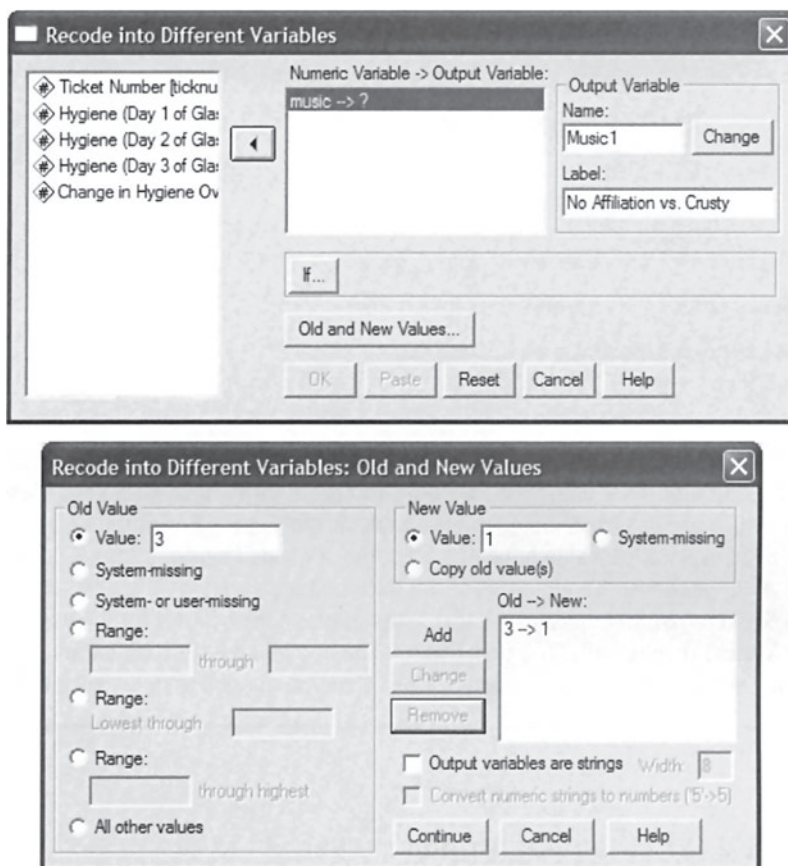


Figura 5.20 Caixa de diálogo para a função *Recode* (recodificar).

tam o valor 0 para a primeira variável *dummy*. Para tanto, selecione ☒ **All other values** e digite o valor 0 na seção denominada *New Value* (Valor Novo) no espaço *Value* (Valor). Quando tiver feito isso, clique em **Add** a fim de adicionar essa alteração à lista de mudanças (essa lista está mostrando *ELSE* → 0). Depois, clique em **Continue** para retornar à caixa de diálogo principal e clique em **OK** a fim de criar a primeira variável *dummy*. Essa variável irá aparecer agora em uma nova coluna no editor de dados e você irá notar que ela tem um valor de 1 para qualquer pessoa originalmente classificada como *punk* e o valor 0 para os demais. Agora, tente criar as próximas duas variáveis *dummy* (denomine-as de **Music2** e **Music3**) utilizando os mesmos princípios.

5.10.2 Saída do SPSS para variáveis auxiliares (*dummy*) ③

Vamos assumir que criamos as variáveis de código ou auxiliares (se você ainda não conseguiu, use o arquivo de dados denominado **GlastonburyDummy.sav**). Você precisa entrar com todas as variáveis auxiliares relacionadas no mesmo bloco (assim, utilize o método *Enter*). Neste caso, teremos que entrar com as variáveis *dummy* no mesmo bloco; contudo, se tivéssemos outra variável (por exemplo, status socioeconômico) que tivesse sido transformada em uma variável *dummy*, poderíamos inseri-la em um bloco diferente (assim, apenas as variáveis *dummy* recodificadas de uma mesma variável precisam ser

Quadro 5.3

Utilizando a sintaxe para recodificar ③

Se você está fazendo muitas recodificações, logo irá cansar de utilizar a caixa de diálogo todas as vezes. No *site* www.artmed.com.br, há um arquivo de sintaxe, **RecodeGlastonburyData.sps**, que eu escrevi para criar todas as variáveis *dummy* que foram discutidas. Carregue esse arquivo e rode a sintaxe, ou abra a janela da sintaxe (veja a Seção 2.6) e digite o seguinte:

```
RECODE music (3 = 1) (ELSE = 0) INTO Music1.
RECODE music (2 = 1) (ELSE = 0) INTO Music2.
RECODE music (1 = 1) (ELSE = 0) INTO Music3.
VARIABLE LABELS Music1 'No affiliation vs. Crusty'. Ecleticos vs. Punks
VARIABLE LABELS Music2 'No affiliation vs. Metaller'. Ecleticos vs. Metaleiros
VARIABLE LABELS Music3 'No affiliation vs. Indies Kid'. Ecleticos vs. Pagodeiros
EXECUTE.
```

Cada comando RECODE realiza o equivalente ao que eu fiz quando utilizei a caixa de diálogo apresentada na Figura 5.20. Assim, as primeiras três linhas apenas criam três novas variáveis (**Music1**, **Music2** e **Music3**), baseadas na variável original **Music**. Para essa primeira variável, se **music** é igual a 3, então ela torna-se 1, e todos os demais valores tornam-se 0. Para a segunda, se **music** é igual a 2, ela torna-se 1 e os demais valores ficam iguais a 0. A mesma lógica vale para a terceira variável *dummy*. As linhas iniciando com **VARIABLE LABELS** dizem ao SPSS para atribuir o texto entre apóstrofes como rótulos para as variáveis **Music1**, **Music2** e **Music3**, respectivamente. A linha final tem o comando **EXECUTE**, sem o qual nenhum dos comandos anteriores será executado. Observe ainda que cada linha termina com um ponto final.

colocadas em blocos). Use o que você aprendeu neste capítulo para rodar uma regressão múltipla utilizando os escores alterados como saída e as três variáveis auxiliares (colocadas em blocos) como previsores. Vamos dar uma olhada na saída.

A Saída 5.12 do SPSS mostra as estatísticas do modelo. Isso mostra que entrando com as três variáveis *dummy*, podemos explicar 7,6% da variância dos escores da higiene (o valor R^2). Em outras palavras, 7,6% da variância nos escores da higiene pode ser explicada pela afiliação musical da pessoa. A ANOVA (que mostra o mesmo que a estatística de mudança R^2 porque existe apenas um passo nessa regressão) nos informa que o modelo é significativamente melhor em prever a mudança nos escores da higiene do que não ter um modelo (ou seja, 7,6% da variância explicada é uma quantia significativa). Isso deve ficar

claro com base no que você leu neste capítulo; o que é mais interessante é como interpretar as variáveis *dummy* individualmente.

A Saída 5.13 do SPSS mostra a tabela básica dos coeficientes para as variáveis *dummy* (excluí os intervalos de confiança e o diagnóstico de colinearidade). A primeira coisa a observar é que cada variável *dummy* aparece na tabela com um rótulo útil (tal como *punks versus ecléticos*). Isso acontece porque quando codificamos as nossas variáveis, fornecemos um rótulo a cada uma; se não tivéssemos feito isso, a tabela conteria os nomes bem menos úteis das variáveis (*music1*, *music2* e *music3*). Os rótulos que eu sugeri que fossem fornecidos para as variáveis nos dão uma boa dica sobre o que cada variável representa. A primeira variável *dummy* (ecléticos *versus* punks) mostra a diferença entre a mudança nos escores da higiene para os grupos ecléticos e punks.

Saída 5.12 do SPSS

Model Summary^b (Resumo do Modelo)

Model (Modelo)	R	R Square (R quadrado)	Adjusted R Square (R Quadrado Ajustado)	Std. Error of the Estimate (Erro Padrão da Estimativa)	Change Statistics (Estatísticas de Mudança)					Durbin-Watson
					R Square Change (Mudança do R quadrado)	F Change (Mudança no F)	df1 (gl1)	df2 (gl2)	Sig. F Change (Sig. da Mudança no F)	
1	0.276 ^a	0.076	0.053	0.68818	0.076	3.270	3	119	0.024	1.893

a. Previsores (Constante) Ecléticos vs. Pagodeiros, Ecléticos vs. Punks, Ecléticos vs. Metaleiros.
b. Variável Dependente: Mudança na higiene ao longo do festival.

ANOVA^b (Análise de Variância)

Model (Modelo)		Sum of Squares (Soma dos quadrados)	df (gl)	Mean Square (Média dos Quadrados)	F	Sig.
1	Regression (Regressão)	4.646	3	1.549	3.270	0.024 ^a
	Residual (Resíduos)	56.358	119	0.474		
	Total	61.004	122			

a. Previsores (Constante) Sem Afiliação vs. pagodeiros, Sem afiliação vs. punks, Sem Afiliação vs. Metaleiros.
b. Variável Dependente: Mudança na higiene ao longo do festival.

Lembre-se de que os valores beta nos informam a mudança na saída devido a uma unidade no predictor. Nesse caso, a mudança de uma unidade no predictor é uma alteração de 0 para 1. Ela mostra a mudança nos escores da higiene como resultado mudança da variável *dummy* de 0 para 1 (*punk*). Pela inclusão das três variáveis ao mesmo tempo, nossa categoria base é sempre 0; assim, isso representa a mudança nos escores da higiene se uma pessoa eclética é comparada com um *punk*. Essa mudança é a diferença entre as médias dos dois grupos. Para ilustrar isso, produzi uma tabela com as médias de cada um dos quatro grupos. Essas médias representam a mudança média nos escores da higiene para os três grupos (isto é, os homens de cada grupo na nossa variável de saída). Se calcularmos as diferenças entre essas médias para o grupo dos ecléticos e o dos *punks*, obteremos: *Punks* – Ecléticos = $(-0,9658) - (-0,5543) = -0,4120$. Ou seja, a mudança nos escores da higiene é maior para o grupo dos *punks* do que para os ecléticos (a higiene dos *punks* decresce mais ao longo dos três dias do festival do que a dos

ecléticos). Esse valor é o mesmo que o beta na Saída 5.13 do SPSS! Assim, os valores beta nos informam a diferença relativa entre cada grupo e o grupo que foi escolhido como categoria base. Esse valor beta é convertido em uma estatística *t* e a significância dessa estatística é calculada. Essa estatística está testando, como já foi visto antes, se o valor é 0, e quando nós temos duas categorias codificadas com 0 e 1, isso significa que o que está sendo testado é se a diferença entre grupos é zero. Se ela é significativa, isso quer dizer que o

Report (Relatório)

Change in Hygiene Over The Festival
(Mudança da Higiene ao Longo do Festival)

Musical Affiliation (Afiliação Musical)	Mean (Médias)	N	Std Deviation (Desvios Padrão)
Indie Kid (Pagodeiros)	- 0.9643	14	0.67020
Metaller (Metaleiros)	- 0.5259	27	0.57583
Crusty (Punks)	- 0.9658	24	0.76038
No Musical affiliation (Ecléticos)	- 0.5543	58	0.70834
Total	- 0.6750	123	0.70713

Saída 5.13 do SPSS

Coefficients ^a (Coeficientes)					
Model (Modelo)	Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
	B	Std. Error (Erro Padrão)	Beta		
1 (Constant) (Constante)	– 0.554	0.090		– 6.134	0.000
No affiliation vs. Crusty (Ecléticos vs. Punks)	– 0.412	0.167	– 0.232	– 2.464	0.015
No affiliation vs. Metaller (Ecléticos vs. Metaleiros)	0.028	0.160	0.017	0.177	0.860
No affiliation vs. Indie Kid (Ecléticos vs. Pagodeiros)	– 0.410	0.205	– 0.185	– 2.001	0.048

a. Variável dependente: mudança na higiene ao longo do festival.

grupo codificado como 1 é significativamente diferente da categoria base – assim, ela está testando a diferença entre duas médias, que é o contexto onde você estará mais familiarizado com a estatística *t* (veja o Capítulo 7). Para a nossa primeira variável *dummy*, o teste *t* é significativo, e o valor beta é negativo; assim, podemos dizer que a mudança na higiene decresce quando uma pessoa passa do grupo dos ecléticos para o grupo dos *punks*. Tenha em mente que um decréscimo nos escores da higiene representa uma mudança maior (você está ficando mais fedorento), ou seja, a higiene decresce significativamente mais nos *punks* do que nos ecléticos.

A próxima variável *dummy* compara metaleiros a ecléticos. O valor beta novamente representa a mudança na higiene se uma pessoa eclética é comparada com um metaleiro. Se calcularmos a diferença entre as médias dos grupos para o grupo eclético e o dos metaleiros, obteremos: metaleiro – eclético = $(-0,5259 - (-0,5543)) = 0,028$. Esse valor é novamente igual ao valor beta na Saída 5.13 do SPSS! Para essa segunda variável *dummy*, o teste *t* não é significativo, assim, podemos dizer que a mudança dos escores da higiene é o mesmo se uma pessoa muda do grupo eclético para o metaleiro. Em outras palavras, a mudança nos escores da higiene não é previsível se uma pessoa é metaleira comparada com uma eclética.

Em nossa última variável *dummy*, comparamos *punks* com ecléticos. O valor beta novamente representa a alteração nos escores da higiene se uma pessoa sem afiliação musical é comparada com um pagodeiro. Se calcularmos a diferença nas médias dos grupos entre os sem afiliação e os pagodeiro, obteremos: pagodeiro – sem afiliação = $(-0,9643) - (-0,5543) = -0,410$. Não deve ser mais surpresa agora que esse valor é o beta da Saída 5.13 do SPSS! O teste *t* é significativo e o beta tem um valor negativo, assim, como ocorreu com a primeira variável *dummy*, podemos dizer que a mudança nos escores da higiene diminui se uma pessoa muda de sem afiliação para pagodeiro. Lembre-se de que um decréscimo nos escores da higiene representa mais alterações (você está se tornando mais fedido); ou seja, pagodeiros são significativamente mais fedorentos do que os ecléticos!



De forma geral, essa análise mostrou que, comparado com o grupo ecléticos, os grupos de *punks* e pagodeiros ficam mais fedidos ao longo dos três dias do festival, mas o grupo de metaleiros não. Esta seção introduziu algumas ideias complexas que irei detalhar nos Capítulos 7 e 8. Se você estiver muito confuso ou quiser saber mais sobre como variáveis *dummy* funcionam, sugiro que você leia as

Seções 7.8 e 8.2.2 e depois volte para cá. Outra dica é ler a excelente monografia de Hardy (1993)!

5.11 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Este é o capítulo mais longo do livro, e se você sente que envelheceu alguns anos enquanto o lia, isso provavelmente aconteceu (olhe ao redor: existem teias de aranha na sala, sua barba cresceu e se você olhar pela janela irá ver que uma segunda era glacial aconteceu e no planeta restam somente a sua pessoa e alguns mamutes). Contudo, analisando pelo lado bom, agora você sabe um bocado sobre regressão, que é a base de praticamente toda a estatística! Os próximos conteúdos deste livro são apenas variações sobre o tema deste capítulo.

Começamos este capítulo analisando a situação quando se tem um preditor e uma saída. Isso nos permitiu verificar alguns princípios básicos, como a equação da linha, o método dos mínimos quadrados e o modo de avaliar se o nosso modelo adere aos dados utilizando algumas quantidades importantes que você verá novamente nos próximos capítulos: a soma dos quadrados do modelo, SS_M , a soma dos quadrados dos resíduos, SS_R , e a soma total dos quadrados, SS_T . Utilizamos esses valores para calcular várias estatísticas importantes, tais como o R^2 e a razão F. Também aprendemos a executar uma regressão no SPSS e a colocar o valor beta resultante na equação da linha para fazer uma previsão sobre nossa saída.

Depois, descobrimos que **a equação da linha pode ser generalizada para incluir vários preditores e verificamos diversos métodos de entrar com as variáveis preditoras no modelo (hierárquico, entrada forçada e passo a passo)**. Vimos os fatores que podem afetar a precisão de um modelo (valores atípicos e casos influentes) e formas de identificar esses fatores. Em seguida, analisamos as suposições necessárias a fim de generalizar nosso modelo para além da amostra de dados que coletamos antes de descobrir como fazer a análise no

SPSS e como interpretar a saída, criar nosso modelo de regressão múltipla e testar sua confiabilidade e generalidade. Terminei o capítulo ensinando como podemos utilizar preditores categóricos na regressão (e de passagem descobrimos sobre a função de recodificação [*recode*]). Em geral, a regressão múltipla é um longo processo que deve ser feito com cuidado e com atenção aos detalhes. Existem muitos pontos importantes a considerar e você deve utilizar uma abordagem sistemática. Espero que este capítulo o ajude a fazer isso. Um aspecto que não consideramos foi se podemos utilizar variáveis categóricas como saída (em vez de preditores) e o próximo capítulo falará sobre um tipo específico de regressão que pode fazer exatamente isso! Antes, porém, preciso tomar uma dose de Valium.

5.12 TERMOS-CHAVE QUE DESCOBRIMOS

- Aderência
- Autocorrelação
- b_i
- β_i
- Colinearidade perfeita
- DFBeta
- DFBeta padronizado
- DFFit
- DFFit padronizado
- Distância de Cook
- Distância de Mahalanobis
- Efeitos supressores
- Encolhimento
- Erros independentes
- Estatística t
- Fator de inflação da variância (FIV)
- Generalização
- Heterocedasticidade
- Homocedasticidade
- Influência
- Multicolinearidade
- Quadrados médios
- R múltiplo
- R^2 Ajustado
- Razão de Covariância (CVR)
- Razão F

- Regressão múltipla
- Regressão passo a passo
- Regressão simples
- Resíduo
- Resíduo excluído
- Resíduos estudentizados
- Resíduos excluídos estudentizados
- Resíduos não-padronizados
- Resíduos padronizados
- Soma dos quadrados do modelo
- Soma dos quadrados dos resíduos
- Soma total dos quadrados
- Teste de Durbin-Watson
- Tolerância
- Validação cruzada
- Valor previsto ajustado
- Valores Chapéu
- Variável de saída
- Variável *dummy*
- Variável previsora

5.13 TAREFAS DO ALEX ESPERTO

As respostas para essas tarefas podem ser encontradas no *site* www.artmed.com.br no arquivo **(Chapter 5).pdf**.

- Uma estudante de moda está interessada em fatores que possam prever os salários de modelos de passarela. Ela coletou dados de 231 modelos. Para cada modelo, ela perguntou o salário diário nos dias em que elas estavam trabalhando (**salary**), a idade (**age**), quanto tempo elas estão trabalhando como modelos (**years**). A estudante também reuniu um painel de especialistas de agências de modelo a fim de dar uma nota para a atratividade de cada modelo, com uma percentagem de 100% representando o máximo de atratividade (**beauty**). Os dados estão no *site* www.artmed.com.

br no arquivo **Supermodel.sav**. Infelizmente, essa estudante de moda comprou alguns textos de estatística não muito bons e não sabe como analisar os seus dados. ☺ Você pode ajudá-la conduzindo uma regressão múltipla para verificar quais fatores preveem o salário da modelo? Quão válido é o modelo de regressão? ②

- Utilizando os dados de Glastonbury deste capítulo (com os códigos *dummy* em **GlastonburyDummy.sav**), que você já analisou, comente se você julga que o modelo é confiável e generalizável. ③

5.14 LEITURAS COMPLEMENTARES

- BOWERMAN, B. L., O'CONNELL, R. T. *Linear statistical models: an applied approach*. Belmont (CA): Duxbury, 1990. 2ª ed. Esse texto é destinado somente para os que gostam de matemática ou estudantes de pós-graduação, mas fornece uma exposição extremamente abrangente da análise de regressão.
- HARDY, M. A. *Regression with dummy variables*. Sage university paper series on quantitative applications in the Social Sciences. 07-093. Newbury Park (CA): Sage, 1993.
- HOWEL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 2002. 5ª ed. Os Capítulos 9 e 15 são excelentes introduções à matemática por trás dos modelos de regressão.
- MILES, J., SHEVILIN, M. *Applying regression and correlation: a guide for students and researchers*. London: Sage, 2001. Esse texto bastante simples aborda a regressão em detalhes com um mínimo de esforço. Altamente recomendável.
- STEVENS, J. *Applied multivariate statistics for the social sciences*. Hillsdale (NJ): Erlbaum, 1992. 2ª ed. Capítulo 3.

REGRESSÃO LOGÍSTICA

6.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

No capítulo anterior, vimos como modelar o relacionamento entre uma ou mais variáveis previsoras e uma saída. Este capítulo amplia esse assunto para verificar como podemos prever uma variável de saída categórica utilizando a *regressão logística*. É um tema difícil (de fato, eu poderia tê-lo enchido de quadros do Alex Esperto, mas não o fiz!) e a regressão logística não é um dos meus pontos fortes, assim, provavelmente escrevi algumas bobagens. No entanto, começamos com um pouco de teoria para ajudá-lo a entender a regressão logística, antes de irmos direto para um exemplo. Você irá aprender a executar a regressão logística no SPSS e a interpretar a saída resultante. Concluímos o capítulo com um segundo exemplo e usaremos esse exemplo a fim de ver como combater a multicolinearidade. Ao longo do texto, também descobriremos por que a seleção de futebol inglesa não sabe cobrar pênaltis.

6.2 PRESSUPOSTOS DA REGRESSÃO LOGÍSTICA ①

Em poucas palavras, a regressão logística é uma regressão múltipla, mas com uma variá-

vel de saída categórica dicotômica e variáveis previsoras contínuas ou categóricas. Simplificando, isso quer dizer que podemos prever a qual de duas categoriais é provável que uma pessoa pertença dado certas informações. Um exemplo trivial é determinar que variáveis podem definir se uma pessoa é homem ou mulher. Você poderá mensurar a preguiça, a estupidez, o consumo de álcool, o número de arrotos que a pessoa dá em um dia. Utilizando a regressão logística, podemos verificar que todas essas variáveis são capazes de prever o gênero de uma pessoa, e a técnica nos apontará também se uma pessoa tem certa probabilidade de ser homem ou mulher. Assim, se você pegar uma pessoa ao acaso e descobrir que ela teve um escore alto em preguiça, estupidez, consumo de álcool e número de arrotos, então o modelo de regressão poderá nos dizer, com base nessa informação, quão provável é que essa pessoa seja homem.

É improvável que um pesquisador esteja interessado no relacionamento entre flatulência e gênero (isso é suficientemente comprovado pela experiência para justificar uma pesquisa!), mas a regressão logística pode ter aplicações vitais. Na literatura biomédica (isto é, pesquisa médica), a regressão logística tem aplicações tais como a de formular modelos sobre os tipos

de fatores que determinam se um tumor é cancerígeno ou benigno. Uma base de dados de pacientes pode ser utilizada para identificar as variáveis que são influentes na previsão do tipo de um tumor. Essas variáveis podem então ser mensuradas em um novo paciente e seus valores colocados no modelo de regressão logística a partir do qual é possível estimar uma probabilidade de o tumor ser maligno. Se a probabilidade do tumor ser maligno é baixa, o médico pode decidir não executar uma cara e dolorosa cirurgia que seria desnecessária.

Nas ciências sociais, raramente enfrentamos decisões de vida ou morte, mas mesmo assim a regressão logística é uma ferramenta bastante útil. Por esse motivo, é triste que muitos livros-texto abordem o tema superficialmente. Neste capítulo, explicarei os princípios por trás da regressão logística e como executar os procedimentos no SPSS.

6.3 QUAIS SÃO OS PRINCÍPIOS POR TRÁS DA REGRESSÃO LOGÍSTICA? ③

Não pretendo entrar nos princípios subjacentes da regressão logística porque eles não são necessários para entender o teste (eu sou uma prova viva disso). Contudo, quero fazer algumas comparações com a regressão usual a fim de inserir a regressão logística em uma estrutura que seja familiar a qualquer um que tenha chegado até aqui no livro. (O quê? Você não leu o capítulo de regressão ainda?!) Se você tem fobia de equações, é melhor olhar para o outro lado. Em uma regressão linear simples, vimos que a variável de saída Y é prevista a partir da equação da linha:

$$Y = b_0 + b_1X + \varepsilon \quad (6.1)$$

Onde b_0 é o intercepto e b_1 é o gradiente da linha, X é o valor da variável previsora e ε é o termo resíduo. Dados os valores de Y e X , os parâmetros desconhecidos na equação podem ser estimados encontrando-se a solução para a qual as distâncias ao quadrado entre os valores observados e previstos da variável dependente sejam mínimas (método dos mínimos quadrados).

Esse assunto deve ser bem familiar. Na regressão múltipla – em que existem vários previsores, – uma equação semelhante é derivada na qual cada predictor tem seu próprio coeficiente. Como tal, Y é previsto a partir de uma combinação de cada variável previsora multiplicada pelo seu respectivo coeficiente de regressão:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon \quad (6.2)$$

Onde b_n é o coeficiente de regressão da correspondente variável X_n . Na regressão logística, em vez de se prever o valor da variável Y a partir de um predictor X ou diversas variáveis predictoras (X s), prevemos a *probabilidade* de Y ocorrer conhecidos os valores de X ou X s. A equação de regressão logística apresenta várias semelhanças com a equação de regressão recém-descrita. Na sua forma mais simples, quando existe um único predictor X , a equação de regressão logística a partir da qual a probabilidade da variável Y é prevista é dada por:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1X + \varepsilon_i)}} \quad (6.3)$$

Onde $P(Y)$ é a probabilidade de Y ocorrer, e é a base dos logaritmos naturais e os demais coeficientes da equação formam uma combinação linear muito semelhante à regressão simples. De fato, você pode ter notado que a parte da equação dentro dos parênteses é idêntica à regressão linear, pois existe uma constante (b_0), uma variável previsora (X) e um coeficiente (ou peso) agregado ao predictor (b_1). Da mesma forma que para a regressão linear, é possível estender essa equação para incluir diversas variáveis predictoras. Quando isso ocorre, a equação toma a seguinte forma:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon)}} \quad (6.4)$$

A equação (6.4) é a mesma utilizada quando existe um único predictor, exceto que a combinação linear foi estendida para incluir qualquer número de previsores. Assim, enquanto que a versão com um predictor da equação de regressão logística contém a equação

de regressão linear simples em seu interior, a versão com múltiplos previsores contém a equação de regressão múltipla.

Por que não posso usar a regressão linear?



A despeito das semelhanças entre a regressão linear e a logística, existe uma boa razão porque não podemos aplicar a regressão linear diretamente a uma situação onde a variável de saída é dicotômica: uma das hipóteses da re-

gressão linear é que o relacionamento entre as variáveis é linear (veja a Seção 5.6.2.1). Vimos anteriormente a importância das hipóteses para se obter um modelo preciso. Portanto, para o modelo de regressão linear ser válido, os dados observados devem ter um relacionamento linear. Quando a variável de saída é dicotômica, essa hipótese é normalmente violada (veja Berry, 1993). Uma forma de contornar esse problema é alterar os dados por meio de uma transformação logarítmica (veja Berry e Feldman, 1985, Capítulo 3). Isso mantém a *forma* do relacionamento linear enquanto o próprio relacionamento em si é não-linear (assim, é um modo de expressar um relacionamento não-linear em uma maneira linear). A equação de regressão logística descrita anteriormente está baseada no seguinte princípio: ela expressa uma equação de regressão linear múltipla em termos logarítmicos e dessa forma resolve o problema da violação da hipótese de linearidade.

A forma exata da equação pode ser arranjada de várias maneiras, mas a versão que eu escolhi expressa a equação em termos da probabilidade de Y ocorrer (isto é, a probabilidade de que um caso pertença a uma determinada categoria). O valor resultante da equação é a uma probabilidade e, como tal, varia sempre entre 0 e 1. Um valor próximo de 0 significa que a ocorrência de Y é bastante improvável e um valor próximo de 1, que ela é bem provável. Além disso, da mesma forma que na regressão linear, na equação logística cada variável preditora tem seu próprio coeficiente. Quando executamos a análise precisamos estimar os valores desses coeficientes para que possamos utilizar a

equação. Esses parâmetros são estimados pelo ajustamento de modelos, com base nos previsores disponíveis, aos dados observados. O modelo escolhido será aquele que quando os valores das variáveis predictoras forem utilizados resulta em um valor de Y mais próximo do valor observado. Especificamente, os valores dos parâmetros são estimados utilizando a **estimação de máxima verossimilhança** que seleciona os coeficientes que tornam os valores observados mais prováveis de terem ocorrido. Assim como ocorreu com a regressão múltipla, tentamos ajustar um modelo aos nossos dados que nos permita estimar valores da variável de saída a partir de valores das variáveis predictoras.

6.3.1 Avaliando o modelo: a estatística de verossimilhança-log ③

Como vimos, o modelo de regressão logística prevê a probabilidade de um evento ocorrer para uma dada pessoa (representaríamos isso como $P(Y_i)$, a probabilidade de que Y ocorra para a pessoa i), baseado em observações de se um evento ocorreu ou não para essa pessoa (representaríamos isso como Y_i , a saída real para a i -ésima pessoa). Assim, para uma dada pessoa, Y poderá ser 0 (a saída não ocorre) ou 1 (a saída ocorre), e o valor previsto, $P(Y)$, será um valor entre 0 (não existe possibilidade de que a saída ocorra) e 1 (a saída ocorrerá). Vimos na regressão múltipla (consulte o capítulo anterior) que para avaliar se um modelo adere aos dados, comparamos os valores observados e os previstos (se você lembrar, utilizamos R^2 , que é a correlação de Pearson ao quadrado entre os valores observados de saída e os valores previstos pelo modelo de regressão). Da mesma forma, na regressão logística, podemos utilizar os valores observados e previstos para avaliar a aderência do modelo. Para fazer isso, utilizamos a **verossimilhança-log (VL)**:

verossimilhança-log =

$$\sum_{i=1}^N \{Y_i \ln(P(Y_i)) + (1 - Y_i) \ln[1 - P(Y_i)]\} \quad (6.5)$$

A verossimilhança-log é, portanto, baseada na soma das probabilidades associadas com a saída real e a prevista (ver Tabachnick e Fidell, 2001). A estatística de verossimilhança-log é análoga à soma dos resíduos ao quadrado na regressão múltipla, no sentido de que ela é um indicador de quanta informação não explicada ainda existe após o modelo ter sido ajustado. Como consequência, tem-se que valores altos da estatística de verossimilhança-log indicam uma aderência pobre do modelo, porque quanto maior for esse valor, mais observações não explicadas existirão.

É possível calcular a verossimilhança-log para diferentes modelos e comparar esses modelos por meio das diferenças entre os resultados da estatística de verossimilhança-log. Um uso para isso é comparar o estado do modelo de regressão contra algum tipo básico de estado. O estado básico utilizado é normalmente o modelo quando apenas a constante está incluída. Na regressão múltipla, o modelo básico utilizado foi a média de todos os valores da variável Y (ela era a nossa melhor estimativa quando nenhuma outra informação estava disponível). Na regressão logística, se quiséssemos prever um resultado, qual seria o nosso melhor palpite? Bem, não podemos utilizar os valores de saída porque eles são formados por 0s e 1s e, assim, a média não tem sentido! Contudo, se soubermos a frequência de 0s e 1s, nosso melhor palpite seria a categoria com o maior número de casos. Dessa forma, se um valor de saída ocorre 107 vezes e não ocorre em 72 vezes, nossa melhor estimativa seria de que ele irá ocorrer (porque temos 107 ocorrências contra 72 não-ocorrências). Como tal, em uma regressão múltipla, nosso modelo base nos dá a melhor previsão na falta de qualquer outra informação: na regressão logística, essa mesma situação seria prever a saída que ocorre com maior frequência. Esse é o modelo de regressão logística quando apenas a constante for incluída. Se adicionarmos um ou mais previsores ao modelo, podemos calcular a melhoria do modelo da seguinte forma:

$$\chi^2 = 2[\text{VL}(\text{Novo}) - \text{VL}(\text{Básico})] \quad (6.6)$$

$$(gl = k_{\text{Novo}} - k_{\text{Básico}})$$

Assim, simplesmente pegamos o novo modelo e o subtraímos do modelo básico (o modelo quando somente a constante está presente). Você deve ter notado que multiplicamos esse valor por 2; fazemos isso porque ele é um resultado da distribuição qui-quadrado (veja o Capítulo 16 e o Apêndice A4) e, dessa forma, fica fácil calcular a significância do valor. A distribuição qui-quadrado que utilizamos tem graus de liberdade igual ao número de parâmetros no novo modelo menos o número de parâmetros no modelo básico. O número de parâmetros, k , no modelo básico é sempre igual a 1 (a constante é o único parâmetro a ser estimado) e qualquer modelo subsequente terá um número de graus de liberdade igual ao número de previsores mais 1 (isto é, o número de previsores mais o parâmetro representando a constante).

6.3.2 Avaliando o modelo: R e R^2 ③

Quando falamos sobre regressão linear, vimos que o coeficiente de correlação múltiplo R e o correspondente R^2 foram medidas úteis para avaliar como o modelo se ajustava aos dados. Vimos também que a razão

de verossimilhança é semelhante no sentido de que ela está baseada no nível de correspondência entre os valores de saídas previstos e observados. Contudo, é possível calcular uma versão mais adequada da correlação múltipla na regressão logística por um valor conhecido como estatística- R . Essa estatística- R é a correlação parcial entre a variável de saída e cada uma das variáveis predictoras e podem variar de -1 a 1 . Um valor positivo indica que quando a variável predictoradora aumenta, também aumenta a probabilidade da ocorrência do evento. Um valor negativo indica que se a variável predictoradora aumenta, a probabilidade do valor de saída ocorrer diminui. Se uma variável tem

Existe um valor equivalente ao R^2 para a regressão logística?



um valor pequeno de R , então ela contribui pouco para o modelo.

O cálculo de R é fornecido na equação (6.7). O $-2VL$ é o -2 verossimilhança-log para o modelo original, a estatística de Wald é calculada como o descrito na próxima seção e os graus de liberdade podem ser lidos na tabela resumo das variáveis na equação. Contudo, em virtude do valor de R ser dependente da estatística de Wald, ele não é uma medida precisa (veremos na próxima seção que a estatística de Wald pode não ser precisa em certas circunstâncias). Por esse motivo, o valor de R deve ser tratado com alguma cautela e é inválido elevá-lo ao quadrado e interpretá-lo como foi feito na regressão linear.

$$R = \pm \sqrt{\frac{\text{Wald} - (2 \times gl)}{-2VL(\text{Original})}} \quad (6.7)$$

Existe controvérsia sobre o que seria um bom valor semelhante ao R^2 da regressão linear, mas uma medida descrita por Hosmer e Lemeshow (1989) pode ser facilmente calculada. Na terminologia do SPSS, o R_V^2 de **Hosmer e Lemeshow** é calculado como na equação (6.8):

$$R_V^2 = \frac{-2VL(\text{Modelo})}{-2VL(\text{Original})} \quad (6.8)$$

Como tal, o R_V^2 é calculado pela divisão do qui-quadrado do modelo (baseado na verossimilhança-log) pelo $-2VL$ -original (a verossimilhança-log para o modelo antes que qualquer preditor tenha sido adicionado). O R_V^2 é a redução proporcional no valor absoluto da medida verossimilhança-log e, desse modo, é uma medida de quanto a não-aderência aumenta como resultado da inclusão de uma variável preditora. Ele pode variar de 0 (indicando que os preditores são inúteis na previsão da variável de saída) a 1 (indicando que o modelo prevê perfeitamente a variável de saída).

Contudo, essa não é a medida utilizada pelo SPSS. O SPSS utiliza o R_{CS}^2 de **Cox e Snell** (1989), que é baseado na verossimilhança-log do modelo ($VL(\text{Novo})$) e a verossimi-

lhança-log do modelo original ($VL(\text{Básico})$) e o tamanho da amostra, n :

$$R_{CS}^2 = 1 - e^{\left[\frac{-2}{n} (VL(\text{Novo}) - VL(\text{Básico})) \right]} \quad (6.9)$$

Contudo, essa estatística nunca alcança o seu valor teórico máximo, 1. Portanto, Nagelkerke (1991) sugeriu a seguinte correção (R_N^2 de **Nagelkerke**):

$$R_N^2 = \frac{R_{CS}^2}{1 - e^{\left[\frac{2(VL(\text{Básico}))}{n} \right]}} \quad (6.10)$$

Embora todas essas medidas tenham diferenças na forma de cálculo (e nos resultados), coletivamente é possível considerá-las como praticamente as mesmas. Assim, em termos de interpretação elas podem ser vistas como similares ao R^2 da regressão linear no sentido de fornecer uma medida do grau de aderência do modelo.

6.3.3 Avaliando a contribuição dos preditores: a estatística de Wald ②

Como na regressão linear, queremos saber não apenas se o modelo se ajusta bem aos dados, mas também a contribuição individual de cada um dos preditores. Na regressão linear, utilizamos os coeficientes estimados da regressão (b) e seus erros padrão para calcular a estatística t . Na regressão logística, existe uma estatística análoga a de Wald que apresenta uma distribuição especial conhecida como **qui-quadrado**. Da mesma forma que o teste t na regressão linear, a estatística de Wald nos informa se o coeficiente b de cada preditor é significativamente diferente de zero. Se isso ocorrer, poderemos assumir que o preditor está contribuindo de forma significativa para a previsão da variável de saída (Y). A equação (6.11) mostra como a estatística de Wald é calculada e é possível ver que basicamente ela é igual à estatística t na regressão linear (veja a equação (5.6)); ela é o valor do coeficiente de regressão dividido pelo seu erro padrão associado. A estatística de Wald é geralmente utilizada para determinar se uma variável é um preditor signi-

ficativo da saída; contudo, ela é provavelmente mais precisa para examinar a estatística da razão de verossimilhança. A estatística de Wald (Figura 6.1) deve ser vista com cautela porque quando o coeficiente de regressão (b) é grande, o erro padrão tende a ficar inflacionado, resultando em uma estatística de Wald subestimada (veja Menard, 1995). A inflação do erro padrão aumenta a probabilidade de que o previsor seja rejeitado quando na realidade ele contribui para o modelo (isto é, você tem uma probabilidade maior de cometer o erro do Tipo II).

$$\text{Wald} = \frac{b}{\text{EP}_b} \quad (6.11)$$

6.3.4 Exp b ③

O mais importante para a interpretação da regressão logística é o valor da exp b ($\text{Exp}(B)$ na saída do SPSS), é um indicador da mudança nas probabilidades resultantes da mudança de uma unidade no previsor. Ele é similar aos coeficientes b da regressão logística, mas mais fácil de entender (porque não requer uma transformação logarítmica). Quando a variável previsor é categórica, o exp b é mais fácil de explicar; assim, imagine que temos um exemplo simples no qual estamos tentando prever a possibilidade de gravidez de uma mulher a par-

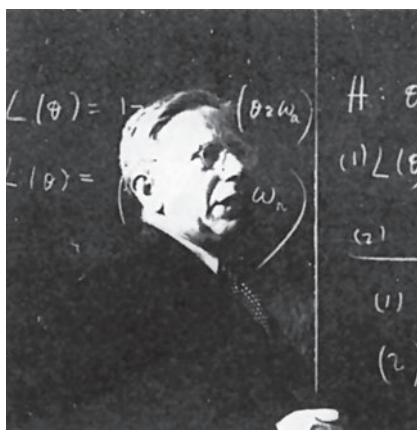


Figura 6.1 Abraham Wald escrevendo “eu não devo construir estatísticas teste propensas a ter erros padrão inflados” no quadro-negro cem vezes.

tir da informação do uso ou não de preservativo pelo parceiro com quem ela se relacionou. A chance de um evento ocorrer é definida como a probabilidade dele ocorrer dividida pela probabilidade dele não ocorrer (veja a equação 6.12) e não deve ser confundida com a forma mais coloquial da palavra que se refere à probabilidade. Assim, por exemplo, a chance de engravidar é a probabilidade de engravidar dividida pela probabilidade de não engravidar:

$$\text{Chance} = \frac{P(\text{evento ocorrer})}{P(\text{evento não ocorrer})}$$

$$P(\text{evento } Y) = \frac{1}{1 + e^{-(b_0 + b_1 X)}} \quad (6.12)$$

$$P(\text{evento } Y \text{ não ocorrer}) = 1 - P(\text{evento } Y \text{ ocorrer})$$

Para calcular a mudança na chance que resulta da mudança de uma unidade no previsor, primeiro precisamos calcular as chances de engravidar quando não foi utilizado preservativo. Depois calculamos as chances de engravidar quando foi usado preservativo. Finalmente, calculamos a proporção entre essas duas chances.

Para calcular o primeiro conjunto de chances, precisamos utilizar a equação (6.3) a fim de calcular a probabilidade de engravidar quando não foi utilizado preservativo. Se tivéssemos mais do que um previsor, usaríamos a equação (6.4). Existem três quantidades desconhecidas nessa equação: o coeficiente da constante (b_0), o coeficiente do previsor (b_1) e o valor do próprio previsor (X). Saberemos o valor de X a partir de como foi codificada a variável de utilização do preservativo (as chances são: não utilizado = 0 e utilizado = 1). Os valores de b_0 e b_1 serão estimados por nós. Podemos calcular as chances como na equação (6.12).

Em seguida, fazemos o mesmo cálculo após a variável previsor ter mudado de uma unidade. Nesse caso, em virtude do previsor ser dicotômico, precisamos calcular a chance de gravidez dado que o preservativo foi utilizado. Assim, o valor de X agora é 1 (em vez de 0).

Nós conhecemos as chances antes e depois da mudança de uma unidade na variável previsor. É apenas uma questão de calcular a

mudança proporcional nas chances que pode ser feita dividindo as chances após a mudança de uma unidade na variável previsora pelas chances antes de tal mudança.

$$\Delta\text{Chance} = \frac{\text{Probabilidade após a mudança de uma unidade no predictor}}{\text{Probabilidade original}} \quad (6.13)$$

Essa proporção de mudança na chance é o $\exp b$, assim, podemos interpretar o $\exp b$ em relação à mudança na chance: se o valor é maior do que 1 significa que quando o predictor aumenta, as chances da saída ocorrer aumentam. Já um valor menor do que 1 indica que quando o predictor aumenta, as chances da saída ocorrer diminuem. Logo fornecerei um exemplo real de como isso funciona.

6.3.5 Métodos de regressão logística ②

Assim como na regressão múltipla, existem vários métodos diferentes que podem ser utilizados na regressão logística.

6.3.5.1 O método da entrada forçada

O método padrão de conduzir essa regressão é o “*enter*”. Ele é o mesmo que a entrada forçada da regressão múltipla, em que todas as covariáveis são colocadas no mesmo modelo de regressão em um único bloco e as estimativas dos parâmetros são calculadas para cada bloco. Alguns pesquisadores acreditam que esse método é o único apropriado para testar teorias (Studenmund e Cassidy, 1987) porque a técnica passo a passo é influenciada por variações aleatórias nos dados e raramente fornece resultados replicáveis se o modelo é novamente executado com a mesma amostra.

6.3.5.2 Métodos passo a passo (*stepwise*)

Se você não ficou assustado com as críticas aos métodos passo a passo no capítulo anterior, pode selecionar tanto o método para frente quanto o para trás. Quando o método para frente é empregado, o computador inicia com um modelo que inclui apenas a constan-

te e depois adiciona os previsores um a um no modelo com base em critérios específicos. Esse critério é o valor da estatística *escore*: a variável com a estatística *escore* mais significativa é adicionada ao modelo. O computador segue até que nenhum dos previsores restantes tenha uma estatística *escore* significativa (o ponto de corte da significância inicia com 0,05). A cada passo, o computador também examina as variáveis no modelo para ver se alguma deve ser removida. Ele faz isso em uma das três seguintes maneiras. A primeira é utilizar a estatística Razão de Verossimilhança descrita em 16.2.2 (o método RV para a frente), em que o modelo atual é comparado ao modelo quando o predictor é removido. Se a remoção do predictor acarreta uma diferença significativa para o ajuste do modelo aos dados observados, o computador retém o predictor (porque o modelo fica melhor quando o predictor é incluído). Se, contudo, a remoção do predictor acarreta uma diferença muito pequena no modelo, o computador rejeita o predictor. Em vez de utilizar a estatística da razão de verossimilhança, que estima o quão bem o modelo adere aos dados observados, o computador pode utilizar uma estatística condicional como um critério de remoção (*Condicional: para frente*). Essa estatística é uma versão aritmeticamente menos intensa que a estatística da razão de verossimilhança, assim, não há muita vantagem em utilizá-la. O critério final é a estatística de Wald, na qual cada predictor no modelo que tem um valor significativo da estatística de Wald (acima do critério padrão de remoção de 0,1) será removido. Entre esses métodos, o melhor é o da estatística da razão de verossimilhança, pois a estatística de Wald pode, às vezes, não ser confiável (Seção 6.3.3).

O oposto do método para a frente é o para trás. Esse método utiliza os mesmos três critérios de remoção, mas em vez de começar o modelo somente com uma constante ele começa com todos os previsores incluídos. O computador então testa se qualquer um dos previsores poderá ser removido sem causar um efeito substancial no grau de aderência do modelo aos dados observados. O predictor que

apresentar o menor impacto no ajuste do modelo aos dados será removido primeiro.

6.3.5.3 Como selecionamos um método? ②



Da mesma forma que a regressão linear (capítulo anterior), o método de regressão escolhido dependerá de vários fatores. A principal consideração é se

você está testando uma teoria ou está fazendo um trabalho exploratório. Como já foi dito, algumas pessoas acreditam que os métodos passo a passo (*stepwise*) não são bons para testar teorias. Contudo, os métodos passo a passo são apropriados quando não existem pesquisas prévias que podem ser tomadas por base para testar hipóteses e em situações onde a causalidade não é de interesse e você quer somente achar um modelo para ajustar os dados (Menard, 1995; Agresti e Finlay, 1986). Também, como foi colocado para a regressão linear, se você decidir utilizar um método passo a passo, o método para trás (*backward*) é preferível ao método para frente (*forward*). Isso é em virtude do **efeito supressor**, que ocorre quando um previsor tem um efeito significativo somente quando outra variável é mantida constante. A seleção para frente é mais provável que a eliminação para trás para excluir previsores envolvidos em efeitos supressores. Desse modo, o método oferece um alto risco de se cometer o Erro do Tipo II. Em termos da estatística teste utilizada em métodos passo a passo, a estatística de Wald tende a ser pouco precisa em certas circunstâncias (veja a Seção 6.3.3), portanto, a razão de verossimilhança é melhor.

6.4 EXECUTANDO A ANÁLISE: UM EXEMPLO DE PESQUISA ②

Como meu primeiro exemplo de pesquisa, vou apresentar um de psicologia do desenvolvimento. Um bom exemplo de uma variável dicotômica é passar ou não em um teste, que

em termos de psicologia do desenvolvimento é referido como tendo ou não uma determinada habilidade cognitiva. Esse exemplo está relacionado com o entendimento de regras de desenvolvimento por uma

criança com base na sua idade e se a criança apresenta uma teoria da mente. Simplificando, uma regra de desenvolvimento é uma convenção de demonstração de uma emoção apropriada em uma dada situação. Por exemplo, se você ganha um presente de Natal que não gosta, a regra apropriada de comportamento é sorrir polidamente e dizer “obrigado tia Kate, eu sempre quis um repolho podre”. O comportamento emocional inapropriado é começar a chorar e gritar “por que você me comprou um repolho podre, sua velha chata?”. Existem algumas evidências de que crianças pequenas não têm um entendimento do comportamento apropriado e isso está conectado à teoria do desenvolvimento do cérebro (que é simplesmente a habilidade de entender o que a outra pessoa pode estar pensando).

Para esse exemplo, nossos pesquisadores estão interessados em saber se o entendimento das regras de desenvolvimento emocionais está relacionado a ter uma teoria da mente. Eles acreditam que uma criança precisa entender como outra pessoa pensa para perceber como o seu comportamento emocional irá afetar essa pessoa: se você não conseguir se imaginar no lugar da tia Katie, então você não irá entender que ela poderá ficar muito chateada se você chamá-la de velha chata. Para testar essa teoria, foi atribuída a várias crianças uma tarefa padrão de falsas crenças (uma tarefa usada para medir se alguém tem uma teoria da mente) em que elas poderiam passar ou não e suas idades em meses foram medidas. Além disso, a cada criança foi dada uma tarefa com uma regra de comportamento que elas poderiam também passar ou não. Assim, as seguintes variáveis foram mensuradas:

Por que você me deu este livro de péssima qualidade como presente de Natal, tia Kate?



- **Saída (variável dependente):** ter um entendimento das regras de comportamento (a criança passou no teste: Sim/Não?).
- **Previsor (variável independente):** ter ou não uma teoria da mente (a criança passou na tarefa de falsas crenças: Sim/Não?).
- **Previsor (variável independente):** Idade em meses.

Nesse experimento, existe uma variável de saída dicotômica, um previsor categórico e um previsor contínuo. O cenário é ideal para uma regressão logística.

6.4.1 A análise principal ②

Para executar uma regressão logística, a entrada dos dados deve ser feita como na regressão linear: eles são arranjados no editor de dados em três colunas (uma para cada variável). Os dados podem ser encontrados no arquivo **display.sav** na pasta do Capítulo 6 no site www.artmed.com.br. Ao observar o editor de dados, você deverá notar que as duas variáveis categóricas entraram como variáveis codificadas (veja a Seção 2.4.4); isto é, os números foram especificados como representantes de categorias. Para facilitar a interpretação, a variável de saída deve ser codificada como 1 (evento ocorreu) e 0 (evento não ocorreu); neste caso, 1 representa ter um entendimento das regras de comportamento e 0 representa a ausência de compreensão das

regras. Para a tarefa das falsas crenças, um código semelhante foi utilizado (1 = passou na tarefa das falsas crenças e 2 = falhou na tarefa das falsas crenças). A regressão logística está localizada em **Regression** (Regressão), que é acessado pelo menu **Analyze** (Analisar): **Analyze**⇒**Regression**⇒**Binary logistic** (Analisar⇒Regressão⇒Logística Binária). Seguindo esse caminho a caixa de diálogo principal para a regressão logística será ativada conforme mostrado na Figura 6.2.

A caixa de diálogo principal é semelhante à da regressão padrão. Existe um espaço para colocar a variável dependente (variável de saída). Nesse exemplo, a saída é o entendimento ou não das regras de comportamento, assim, você simplesmente pode clicar em **display** e transferi-la para o quadro **Dependent** (Dependente) clicando em **►**. Existe também um quadro para especificar as covariáveis (as variáveis previsoras). É possível especificar os efeitos principais e as interações na regressão logística. Para especificar um efeito principal, selecione um previsor (por exemplo, **age** [idade]) e transfira essa variável para o quadro denominado **Covariates** (Covariáveis) clicando em **►**. Para entrar com uma interação, clique em mais de uma variável no lado esquerdo da caixa de diálogo (por exemplo, marque duas ou mais variáveis) e depois clique em **>a*b>** a fim de movê-las para o quadro das **Covariates** (Covariáveis). Nesse exemplo, existem apenas

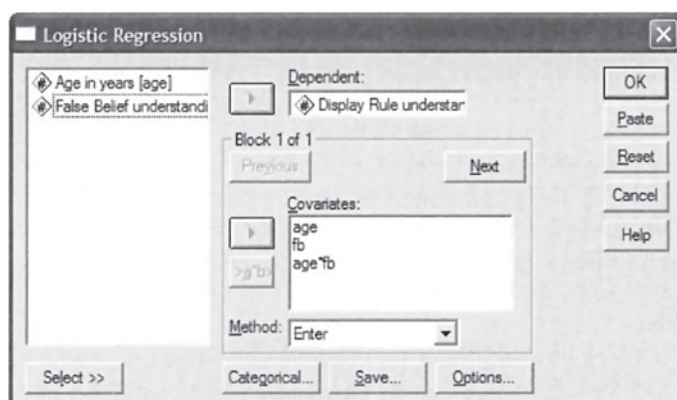


Figura 6.2 Caixa de diálogo principal da regressão logística.

dois previsores, portanto, existe apenas uma possível interação (a interação entre idade [age] \times falsas crenças [fb]), mas se você tiver três ou mais previsores, poderá selecionar várias interações dois a dois assim como a interação das três variáveis.

6.4.2 Métodos de regressão ②

Assim como na regressão múltipla, existem vários métodos diferentes que podem ser utilizados na regressão logística (veja a Seção 6.3.5). Você pode selecionar um método específico de regressão clicando na seta que aponta para baixo próxima ao quadro denominado *Method* (Method). A Figura 6.3 mostra parte do menu da regressão logística quando os métodos de regressão estão ativos. Para essa análise, selecione o método *Forward:LR*. Depois de eu ter escrito que você nunca deveria fazer análises passo a passo, você deve estar achando um tanto estranho eu sugerir a utilização do método para a frente (forward) aqui. Bem, esse estudo é o primeiro na área e, assim, não existem pesquisas prévias que nos apontem quais previsores são confiáveis. Também, ainda não lhe mostrei um exemplo de regressão passo a passo, portanto, essa será uma maneira útil de demonstrar como o procedimento funciona!

6.4.3 Previsores categóricos ②

Nesse exemplo, existe uma variável previsora categórica. Uma das qualidades da regressão logística é que ela aceita previsores categóricos. Contudo, é necessário indicar ao SPSS quais variáveis, se existirem, são categóricas, clicando em *Categorical* na caixa de diálogo da regressão logística para ativar uma nova caixa de diálogo semelhante a da Figura 6.4.

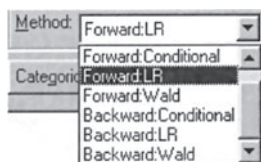


Figura 6.3 Métodos de regressão.

Nessa caixa de diálogo, as covariáveis estão listadas no lado esquerdo e existe um quadro na direita no qual as covariáveis categóricas podem ser colocadas. Destaque as variáveis categóricas que você tem (nesse exemplo, clique na **fb**) e transfira-as para o quadro *Categorical Covariates* (Covariáveis Categóricas) clicando em . Na Figura 6.4, a **fb**, já está selecionada e transferida.

Existem várias maneiras de tratar previsores categóricos; por padrão, o SPSS irá utilizar os desvios dos contrastes em todas as variáveis categóricas. Na Seção 5.10 do capítulo anterior, vimos que os previsores categóricos poderiam ser incorporados na regressão pela codificação dos seus valores em zeros e uns (conhecido como código auxiliar – *dummy*). Agora existem maneiras diferentes de arranjar esses códigos dependendo do que você quer comparar. O SPSS tem várias formas padronizadas que poderão ser selecionadas (apresentaremos esses métodos em mais detalhes no Capítulo 8). Por padrão, o SPSS utiliza a codificação *Indicator* (Indicador), que é a variável auxiliar (*dummy*) padrão que já foi explicada na Seção 5.10 (e você pode escolher ainda a sua primeira ou última categoria como base). Se quiser mudar para um tipo diferente de contraste, clique na seta que aponta para baixo no quadro *Change Contrast* (Mudar contraste). A Figura 6.5 mostra que é possível selecionar contrastes simples, de diferenças, de Helmert, repetidos, polinomiais e contrastes de desvios

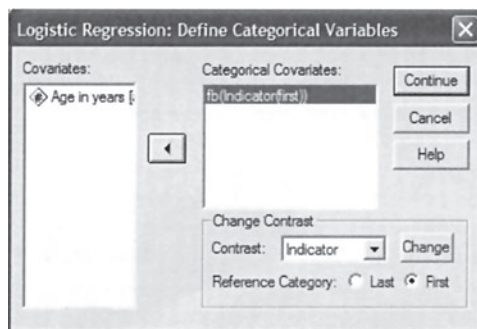


Figura 6.4 Definindo as variáveis categóricas na regressão logística.

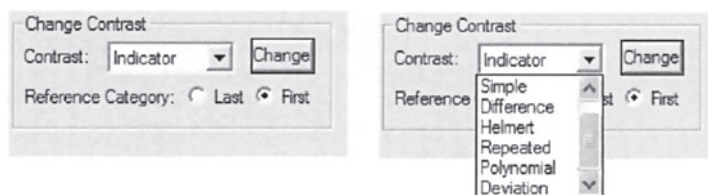


Figura 6.5 Selecionando contrastes para previsores categóricos.

da mesma forma que na ANOVA (ver Tabela 8.6). Essas técnicas serão discutidas em detalhes no Capítulo 8, onde será explicado o que um contraste faz. Contudo, você não irá precisar dos contrastes indicadores naquele capítulo, portanto, irei utilizá-lo aqui (como na Figura 6.4). Quando um contraste indicador é utilizado, os níveis da variável categórica são recodificados utilizando uma variável *dummy* padrão (veja as Seções 5.10 e 7.8).

6.4.4 Obtendo resíduos ②

Como na regressão linear, é possível solicitar ao SPSS que determine um conjunto de resíduos como novas variáveis no editor de dados. Essas variáveis-resíduos podem então ser examinadas para verificar a qualidade do ajuste do modelo aos dados observados. Para salvar os resíduos, clique em **Save...** na caixa de diálogo principal da regressão logística (Figura 6.2). O SPSS salva cada uma das variáveis selecionadas no editor de dados, mas elas podem ser listadas no visualizador de saídas utilizando o comando *Case Summaries* (Resumo de Casos) (veja a Seção 5.8.6) e selecionando as variáveis de interesse. A caixa de diálogo dos resíduos ilustrada na Figura 6.6 nos fornece várias opções; a maioria delas são as mesmas da regressão múltipla (veja a Seção 5.7.4). Dois tipos de resíduos exclusivos da regressão logística são as probabilidades previstas (*predicted probabilities*) e a pertinência prevista ao grupo (*predicted group memberships*). Essas probabilidades previstas são as probabilidades de Y ocorrer dado o valor de cada preditor para um determinado participante. Elas são derivadas da equação (6.4)

para um dado caso. A permanência prevista ao grupo é autoexplicativa no sentido de que ela prevê a qual das duas categorias de Y o participante tem a maior probabilidade de pertencer com base no modelo. A permanência ao grupo é baseada nas probabilidades previstas e explicarei esses valores mais detalhadamente quando formos interpretar os resíduos. Vale a pena selecionar todas as opções disponíveis ou pelo menos as mesmas opções que aparecem na Figura 6.6.

O exame dos resíduos depois de qualquer análise é extremamente importante. Uma das muitas vantagens dos pacotes computacionais em relação à análise manual é que eles agilizam a obtenção de diagnósticos residuais que de outra forma seriam difíceis de conseguir devido ao tempo gasto em cálculos e pela álgebra envolvida. No entanto, o avanço computacional ainda não atingiu o ponto ótimo na consideração dos resíduos. Infelizmente, cientistas sociais, treinados a considerar apenas valores de probabilidade abaixo de 0,05 desconsideram a análise dos resíduos! Falei sobre a importância de construir modelos precisos e, também, que executar uma regressão sem conferir se o modelo se ajusta aos dados é como comprar calças sem experimentá-las – eles podem parecer boas no cabide, mas quando você as veste, em casa, ficam apertadas. As calças podem cumprir seu objetivo (elas cobrem as pernas e esquentam), mas elas não têm um valor real (porque elas cortam a circulação do sangue nas suas pernas e em outras partes importantes). Da mesma forma, a regressão executa o seu trabalho a despeito dos dados – ela irá criar um modelo – mas o valor real do modelo será limitado.

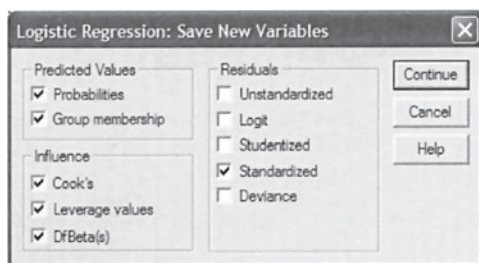


Figura 6.6 Caixa de diálogo para a determinação de resíduos na regressão.

Testar o grau de aderência com que o modelo se ajusta aos dados coletados é a essência das estatísticas de diagnóstico. Se o modelo adere bem aos dados, podemos ter uma maior segurança de que os coeficientes do modelo serão mais precisos e não serão influenciados por uns poucos pontos que se desviam. Na melhor das hipóteses, falhas no exame dos resíduos podem levá-lo a ignorar coeficientes pouco precisos no modelo. Na pior, esse modelo não acurado pode ser utilizado para aceitar hipóteses teóricas que são, de fato, falsas.

6.4.5 Opções adicionais ②

Existe uma caixa de diálogo final que oferece opções adicionais. Essa caixa é mostrada na Figura 6.7 e é acessada clicando em

Options... na caixa de diálogo principal da regressão logística. Em geral, as opções padrão são suficientes. Mencionei na Seção 6.4.2 que quando o método passo a passo é utilizado, existe um critério padrão para a seleção e a remoção de previsores no modelo. Esses critérios estão apresentados na caixa de diálogo *options* (opções) sob o título *Probability for Stepwise* (Probabilidade para o passo a passo). A probabilidade limite pode ser alterada, mas há necessidade, a menos que você tenha uma boa razão para querer um critério mais rígido para a seleção de variáveis. Outro valor padrão é a chegada ao modelo após um máximo de 20 iterações. Sempre que tentamos ajustar o melhor modelo possível aos nossos dados, e o número máximo de iterações pode ser imaginado como o número máximo de tentativas que o computador fará para encontrar esse melhor modelo. A menos que você tenha um modelo bastante complexo, 20 iterações serão mais do que suficientes. Vimos, no Capítulo 5, que as equações de regressão contêm uma constante que representa o intercepto com Y (isto é, o valor de Y quando os valores dos previsores são iguais a zero). Por padrão, o SPSS inclui a constante no modelo, mas é possível realizar a análise sem a constante, e isso tem o efeito de fazer o modelo passar pela origem (isto é, Y é zero quando X é zero). Dado que estamos interessados em produzir um modelo

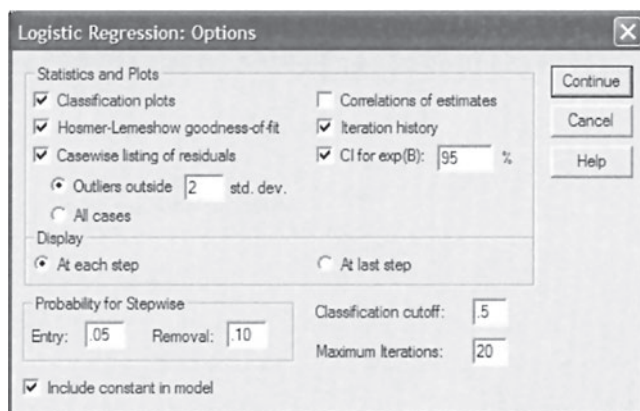


Figura 6.7 Caixa de diálogo para as opções da regressão logística.

com o máximo de aderência aos dados que coletamos, não há motivo para fazer uma análise sem a constante incluída.

Uma opção útil é o diagrama de classificação, que é um histograma dos valores reais e previstos da variável de saída. Esse diagrama serve para avaliar o ajuste do modelo aos dados observados. É também possível fazer uma lista de resíduos por caso (**casewise**) tanto para casos em que o resíduo padronizado é maior do que dois desvios padrões (esse valor pode ser alterado, mas o padrão é razoável) quanto para todos os casos. Recomendo um exame mais detalhado dos resíduos, mas essa opção pode ser útil para uma inspeção rápida. Você pode solicitar ao SPSS que apresente um intervalo de confiança (veja a Seção 1.6.2) para a estatística $\exp b$; o SPSS versão 12 apresenta um intervalo de confiança de 95% que é apropriado e, também, uma estatística útil para se ter. Mais importante ainda, você pode solicitar a estatística de aderência de *Hosmer-Lemeshow*, que pode ser utilizada para avaliar quão bem o modelo escolhido se ajusta aos dados. As opções restantes não são importantes: você pode escolher apresentar todas as estatísticas e gráficos a cada estágio de uma análise (o padrão) ou somente ao final, quando o modelo estiver determinado. Finalmente, você pode obter a matriz dos coeficientes das estimativas dos parâmetros em termos do modelo, coeficientes e valores de verossimilhança-log

a cada iteração do processo de estimação dos parâmetros – a praticidade disso é desperdiçada pela maioria das pessoas!

6.5 INTERPRETANDO A REGRESSÃO LOGÍSTICA ②

Quando você tiver selecionado todas as opções, clique em **OK** e observe a saída aparecer na janela do visualizador. Explicarei cada parte da saída mais adiante. Em versões anteriores do SPSS, a saída da regressão logística não era formatada (ela aparecia apenas como texto – veja a primeira edição deste livro se você estiver curioso!), mas atualmente a saída aparece formatada em belas tabelas!

6.5.1 O modelo inicial ②

A Saída do SPSS 6.1 nos informa sobre os códigos do parâmetro atribuídos à variável categórica previsora. Códigos indicadores foram escolhidos com duas categorias e, assim, o código é o mesmo dos valores no editor de dados. Se o código do desvio for escolhido, ele será -1 (**fb** Sim) e 1 (**fb** Não). Com o contraste simples, os códigos serão $-0,5$ (**fb** Não) e $0,5$ (**fb** Sim) se a última categoria for selecionada como referência. Os códigos do parâmetro são importantes para calcular a probabilidade da variável de saída ($P(Y)$), mas voltaremos a isso mais tarde.

Saída do SPSS 6.1

Dependent Variable Encoding (Codificação da variável dependente)

<i>Original Value</i> (Valor Original)	<i>Internal Value</i> (Valor Interno)
No (Não)	0
Yes (Sim)	1

Categorical Variable Codings (Códigos da variável categórica)

		<i>Frequency</i> (Frequência)	<i>Parameter</i> (Parâmetro) (1)
<i>False Belief Understanding</i> (Entendimento de falsas crenças)	Yes (Sim)	29	0.000
	No (Não)	41	1.000

Para a primeira análise, requisitamos o método passo a passo para frente (*forward stepwise*), portanto, o modelo inicial é determinado utilizando somente a constante na equação de regressão. A Saída 6.2 do SPSS nos informa sobre o modelo quando somente a constante é incluída (isto é, todas as variáveis predictoras são omitidas). Embora o SPSS não apresente esse valor, a verossimilhança-log do modelo base (veja a Seção 6.3.1) é 96,124 (acredite em mim por enquanto!). Isso representa o valor da aderência quando o modelo mais básico é ajustado aos dados. Quando inclui somente a constante, o computador cria um modelo que atribui todos os participantes a uma única categoria da variável de saída. Nesse exemplo, o SPSS pode decidir tanto prever que cada criança tem um entendimento das regras

de comportamento ou prever que cada criança não tem entendimento das regras de comportamento. Ele pode tomar essa decisão arbitrariamente, mas porque é crucial maximizar quão bem o modelo prevê os valores observados, o SPSS irá prever que todas as crianças pertencem à categoria onde a maioria dos dados estaria. Nesse exemplo, 39 crianças têm um entendimento das regras de comportamento e somente 31 não. Dessa forma, se o SPSS prevê que todas as crianças têm um entendimento das regras de comportamento, essa previsão estará correta 39 vezes em 70 (isto é, 56% aproximadamente). Contudo, se o SPSS prevê que cada criança não apresenta entendimento das regras de comportamento, a previsão estará correta 31 vezes em 70 (44% aproximadamente). Assim, das duas opções, é melhor prever que todas as

Saída do SPSS 6.2

Classificação Table^{a, b} (Tabela de classificação)

Observed (Observado)			Predicted (Previstos)		
			Display Rule Understanding (Entendimento das Regras de Comportamento)		Percentage Correct (Percentual Correto)
			No (Não)	Yes (Sim)	
Step 0 (Passo 0)	Display Rule Understanding (Entendimento das Regras de Comportamento)	No (Não)	0	31	0.0
	Overall Percentage (Estatísticas Gerais)	Yes (Sim)	0	39	100.0 55.7

a A constante está incluída no modelo.
b O valor de corte é 0,500.

Variables in the Equation (Variáveis na equação)

		B	S. E. (Erro Padrão)	Wald	df (gl)	Sig. (Sig.)	Exp(B)
Step 0 (Passo 0)	Constant (Constante)	0.230	0.241	0.910	1	0.340	1.258

Variables not in the Equation (Variáveis que não estão na equação)

			Score (Escore)	df (gl)	Sig. (Sig.)
Step 0 (Passo 0)	Variables (Variáveis)	AGE (Idade)	15.956	1	0.000
		FB(1)	24.617	1	0.000
		AGE by FB(1)	23.987	1	0.000
	Overall Statistics (Estatísticas Gerais)		26.257	3	0.000

crianças apresentam entendimento das regras de comportamento porque isso resulta em um número maior de previsões corretas. A saída do SPSS mostra uma tabela de contingência para o modelo em seu estado básico. O SPSS previu que todas as crianças têm entendimento das regras de comportamento, o que resulta numa acurácia de 0% para as crianças observadas como não tendo entendimento das regras de comportamento, e 100% para as que passaram no teste da tarefa das regras de comportamento. No todo, o modelo classifica corretamente 55,71% das crianças. A próxima parte da saída resume o modelo, e isso requer citar o valor da constante (b_0), que é igual a 0,23.

A tabela final da saída é denominada Variáveis que não estão na Equação (*Variables not in the Equation*). A última linha dessa tabela apresenta a estatística qui-quadrado dos resíduos como 26,257, que é significativa a $p < 0,0001$ (ela é chamada de estatística global – *Overall Statistics*). Essa estatística informa que os coeficientes para as variáveis que não estão no modelo são significativamente diferentes de zero; em outras palavras, que a adição de uma ou mais dessas variáveis ao modelo irá afetar significativamente o seu poder de previsão. Se a probabilidade para o qui-quadrado residual fosse maior do que 0,05, isso implicaria que nenhuma das variáveis excluídas do modelo contribuiria de forma significativa para o poder preditivo do modelo. Desse modo, a análise terminaria nesse estágio.

O restante dessa tabela lista cada um dos previsores com o valor da **estatística do escore eficiente de Roa** para cada uma (coluna denominada *Escore [Score]*). Em amostras grandes, quando a hipótese nula é verdadeira, a estatística escore é idêntica à estatística de Wald e à estatística da razão de verossimilhança. Ela é utilizada nesse estágio da análise porque é computacionalmente menos intensiva que a estatística de Wald e, assim, pode ser calculada em situações em que esta seria proibitiva. Como qualquer estatística teste, a estatística escore de Roa tem uma distribuição específica a partir da qual a sua significância pode ser determinada. Nesse exemplo, todas as variáveis

excluídas têm estatísticas escore significativas a $p < 0,001$ e, dessa forma, todas podem potencialmente contribuir para o modelo. Contudo, como mencionado na Seção 6.4.2, os cálculos passo a passo são relativos, portanto, a variável que será selecionada para inclusão é a que apresentar o maior valor para a estatística escore e que seja significativa ao nível de 0,05. Nesse exemplo, tal variável é a **fb** porque ela apresenta a estatística escore de valor mais alto. A próxima parte da saída lida com o modelo após os previsores terem sido adicionados.

6.5.2 Passo 1: Entendendo falsas crenças ③

No primeiro passo, o entendimento de falsas crenças (**fb**) é adicionado ao modelo como um preditor. Assim, uma criança é classificada como tendo um entendimento das regras de comportamento com base no sucesso ou não no teste da tarefa das falsas crenças. Agora, isso pode ser facilmente explicado se olharmos a tabulação cruzada para as variáveis **fb** e **display**.¹ O modelo utilizará o entendimento de falsas crenças para prever se uma criança compreende as regras de comportamento aplicando a tabela de tabulação cruzada mostrada na Tabela 6.1. Desse modo, o modelo prevê que todas as crianças que mostraram entender falsas crenças apresentarão compreensão das regras de comportamento. Existem 41 crianças com entendimento de falsas crenças, assim, o modelo prevê que 41 crianças têm entendimento das regras de comportamento (ele estará correto 33 vezes em 41 e incorreto 8 vezes em 41). Além disso, esse novo modelo prevê que todas as 29 crianças que não apresentaram entendimento de falsas crenças não apresentam entendimento das regras de comportamento (nessa situação, ele estará correto 23 vezes em 29 e errado 6 vezes em 29).

A Saída 6.3 do SPSS mostra um resumo das estatísticas sobre o novo modelo (que,

¹ A caixa de diálogo para produzir essa tabela pode ser obtida acionando os menus **Analyze⇒Descriptive Statistics⇒Crosstabs** (Analisar⇒Estatística Descritiva⇒Tabulação cruzada).

Table 6.1 Tabulação cruzada das regras de comportamento com o entendimento de falsas crenças

		False Belief Understanding (Fb) (Entendimento de Falsas Crenças)	
		No (Não)	Yes (Sim)
Display Rule (Regra de Comportamento)	No (Não)	23	8
Understanding (display) (Entendimento)	Yes (Sim)	6	33

Saída do SPSS 6.3

Omnibus Tests of Model Coeficientes
(Testes em etapas dos coeficientes do modelo)

		Chi-square (Qui-quadrado)	df (gl)	Sig. (Sig.)
Step (Passo 1)	Step (Passo)	26.083	1	0.000
	Block (Bloco)	26.083	1	0.000
	Model (Modelo)	26.083	1	0.000

Model Summary (Resumo do Modelo)

Step	-2 Log likelihood (-2 Verossimilhança-log)	Cox & Snell R Square (R quadrado de Cox e Snell)	Nagelkerke R Square (R quadrado de Nagelkerke.)
1	70.042	0.311	0.417

Classification Table^a (Tabela de Classificação)

			Predicted (Previstos)		
			Display Rule Understanding (Entendimento das Regras de Comportamento)		Percentage Correct (Percentual Correto)
			No (Não)	Yes (Sim)	
Step 1 (Passo 1)	Observed (Observado)	No (Não)	23	8	74.2
	Display Rule Understanding (Entendi- mento das Regras de Comportamento)	Yes (Sim)	6	33	84.6
	Overall Percentage (Estatísticas Gerais)				80.0

a O valor de corte é 0,500.

como já vimos, contém **fb**). A aderência global do novo modelo é determinada utilizando a estatística de verossimilhança-log (veja a Seção 6.3.1). No SPSS, em vez de relatar o valor da verossimilhança-log, o valor é multiplicado por -2 (e algumas vezes denominado $-2VL$ ($-2VL$)): essa multiplicação é feita porque $-2VL$ tem uma distribuição aproximadamente qui-quadrado e, dessa forma, é possível

comparar os seus valores contra aqueles que esperaríamos que fossem obtidos apenas por acaso. Lembre-se de que valores grandes da estatística VL indicam um modelo com uma má aderência.

Nesse estágio da análise, o valor de $-2x$ verossimilhança-log deve ser menor do que quando somente a constante foi incluída no modelo (porque valores baixos de $-2VL$ in-

dicam que o modelo está prevendo a variável de saída com maior precisão). Quando somente a constante foi incluída, $-2VL = 96,124$, mas com a adição da variável **fb**, esse valor foi reduzido para 70,042. Essa redução nos informa que o modelo agora é melhor para prever o entendimento das regras de comportamento do que antes da variável **fb** ter sido acrescentada. Você pode determinar quão bem um modelo prevê a variável de saída utilizando a estatística do modelo qui-quadrado (*model chi-square statistic*), que avalia a diferença entre o modelo atual e o modelo quando somente a constante está incluída. Vimos na Seção 6.3.1 que podemos determinar a significância da mudança em um modelo tomando a verossimilhança-log do novo modelo e subtraindo da verossimilhança-log do modelo base. A estatística do modelo qui-quadrado funciona com base nessa ideia e, dessa forma, é igual a $-2VL$ com a variável **fb** incluída menos o valor de $-2VL$ quando somente a constante está presente no modelo ($96,124 - 70,042 = 26,083$). Esse valor tem uma distribuição qui-quadrado, portanto, sua significância estatística pode ser facilmente calculada.² Nesse exemplo, o valor é significativo ao nível de 0,05, assim, podemos dizer que o modelo como um todo está prevendo o entendimento das regras de comportamento melhor do que quando tinha apenas a constante. A estatística do modelo qui-quadrado é análoga ao teste F para a soma dos quadrados da regressão linear (veja o Capítulo 5). O ideal seria ter um $-2VL$ não-significativo (indicando que a quantidade de dados não explicados é mínima) e uma estatística do modelo qui-quadrado alta (indicando que o modelo incluindo os previsores é significativamente melhor do que sem os previsores). Contudo, na realidade é possível que ambas as estatísticas sejam muito significativas.

Existe uma segunda estatística denominada estatística passo (*step*), que indica a melhoria no poder preditivo do modelo a partir do último

estágio. Nesse estágio existiu apenas um passo na análise, portanto, o valor da estatística de melhoria é a mesma do qui-quadrado do modelo. Contudo, em modelos mais complexos em que existem três ou quatro estágios, esta estatística fornece uma medida da melhoria do poder preditivo do modelo desde o último passo. Seu valor é igual ao $-2VL$ no passo atual menos o $-2VL$ no passo anterior. Se a estatística de melhoria é significativa, ela indica que o modelo agora prevê a saída significativamente melhor do que no último passo, e numa regressão para frente (*forward*), isso pode ser entendido como uma indicação da contribuição de um predictor para o poder preditivo do modelo. De forma semelhante, a estatística bloco (*block*) fornece a mudança no $-2VL$ desde o último bloco (para uso em análises hierárquicas ou por blocos).

Finalmente, a tabela de classificação no final da seção de saída indica quão bem o modelo prevê a inclusão em um grupo. Em virtude de o modelo estar utilizando o entendimento de falsas crenças para prever a variável de saída, essa tabela de classificação é a mesma que a Tabela 6.1. O modelo atual classifica corretamente 23 crianças que não apresentam compreensão das regras de comportamento, mas classifica errado outras 8 (isto é, ela classifica corretamente 74,19% dos casos). Para as crianças que têm compreensão das regras de comportamento, o modelo classifica corretamente 33 e erra em 6 casos (isto é, ele classifica 84,62% dos casos corretamente). A acurácia global de classificação é, dessa forma, a média ponderada desses dois valores (80%). Assim, quando somente a constante foi incluída o modelo classificou corretamente 56% das crianças, mas com a inclusão do predictor **fb**, esse valor subiu para 80%.

A próxima parte da saída (Saída do SPSS 6.4) é crucial porque ela nos informa as estimativas dos coeficientes dos previsores incluídos no modelo. Essa seção da saída nos fornece os coeficientes e estatísticas para as variáveis que foram incluídas no modelo nesse ponto (a saber, a constante e a **fb**). O valor *b* é o mesmo valor da regressão linear: eles são os valores que precisamos para substituir na equação (6.4) a fim

² Os graus de liberdade serão o número de parâmetros no novo modelo (o número de previsores mais 1; nesse caso, com um predictor, é 2) menos o número de parâmetros no modelo base (que é 1, a constante). Portanto, nesse caso, $gl = 2 - 1 = 1$.

Saída do SPSS 6.4

Variables in the Equation (Variáveis na equação)

		<i>B</i>	<i>S. E.</i> (Erro Padrão)	<i>Wald</i>	<i>df</i> (gl)	<i>Sig.</i> (Sig.)	<i>Exp(B)</i>	95,0% C.I for EXP(B) (I.C de 95% para EXP(B))	
								<i>Lower</i> (Inferior)	<i>Upper</i> (Superior)
<i>Step</i> (Passo) 1 ^ª	<i>FB(1)</i>	2.761	0.605	20.856	1	0.000	15.812	4.835	51.706
	<i>Constant</i> (Constante)	−1.344	0.458	8.592	1	0.003	0.261		

a Variable(s) entered on Step 1: FB (Variável(is) adicionada(s) no passo 1: FB).

de identificar a probabilidade de que um caso pertença a uma determinada categoria. Vimos na regressão linear que o valor de *b* representa a modificação na saída resultante da mudança de uma unidade na variável previsor. A interpretação desse coeficiente na regressão logística é muito semelhante no sentido de que ele representa a chance na *logit* da variável de saída associada à alteração de uma unidade na variável previsor. A *logit* da saída é simplesmente o logaritmo natural da chance de *Y* ocorrer.

Uma estatística importante é a estatística de Wald,³ que tem uma distribuição qui-quadrado e nos informa se o coeficiente *b* para o dado previsor difere de zero de forma significativa. Se isso acontece, podemos assumir que o previsor está contribuindo de modo significativo para a previsão da saída (*Y*). Vimos, na Seção 6.3.3, que a estatística de Wald deve ser utilizada com cautela porque, quando o coeficiente da regressão (*b*) é grande, o erro padrão tende a ficar inflacionado, resultando em uma subestimativa da estatística (veja Menard, 1995). Contudo, em relação a esses dados ela parece indicar que a compreensão das falsas crenças é um previsor significativo para o entendimento das regras de comportamento (observe que a significância da estatística de Wald é menor do que 0,05).

Na Seção 6.3.2, vimos que podíamos calcular um valor análogo de *R* utilizando a equação (6.7). Para esses dados, a estatística

de Wald e seus *gl* podem ser vistos na Saída do SPSS 6.4 (20,856 e 1, respectivamente) e o valor original do $-2VL$ é 96,12. Dessa forma, *R* pode ser calculado como:

$$R = \pm \sqrt{\left(\frac{20,856 - (2 \times 1)}{96,124} \right)} \quad (6.14)$$
$$= 0,4429$$

Na mesma seção, vimos que a medida de Hosmer e Lemeshow (R^2_V) é calculada dividindo o qui-quadrado do modelo pelo valor original de $-2VL$. Nesse exemplo, o qui-quadrado do modelo após todas as variáveis terem sido adicionadas é 26,083 e o valor original de $-2VL$ (antes que qualquer variável tenha sido adicionada) foi 96,124. Assim, $R^2_V = 26,083/96,124 = 0,271$, que é diferente do valor que seria obtido pela elevação ao quadrado do *R* determinado acima ($R^2 = 0,4429^2 = 0,196$). Na Saída 6.3 do SPSS, foram fornecidas duas outras medidas de R^2 , descritas na Seção 6.3.2. A primeira delas é a medida de Cox e Snell,⁴ que o SPSS fornece como 0,311,

⁴ Ela é calculada a partir da equação (6.9). Lembre-se de que essa equação utiliza a verossimilhança-log e que o SPSS apresenta como $-2 \times$ verossimilhança-log. VL(Nova) é, dessa forma, $70,042/-2 = -35,021$ e a VL(Base) = $96,124/-2 = -48,062$. O tamanho *n* da amostra é 70. Portanto:

$$R^2_{CS} = 1 - e^{\left\{ \frac{-2}{70} [-35,021 - (-48,062)] \right\}}$$
$$= 1 - e^{-0,3726}$$
$$= 1 - 0,6889$$
$$= 0,311$$

³ O SPSS na realidade apresenta a estatística de Wald ao quadrado, assim, para esses dados ela seria $(2,761/0,605)^2 = 20,8$, como o apresentado na tabela – ver equação (6.11).

e a segunda é o valor ajustado de Nagelkerke,⁵ que o SPSS calcula como sendo 0,417. Como você pode ver, existe uma diferença substancial entre os dois valores!

O último item que precisamos analisar é a $\exp b$ ($\text{Exp}(B)$) na saída do SPSS, que foi descrita na Seção 6.3.4. Para calcular a mudança nas chances que resultam da variação de uma unidade no previsor, neste exemplo, precisamos primeiro calcular as chances de uma criança entender as regras de comportamento dado que ela não tem uma compreensão de segunda ordem de falsas crenças. Depois calculamos as chances de uma criança ter entendimento das regras de comportamento dado que ela compreende falsas crenças. Finalmente, calculamos a mudança proporcional nessas duas chances.

Para calcular o primeiro conjunto de chances, precisamos utilizar a equação (6.12) a fim de calcular a probabilidade de uma criança compreender regras de comportamento visto que ela falhou na tarefa das falsas crenças. O código do parâmetro no início da saída indica que a uma criança que falhou na tarefa de falsas crenças foi atribuído o código 0; dessa forma, podemos utilizar esse valor no lugar de X . Estimamos o valor de b_1 como 2,7607 (veja Variáveis na Equação na Saída 6.4 do SPSS) e a constante, que pode ser obtida na mesma tabela, é -1,3437. Podemos calcular as chances como na equação (6.15):

$$\begin{aligned} P(\text{Evento } Y) &= \frac{1}{1 + e^{-(b_0 + b_1 X_1)}} \\ &= \frac{1}{1 + e^{-[-1,3437 + (2,7607 \times 0)]}} \\ &= 0,2069 \\ P(\text{Não Evento } Y) &= 1 - P(\text{Evento } Y) \quad (6.15) \\ &= 1 - 0,2069 \\ &= 0,7931 \\ \text{Chance} &= \frac{0,2069}{0,7931} \\ &= 0,2609 \end{aligned}$$

Em seguida, fazemos o mesmo cálculo após a variável previsor ter mudado em uma unidade. Nesse caso, em virtude do previsor ser dicotômico, precisamos calcular as chances de uma criança que passou na tarefa de entendimento das regras de comportamento, dado que ela tinha passado na tarefa das falsas crenças. Assim, o valor da variável falsas crenças, X , agora é 1 (em vez de 0). Os cálculos resultantes são mostrados na equação (6.16):

$$\begin{aligned} P(\text{Evento } Y) &= \frac{1}{1 + e^{-(b_0 + b_1 X_1)}} \\ &= \frac{1}{1 + e^{-[-1,3437 + (2,7607 \times 1)]}} \\ &= 0,8049 \\ P(\text{Não Evento } Y) &= 1 - P(\text{Evento } Y) \quad (6.16) \\ &= 1 - 0,8049 \\ &= 0,1951 \\ \text{Chance} &= \frac{0,8049}{0,1951} \\ &= 4,1256 \end{aligned}$$

⁵ Ela é calculada a partir da equação (6.10) – existe uma pequena diferença porque eu utilizei valores arredondados na terceira decimal:

$$\begin{aligned} R_N^2 &= \frac{0,311}{1 - e^{\left[\frac{2(-48,062)}{70}\right]}} \\ &= \frac{0,311}{1 - e^{-1,3732}} \\ &= \frac{0,311}{1 - 0,2533} \\ &= 0,416 \end{aligned}$$

Sabemos as chances antes e depois da alteração de uma unidade na variável previsor. Agora, basta calcular a mudança proporcional em chances dividindo a chance após a mudança de uma unidade no previsor pela chance antes da mudança:

Chance após a
mudança de uma
unidade no predictor

$$\Delta \text{Chance} = \frac{\text{unidade no predictor}}{\text{chance original}}$$

(6.17)

$$= \frac{4,1256}{0,2609}$$

$$= 15,8129$$

Você deve notar que o valor da mudança proporcional em chances é a mesma que o SPSS apresenta para o valor exp b (exceto por uma diferença no arredondamento). Podemos interpretar exp b em termos de uma mudança nas chances. Se o valor é maior do que 1, então ele indica que à medida que o predictor aumenta, aumentam as chances de uma saída ocorrer. Já um valor menor do que 1 indica que à medida que o predictor aumenta, as chances de uma saída ocorrer diminuem. Nesse exemplo, podemos afirmar que as chances de uma criança que compreende falsas crenças entender também as regras de comportamento é 15 vezes maior do que uma criança que não compreende falsas crenças.

Nas opções (veja a Seção 6.4.5), requisitamos um intervalo de confiança para o exp b e ele também pode ser visto na saída. Interpretamos um intervalo de confiança da seguinte maneira: se executássemos 100 experimentos e calculássemos os intervalos de confiança para o valor exp b, esses intervalos conteriam

o valor populacional de exp b em 95 ocasiões. Assim, nesse caso, poderíamos estar razoavelmente confiantes de que o valor populacional de exp b ficaria entre 4,84 e 51,71. Contudo, existiria uma probabilidade de 5% de que uma amostra forneceria um intervalo de confiança que não conteria o valor real de exp b.

A estatística teste para **fb** se ela for removida do modelo está na Saída 6.5 do SPSS. Em um momento anterior, falei que a regressão coloca as variáveis na equação e então testa se elas atendem ao critério de remoção. Bem, (*modelo se o termo é removido*) nos informa os efeitos da remoção. O importante a observar é a significância da razão da verossimilhança-log (VR log). O VR Log para esse modelo é altamente significativo ($p < 0,0001$), o que indica que se removermos **fb** do modelo, teremos um efeito significativo na sua habilidade preditiva – em outras palavras, não seria uma boa ideia removê-la!

Finalmente, somos informados sobre as variáveis que não estão atualmente no modelo. Primeiro, o qui-quadrado residual (denominado Estatística Global – *Overall Statistics* – na saída), que não é significativo, informa que nenhuma das variáveis restantes apresenta coeficientes significativamente diferentes de zero. Além disso, cada variável é listada com sua estatística *escore* e o valor da significância e para as duas variáveis os coeficientes não são significativamente diferentes de zero

Saída do SPSS 6.5

Model if Term Removed (Modelo se o Termo é Removido)

	Model LogLikelihood (Verossimilhança-Log do Modelo)	Change in -2logLikelihood. (Mudança no -2xVerossimilhança-Log)	df (gl)	Sig. of Change (Sig. da Mudança)
Step 1 FB (Passo 1)	-48.062	26.083	1	0.000

Variables not in the Equation (Variáveis que não estão na equação)

			Score (Escore)	df (gl)	Sig. (Sig.)
Step 1 (Passo 1)	Variables (Variáveis)	AGE (Idade)	2.313	1	0.128
		AGE by FB(1) (Idade por FB(1))	1.261	1	0.261
	Overall Statistics (Estatísticas Gerais)		2.521	2	0.283



Dica da Samanta Ferrinho

- A aderência global do modelo é mostrada pela estatística $-2x$ verossimilhança-log e a estatística qui-quadrado associada. Se a significância do qui-quadrado é menor do que 0,05, então o modelo se ajusta de forma significativa aos dados.
- Verifique a tabela denominada “Variáveis na equação” a fim de ver quais variáveis são significativas para prever a saída.
- Para cada variável no modelo, verifique a estatística de Wald e sua significância (que deve ser menor que 0,05). Mais importante, utilize o valor do $\text{Exp}(B)$ para a interpretação. Se o valor for maior do que 1, então quando o previsor aumenta, as chances de uma saída ocorrer aumentam. Já um valor menor do que 1 indica que quando o previsor aumenta, as chances de uma saída ocorrer diminuem. Para a interpretação acima mencionada ser confiável, o intervalo de confiança de $\text{Exp}(B)$ não deve ultrapassar 1!
- Confira a tabela denominada “Variáveis que não estão na equação” para ver quais variáveis não prevêm de forma significativa a saída.

zero). Essa situação representa um modelo que está prevendo corretamente os dados de saída. Se, contudo, existirem muitos pontos agrupados no centro do diagrama, isso mostra que em muitos casos o modelo está prevendo uma probabilidade de ocorrência do evento de 0,50. Em outras palavras, para esses casos não existe mais do que uma chance de 50:50 de que o dado será previsto corretamente; o modelo pode prever esses casos tão bem quanto o lançamento de uma moeda (jogar cara ou coroa) o faria. Ainda, um bom modelo irá assegurar que poucos casos sejam mal classificados; nesse exemplo existem dois Ns no lado direito do modelo e um Y no lado esquerdo do modelo. Esses são casos mal classificados; quanto menos desses existirem, melhor será o modelo.

6.5.3 Listando probabilidades previstas ②

É possível listar as probabilidades esperadas da variável de saída que ocorrem com base no modelo final. Na Seção 6.4.4, vimos que o SPSS pode salvar os resíduos e também as probabilidades previstas. O SPSS salva essas probabilidades e a previsão de pertinência a um grupo como variáveis no editor de dados com os nomes PRE_1 e PGR_1, respectivamente. Essas probabilidades po-

dem ser listadas utilizando a caixa de diálogo **Analyze⇒Reports⇒Case Summaries...** (Analisar⇒Relatar⇒Resumo de Casos...) (veja a Seção 5.8.6). A Saída do SPSS 6.7 mostra uma seleção das probabilidades previstas (porque o único previsor significativo é uma variável dicotômica, existirão somente dois valores de probabilidades previstas). Também vale a pena listar as variáveis predictoras como meio de esclarecer de onde as probabilidades previstas surgem.

Vimos, a partir do modelo, que o único previsor significativo da compreensão das regras de comportamento foi o entendimento das falsas crenças. Ele pode ter um valor tanto de 1 (passar na tarefa das falsas crenças) quanto de 0 (falhar na tarefa das falsas crenças). Se esses dois valores são colocados na equação (6.4) com os respectivos coeficientes de regressão, os dois valores das probabilidades são obtidos. De fato, calculamos esses valores como parte da equação (6.15) e da equação (6.16), e você deve ter observado que as probabilidades calculadas correspondem aos valores em PRE_1. Esses valores nos informam que quando uma criança não possui o entendimento de segunda ordem das falsas crenças ($fb = 0$, Não), existe uma probabilidade de 0,2069 de que ela passe na tarefa da compreensão das regras de comportamento – aproximadamente 21% de

Saída do SPSS 6.7

Case Summaries^a (Resumo dos casos)

	<i>Case Number</i> (Número do Caso)	<i>Age in years</i> (Idade em anos)	<i>False Belief Understanding</i> (Entendimento de Falsas Crenças)	<i>Display Rule Understanding</i> (Entendimento das Regras de Comportamento)	<i>Predicted probability</i> (Probabilidade prevista)	<i>Predicted group</i> (Grupo previsto)
1	1	24.00	Não	Não	0.20690	Não
2	5	36.00	Não	Não	0.20690	Não
3	9	34.00	Não	Sim	0.20690	Não
4	10	31.00	Não	Não	0.20690	Não
5	11	32.00	Não	Não	0.20690	Não
6	12	30.00	Sim	Sim	0.80488	Sim
7	20	26.00	Não	Não	0.20690	Não
8	21	29.00	Não	Não	0.20690	Não
9	29	45.00	Sim	Sim	0.80488	Sim
10	31	41.00	Não	Sim	0.20690	Não
11	32	32.00	Não	Não	0.20690	Não
12	43	56.00	Sim	Sim	0.80488	Sim
13	60	63.00	Não	Sim	0.20690	Não
14	66	79.00	Sim	Sim	0.80488	Sim
Total N		14	14	14	14	14

a Limitado aos primeiros 100 casos.

chance (1 em 5 crianças). Contudo, se a criança passa na tarefa das falsas crenças ($fb = 1$, Sim), existe uma probabilidade de 0,8049 de que ela irá passar na regra de comportamento – uma chance de 80,5% (4 em cada 5 crianças). Uma probabilidade de 0 indica que uma criança não tem chance de passar na tarefa do entendimento das regras de comportamento, e uma probabilidade de 1 indica que a criança irá definitivamente passar nessa regra. Portanto, os valores obtidos mostram a importância do entendimento das falsas crenças como pré-requisito para entendimento das regras de comportamento.

Assumindo que o modelo é preciso e que o entendimento das falsas crenças é significativo, podemos concluir que o entendimento das falsas crenças é o único preditor do entendimento das regras de comportamento. Além disso, a idade e a interação da idade com o entendimento das falsas crenças não prevêm de maneira alguma o entendimento das regras de comportamento. Como exercício, tente refazer essa análise utilizando o método da entrada forçada. Quais as diferenças nas conclusões?

Para ter certeza de que o modelo, de fato, é bom, é importante examinar os resíduos. Na

Seção 6.4.4, vimos como o SPSS salva os vários resíduos no editor de dados. Agora podemos listar esses resíduos utilizando a caixa de diálogo **Analyze⇒Reports⇒Case Summaries...** (Analisar⇒Relatar⇒Resumo de Casos...) e interpretá-los.

6.5.4 Interpretando os resíduos ②

Os principais objetivos de examinar os resíduos na regressão logística são (1) isolar pontos em que o modelo tem pouca aderência e (2) isolar pontos que exercem uma influência indevida no modelo. Para avaliar o primeiro, vamos examinar os resíduos, especialmente os estudentizados, os padronizados e as estatísticas de desvio. Todas essas estatísticas tem uma propriedade em comum: 95% dos casos em média, em uma amostra normalmente distribuída, devem ter valores que estão entre $\pm 1,96$, e 99% dos casos devem ter valores que estão entre $\pm 2,58$. Dessa forma, qualquer valor fora do intervalo ± 3 deve ser motivo de preocupação e qualquer um fora do intervalo $\pm 2,5$ deve ser examinado com cuidado. Para avaliar a influência de casos individuais, utilizamos as estatísticas de influência, como a distância de

Cook (que é interpretada da mesma forma que na regressão linear: como uma medida da mudança no coeficiente de regressão se um caso é retirado do modelo). O valor do DFBeta, que é a versão padronizada da estatística de Cook, também nos informa sobre a influência de certos casos – qualquer valor maior do que 1 indica casos potencialmente influentes. Ainda, a estatística de influência de valores chapéu, que devem estar entre 0 (o caso não tem influência alguma) a 1 (o caso exerce completa influência sobre o modelo), nos informa se certos casos estão exercendo influência indevida sobre o modelo. O valor esperado da influência (*leverage*) é definido como na regressão linear. Essas estatísticas foram explicadas em mais detalhes no Capítulo 5.

Se você requisitou as estatísticas residuais, o SPSS salva cada uma como uma nova coluna no editor de dados; a Tabela 6.2 lista cada nome das variáveis e o que elas representam. Existem comentários adicionais resumindo os valores esperados de algumas estatísticas.

A Saída do SPSS 6.8 mostra as estatísticas residuais básicas para esse exemplo (distância de Cook, influência, resíduos padronizados e valores DFBeta). Note que todos os casos

apresentam valores DFBetas menores do que 1 e a estatística de influência (LEV_1) próxima ao valor calculado esperado de 0,03. Também não existem valores anormalmente altos da distância de Cook (COO_1), o que significa que não existem casos influentes afetando o modelo. A distância de Cook é uma medida não-padronizada e, assim, não existe um valor absoluto que nos permita afirmar que um caso está tendo influência. Em vez disso, você deve olhar para valores da distância de Cook que são particularmente altos comparados com outros casos da amostra. Contudo, Stevens (1992) sugere que um valor maior do que 1 é problemático. Aproximadamente metade dos valores da influência são um tanto altos, mas como as outras estatísticas estão bem, isso não deve ser um motivo de preocupação.

Os resíduos padronizados têm valores menores do que $\pm 2,5$ e em geral apresentam valores menores do que ± 2 , portanto, parece não haver motivos para se preocupar.

Você poderá notar que esses resíduos não são usuais porque são baseados em um único preditor categórico. Esse é o motivo de não existir muita variabilidade nos valores dos resíduos. Também, se valores atípicos subs-

Table 6.2 Resumo das estatísticas residuais salvas pelo SPSS

Rótulo	Nome	Comentário
PRE_1	Valor previsto	
PGR_1	Grupo previsto	
COO_1	Distância de Cook	Deve ser menor do que 1
LEV_1	Influência (<i>Leverage</i>)	Deve estar entre 0 (sem influência) e 1 (completa influência). A influência esperada é $(k + 1)/N$, onde k é o número de preditores e N é o tamanho da amostra. Nesse caso, ela deve ser $2/70 = 0,03$.
RES_1	Resíduo	
LRE_1	Resíduo Logit	
SRE_1	Resíduo Estudentizado	5% podem estar além $\pm 1,96$ e 1% além $\pm 2,58$.
ZRE_1	Resíduo padronizado	Casos acima de 3 são motivos de alerta e casos próximos de 3 merecem uma investigação.
DEV_1	Desvios	
DFB0_1	DFBeta para a constante	Deve ser menor do que 1.
DEF1_1	DFBeta para o primeiro preditor (fb)	

Saída do SPSS 6.8

Case Summaries^a (Resumo dos casos)

	<i>Case Number</i> (Número do Caso)	<i>Analog of Cook's influence statistics</i> (Análogo da estatística de influência de Cook.)	<i>Leverage value</i> (Valor da influência)	<i>Normalized residual</i> (Resíduo normalizado)	<i>DFBeta for constant</i> (DFBeta para a constante)	<i>DFBeta for FB(1)</i> (DFBeta para a FB(1))
1	1	0.00932	0.03448	-0.51075	-0.04503	0.04503
2	2	0.00932	0.03448	-0.51075	-0.04503	0.04503
3	3	0.00932	0.03448	-0.51075	-0.04503	0.04503
4	4	0.00932	0.03448	-0.51075	-0.04503	0.04503
5	5	0.00932	0.03448	-0.51075	-0.04503	0.04503
6	6	0.13690	0.03448	1.95789	0.17262	-0.17262
7	7	0.00932	0.03448	-0.51075	-0.04503	0.04503
8	8	0.00932	0.03448	-0.51075	-0.04503	0.04503
9	9	0.13690	0.03448	1.95789	0.17262	-0.17262
10	10	0.00932	0.03448	-0.51075	-0.04503	0.04503
11	11	0.00932	0.03448	-0.51075	-0.04503	0.04503
12	12	0.00606	0.02439	0.49237	0.00000	0.03106
13	13	0.00932	0.03448	-0.51075	-0.04503	0.04503
14	14	0.10312	0.02439	-2.03101	0.00000	-0.12812
15	15	0.00932	0.03448	-0.51075	-0.04503	0.04503
16	16	0.00932	0.03448	-0.51075	-0.04503	0.04503
17	17	0.13690	0.03448	1.95789	0.17262	-0.17262
18	18	0.00932	0.03448	-0.51075	-0.04503	0.04503
19	19	0.00606	0.02439	0.49237	0.00000	0.03106
20	20	0.00932	0.03448	-0.51075	-0.04503	0.04503
21	21	0.00932	0.03448	-0.51075	-0.04503	0.04503
22	22	0.00606	0.02439	0.49237	0.00000	0.03106
23	23	0.00606	0.02439	0.49237	0.00000	0.03106
24	24	0.10312	0.02439	-2.03101	0.00000	-0.12812
25	25	0.00932	0.03448	-0.51075	-0.04503	0.04503
26	26	0.00932	0.03448	-0.51075	-0.04503	0.04503
27	27	0.00932	0.03448	-0.51075	-0.04503	0.04503
28	28	0.00932	0.03448	-0.51075	-0.04503	0.04503
29	29	0.00606	0.02439	0.49237	0.00000	0.03106
30	30	0.00932	0.03448	-0.51075	-0.04503	0.04503
31	31	0.13690	0.03448	1.95789	0.17262	-0.17262
32	32	0.00932	0.03448	-0.51075	-0.04503	0.04503
33	33	0.00606	0.02439	0.49237	0.00000	0.03106
34	34	0.00606	0.02439	0.49237	0.00000	0.03106
35	35	0.00606	0.02439	0.49237	0.00000	0.03106
36	36	0.00606	0.02439	0.49237	0.00000	0.03106
37	37	0.10312	0.02439	-2.03101	0.00000	-0.12812
38	38	0.00606	0.02439	0.49237	0.00000	0.03106
39	39	0.13690	0.03448	1.95789	0.17262	-0.17262
40	40	0.10312	0.02439	-2.03101	0.00000	-0.12812
41	41	0.00606	0.02439	0.49237	0.00000	0.03106
42	42	0.00606	0.02439	0.49237	0.00000	0.03106
43	43	0.00606	0.02439	0.49237	0.00000	0.03106
44	44	0.00606	0.02439	0.49237	0.00000	0.03106
45	45	0.00606	0.02439	0.49237	0.00000	0.03106
Total N		45	45	45	45	45

a Limitado aos primeiros 100 casos.

Case Summaries^a (Resumo dos casos)

	Case Number (Número do Caso)	Analog of Cook's influence statistics (Análogo da estatística de influência de Cook.)	Leverage value (Valor da influência)	Normalized residual (Resíduo normalizado)	DFBeta for constant (DFBeta para a constante)	DFBeta for FB(1) (DFBeta para a FB(1))
1	46	0.10312	0.02439	-2.03101	0.00000	-0.12812
2	47	0.00606	0.02439	0.49237	0.00000	0.03106
3	48	0.00932	0.03448	-0.51075	-0.04503	0.04503
4	49	0.00932	0.03448	-0.51075	-0.04503	0.04503
5	50	0.10312	0.02439	-2.03101	0.00000	-0.12812
6	51	0.00606	0.02439	0.49237	0.00000	0.03106
7	52	0.00606	0.02439	0.49237	0.00000	0.03106
8	53	0.00606	0.02439	0.49237	0.00000	0.03106
9	54	0.00606	0.02439	0.49237	0.00000	0.03106
10	55	0.00606	0.02439	0.49237	0.00000	0.03106
11	56	0.00606	0.02439	0.49237	0.00000	0.03106
12	57	0.10312	0.02439	-2.03101	0.00000	-0.12812
13	58	0.00606	0.02439	0.49237	0.00000	0.03106
14	59	0.00606	0.02439	0.49237	0.00000	0.03106
15	60	0.13690	0.03448	1.95789	0.17262	-0.17262
16	61	0.00606	0.02439	0.49237	0.00000	0.03106
17	62	0.00606	0.02439	0.49237	0.00000	0.03106
18	63	0.00606	0.02439	0.49237	0.00000	0.03106
19	64	0.00606	0.02439	0.49237	0.00000	0.03106
20	65	0.00606	0.02439	0.49237	0.00000	0.03106
21	66	0.00606	0.02439	0.49237	0.00000	0.03106
22	67	0.00606	0.02439	0.49237	0.00000	0.03106
23	68	0.10312	0.02439	-2.03101	0.00000	-0.12812
24	69	0.00606	0.02439	0.49237	0.00000	0.03106
25	70	0.00606	0.02439	0.49237	0.00000	0.03106
Total N		25	25	25	25	25

a Limitado aos primeiros 100 casos.

tanciais ou casos influentes tivessem sido isolados, você não deveria simplesmente eliminar esses casos para melhorar o ajuste do modelo. Eles deveriam ser examinados bem

de perto para tentar descobrir uma boa razão de por que eles são diferentes. Eles podem ser um erro de entrada dos dados ou uma situação especial: por exemplo, se a criança teve difi-

Dica da Samanta Ferrinho



Você precisa examinar casos que podem influenciar o modelo de regressão logística.

- Observe os resíduos padronizados: não mais do que 5% dos casos devem ter valores absolutos acima de 2 e não mais do que 1% devem ter valores absolutos acima de 2,5. Qualquer caso com um valor acima de 3 poderá ser um valor atípico.
- Procure os valores da distância de Cook no editor de dados: qualquer valor acima de 1 indica um caso que pode estar influenciando o modelo.
- Calcule a influência média (o número de previsores mais um, dividido pelo tamanho da amostra) e procure por valores maiores do que duas ou três vezes a influência média.
- Procure por valores absolutos do DFBeta maiores do que 1.

culdade de prestar atenção na tarefa das falsas crenças e você notou isso quando realizou o experimento. Em tal situação, você teria um bom motivo para excluir o caso e saber as razões para tê-lo feito.

6.5.5 Calculando o tamanho de efeito ②

Já vimos (Seção 6.3.2) que o SPSS produz um valor de r para cada preditor, com base na Estatística de Wald. Isso serve para a medida do tamanho de efeito, mas vale a pena ter em mente o que já foi dito sobre ela: ela não será acurada quando a estatística de Wald é imprecisa!

6.6 COMO RELATAR A REGRESSÃO LOGÍSTICA ②

Em virtude de a regressão logística ser pouco utilizada, é difícil encontrar qualquer conjunto de procedimentos sobre como relatá-la! Portanto, o que segue é apenas uma visão pessoal do que deve ser apresentado, usando como base as normas da APA (American Psychological Association) para relatar situações que envolvem a regressão múltipla (veja a Seção 5.9). Assim como na regressão múltipla, sou a favor de tabelar os resultados, a menos que ela seja um modelo muito simples. No mínimo, os valores beta, seus erros padrão e os valores da significância e algumas estatísticas gerais sobre o modelo (como o R^2 e as estatísticas de aderência) devem ser apresentados. Também recomendo relatar o valor de $\text{Exp}(B)$ e o intervalo de confiança associado. Apresentar a constante também é interessante, pois

desse modo os leitores do seu trabalho poderão reconstruir o modelo caso tenham necessidade de fazê-lo. Considere também relatar as variáveis que não foram preditores significativos, pois saber isso pode ser tão útil quanto saber quais as que foram significativas. Para o exemplo deste capítulo, poderíamos apresentar uma tabela semelhante à Tabela 6.3.

Espero que você consiga perceber de onde os valores surgiram com base no que já discutimos até aqui. Como na regressão múltipla, arredondei os valores na segunda decimal. Para o R^2 e os betas padronizados não existe um zero antes do ponto decimal (porque esses valores não podem ser maiores do que 1), mas para todos os demais valores menores do que 1 o zero está presente.* A significância da variável é representada por asteriscos com uma nota indicando o nível de significância utilizado.

6.7 OUTRO EXEMPLO ②

Esse exemplo foi inspirado por eventos ocorridos nas finais da Copa do Mundo de Futebol de 1998 (esse tipo de frustração não é facilmente esquecida!). Infelizmente (para mim, um inglês), fui sujeitoado a ver a Inglaterra ser eliminada do campeonato numa disputa de pênaltis. Seis anos depois, acabo de ver (ontem)

* N. de T.: Como já foi observado anteriormente, no Brasil não se costuma fazer essa diferença, portanto todos os valores apresentam o zero antes da vírgula. Lembre que no sistema inglês e americano o ponto decimal significa, de fato, a vírgula. Como a tabela é, em geral, feita pelo software, foi mantida a notação original.

Tabela 6.3 Como relatar a regressão logística

	Intervalo de Confiança de 95% para a $\text{exp } b$			
	$B(EP)$	Inferior	$\text{Exp } b$	Superior
<i>Incluído</i>				
Constante	-1,34* (0,46)	4,84	15,81	51,71
Entendimento de falsas crenças	2,76** (0,60)			

Nota: $R^2 = 0,27$ (Hosmer e Lemeshow), 0,31 (Cox e Snell), 0,42 (Nagelkerke). $\chi^2(1)$ do Modelo = 26,08, $p < 0,001$, * $p < 0,01$, ** $p < 0,001$



a Inglaterra ser eliminada do Campeonato Europeu em outra disputa de pênaltis. Consolome com o fato de que pelo menos eles têm consistência na sua inépcia.

Se eu fosse o técnico da Inglaterra, estaria interessado em descobrir que fatores podem prever se um jogador vai ou não perder um pênalti. Para os que detestam futebol, pensem nesse exemplo como fatores que podem prever o sucesso em um lance livre do basquete ou um saque no vôlei, um pênalti no hóquei ou no rúgbi⁶ ou um chute no futebol americano. A questão dessa pesquisa é perfeita para uma regressão logística porque a nossa variável de saída é dicotômica: um pênalti pode virar gol ou não. Imagine que pesquisas anteriores (Eriksson, Beckham e Vassell, 2004; Hoddle, Batty e Ince, 1998) apontaram dois fatores confiáveis que podem prever se um batedor irá converter um pênalti em gol ou não. O primeiro fator é se o jogador que irá cobrar o pênalti é ansioso (esse fator pode ser medido utilizando uma escala como o questionário sobre ansiedade da universidade estadual da Pensilvânia).* O segundo fator é o sucesso anterior do jogador em bater pênaltis (se o jogador acerta cobranças de penalidades com frequência). Sabemos que a ansiedade tem um efeito nocivo no desempenho de várias tarefas, portanto, é possível prever que o estado de ansiedade poderá ser responsável por alguma parte da variância não explicada na cobrança de pênaltis.

Esse exemplo é um caso clássico de construção sobre um modelo bem estabelecido, pois dois previsores já são conhecidos e nós queremos testar o efeito de um novo. Assim, 75 jogadores de futebol foram selecionados

ao acaso e antes de cobrar um pênalti em uma competição eles responderam a um questionário sobre a ansiedade, fornecendo uma medida do quanto eles se preocupavam com as coisas em geral. As taxas de sucesso na cobrança de penalidades foram obtidas de uma base de dados. Finalmente, foi anotado se o jogador converteu ou não o pênalti. Os dados podem ser encontrados no arquivo **penalty.sav**, que contém quatro variáveis, cada uma em uma coluna separada.

- **Scored** (convertido): Essa variável é a nossa saída e ela está codificada como 0 = errou e 1 = converteu.
- **Pswq** (preocupação): Essa variável é o primeiro predictor e fornece uma medida do grau de preocupação do jogador em geral.
- **Previous** (anterior): Essa variável é o percentual de cobranças convertidas por um determinado jogador ao longo de sua carreira. Ela representa o sucesso anterior na cobrança de pênaltis.
- **Anxious** (Ansiedade). Essa variável é o nosso terceiro predictor e não foi previamente utilizada para prever o sucesso numa cobrança de penalidades. É uma medida do estado de ansiedade antes de cobrar o pênalti.

6.7.1 Executando a análise: regressão por entrada em blocos ②

Para executar a análise, precisamos primeiro selecionar a caixa de diálogo da regressão logística (*logistic regression*) e utilizar o mouse para especificar o caminho **Analyze⇒Regression⇒Binary Logistic** (Analisar⇒Regressão⇒Logística Binária). Nesse exemplo, sabemos que dois previsores já foram testados e é recomendável entrar com eles no modelo na forma de um único bloco. Podemos adicionar o novo predictor em um segundo bloco (fazendo isso, estamos examinando um modelo antigo e depois adicionando uma nova variável a esse modelo para ver se existe alguma melhoria). Esse método é conhecido como entrada em blocos e a Figura 6.8 mostra como ele é especificado.

⁶ No entanto, esse exemplo não seria realista porque o nosso time de rúgbi, ao contrário do de futebol, tem Johnny Wilkinson, o melhor dos cobradores de pênaltis – palmas pelo seu magnífico pé esquerdo!

* N. de T.: PSWQ (*Penn State Worry Questionnaire*) no original.

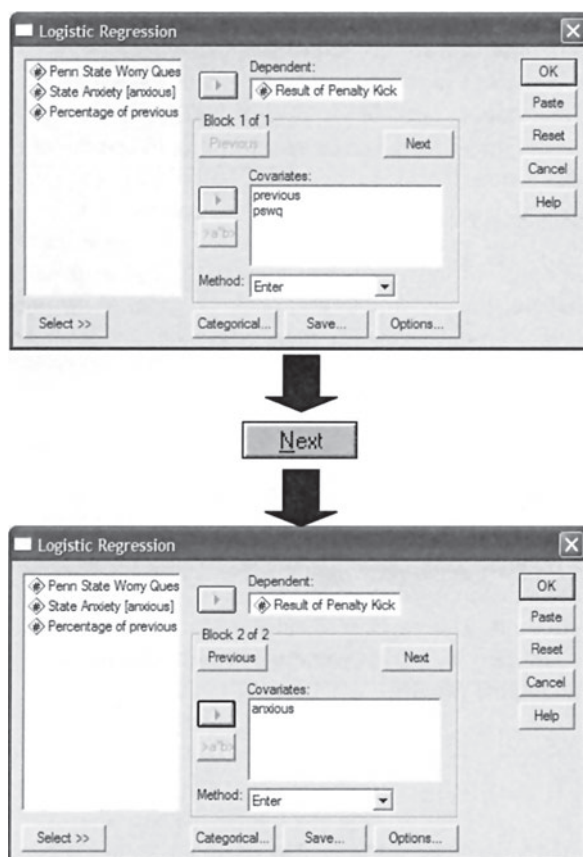


Figura 6.8 Método da entrada por blocos na regressão logística.

É fácil realizar uma regressão com entradas em blocos. Primeiro você deve utilizar o mouse para selecionar a variável **score**d (convertido) da lista de variáveis e depois transferi-la para o painel denominado *Dependent* (Dependente) clicando em . Segundo, você deve selecionar os dois previsores determinados anteriormente. Assim, selecione **pswq** e **previous** (anterior) da lista de variáveis e transfira-as para o painel denominado *Covariates* (Covariáveis) clicando em . Nosso primeiro bloco de variáveis está agora especificado. Para especificar o segundo bloco, clique em (próximo) a fim de limpar o painel das covariáveis, que deve estar rotulado agora como *Block 2 of 2* (Bloco 2 de 2). Selecione a variável **anxious** (ansiedade) da lista de variáveis e faça a transferência para o

painel denominado *Covariates* (Covariáveis) clicando em . Podemos, nesse estágio, selecionar algumas interações para incluir no modelo, mas a menos que exista alguma boa razão teórica para acreditar que os previsores têm interação, não existe a necessidade de fazê-lo. Verifique se o **Enter** está selecionado como método de regressão (esse é o método padrão e já deve estar selecionado).

Uma vez que as variáveis tenham sido especificadas, você deve selecionar as opções descritas nas Seções 6.4.4 e 6.4.5. Como nenhum dos previsores é categórico, não há necessidade de usar a opção **Categorical...** (Categórico). Quando você tiver selecionado as opções e os resíduos que você quer, retorne à caixa de diálogo principal da regressão logística e clique em .

6.7.2 Interpretando saídas ③

A saída da regressão logística será arranjada conforme os blocos que foram especificados. Em outras palavras, o SPSS irá produzir um modelo de regressão para as variáveis especificadas no bloco 1 e depois produzirá um segundo modelo que conterà as variáveis para os dois blocos. Primeiro, a Saída 6.9 do SPSS mostra os resultados do bloco 0: a saída nos informa que 75 casos foram aceitos e que a variável dependente foi codificada como 0 e

1 (porque essa variável foi codificada como 0 e 1 no editor de dados, esses códigos correspondem exatamente aos dados no SPSS). Somos então informados sobre as variáveis que estão na equação e as que não estão na equação. Nesse ponto, somente a constante está incluída no modelo, e sendo totalmente honesto, nenhuma dessas informações é interessante!

Os resultados para o bloco 1 são mostrados na Saída 6.10 do SPSS, e nesta análise nós forçamos o SPSS a adicionar as variáveis **previous**

Saída do SPSS 6.9

Dependent Variable Encoding (Codificação da Variável Dependente)

<i>Original Value</i> (Valor Original)	<i>Internal Value</i> (Valor Interno)
<i>Missed Penalty</i> (Pênalti Perdido)	0
<i>Scored Penalty</i> (Pênalti Convertido)	1

Block 0: Beginning Block (Bloco 0: Inicial)

Classificação Table^{a, b} (Tabela de classificação)

<i>Observed</i> (Observado)			<i>Predicted</i> (Previstos)		
			<i>Result of Penalty Kick</i> (Resultado da cobrança)		<i>Percentage Correct</i> (Percentual Correto)
			<i>Missed Penalty</i> (Pênalti Perdido)	<i>Scored Penalty</i> (Pênalti Convertido)	
<i>Step 0</i> (Passo 0)	<i>Result of Penalty Kick</i> (Resultado da cobrança)	<i>Missed Penalty</i> (Pênalti Perdido)	0	35	0.0
			0	40	100.0
	<i>Overall Percentage</i> (Percentagem Global)	<i>Scored Penalty</i> (Pênalti Convertido)			53.3

a A constante está incluída no modelo.

b O valor de corte é 0,500.

Variables in the Equation (Variáveis na equação)

	<i>B</i>	<i>S. E.</i> (Erro Padrão)	<i>Wald</i>	<i>df</i> (gl)	<i>Sig.</i> (Sig.)	<i>Exp(B)</i>	
<i>Step 0</i> (Passo 0)	<i>Constant</i> (Constante)	0.134	0.231	0.333	1	0.564	1.143

Variables not in the Equation (Variáveis fora da equação)

			<i>Score</i> (Escore)	<i>df</i> (gl)	<i>Sig.</i> (Sig.)
<i>Step 0</i> (Passo 0)	<i>Variables</i> (Variáveis)	<i>PREVIOUS</i> (Anterior)	34.109	1	0.000
	<i>Overall Statistics</i> (Estatísticas Globais)	<i>PSWQ</i>	34.193	1	0.000
			41.558	2	0.000

e **pswq** no modelo de regressão. Dessa forma, essa parte da saída fornece informações sobre o modelo após as variáveis **previous** e **pswq** terem sido adicionadas. Observe que o $-2VL$ é 48,66, ou seja, uma mudança para menos do 54,98 (que é o valor dado pelo qui-quadrado do modelo). Esse valor nos informa sobre o modelo como um todo enquanto que o bloco (*block*) nos informa quanto o modelo melhorou desde o último bloco. A mudança na quantidade explicada pelo modelo é significativa ($p < 0,0001$), assim, utilizar a experiência prévia e a preocupação como preditor aumenta significativamente nossa habilidade de prever o sucesso na cobrança de penalidades. Um pouco abaixo, a tabela de classificação nos mostra que 84% dos casos podem ser corretamente classificados utilizando **pswp** e **previous**.

No exemplo das regras de comportamento, o SPSS não produziu o teste de aderência de Hosmer e Lemeshow. A razão é que esse teste não pode ser calculado quando existe somente um preditor e esse preditor é uma variável categórica dicotômica! Contudo, para esse exemplo o teste pode ser calculado. A parte importante desse teste é a própria estatística teste (7,93) e sua significância (0,339). Essa estatística testa a hipótese de que os dados observados são significativamente diferentes dos valores previstos pelo modelo. Assim, nós queremos um valor não-significativo para esse teste (porque isso indicaria que o modelo não difere significativamente dos dados observados). Aqui temos um valor não-significativo, o que é um indicativo de que temos um modelo que irá prever bem os valores.

A parte da Saída 6.10 do SPSS denominada *Variables in the Equation* (Variáveis na Equação) nos informa quais os parâmetros do modelo quando **previous** e **pswq** são utilizados como preditores. Os valores de significância da estatística de Wald para cada preditor indicam que tanto **pswq** quanto **previous** podem prever significativamente o sucesso da cobrança de pênalti ($p < 0,01$). Os valores de $\exp b$ para **previous** indicam que se o percentual de pênaltis convertidos aumenta em uma unidade, as chances de converter um pênalti

também aumentam (porque $\exp b$ é maior do que 1). O intervalo de confiança para esse valor varia de 1,02 a 1,11, portanto, podemos ter certeza de que o valor populacional de $\exp b$ está em algum ponto entre esses dois valores. Além disso, em virtude dos dois valores serem maiores do que 1, podemos também estar confiantes de que o relacionamento entre a variável **previous** e o sucesso em converter um pênalti encontrado na amostra é verdadeiro para toda a população de jogadores de futebol. Os valores de $\exp b$ para a variável **pswq** indica que se o nível de preocupação aumenta por um ponto ao longo da escala de preocupação da Universidade Estadual da Pensilvânia, então as chances de converter um pênalti decresce (porque $\exp b$ é menor do que 1). O intervalo de confiança para esse valor varia de 0,68 a 0,93, assim, podemos estar bem confiantes de que o valor populacional de $\exp b$ está em algum ponto entre esses dois valores. Além disso, em virtude de os dois valores serem menores do que 1, podemos ter certeza de que o relacionamento entre a variável **pswq** e o sucesso em bater um pênalti encontrado nessa amostra pode ser atribuído para toda a população de jogadores de futebol. Se tivéssemos encontrado que o intervalo de confiança variasse de um valor menor do que 1 a um maior do que 1, isso limitaria a generalização de nossos achados porque o valor de $\exp b$ na população poderia indicar tanto uma relação positiva ($\exp(B) > 1$) quanto uma negativa ($\exp(B) < 1$).

Observando o diagrama de classificação (Saída do SPSS 6.11), percebemos que muitos casos estão agrupados nos extremos do diagrama e poucos casos estão no meio do diagrama. Isso reitera o que já sabemos: o modelo classifica corretamente a maioria dos casos. Podemos, nesse ponto, calcular o R^2 (veja a Seção 6.3.2) dividindo o qui-quadrado do modelo pelo valor original de $-2VL$. O resultado é mostrado na equação (6.18) e significa que o modelo pode explicar 53% da variância do sucesso na conversão de penalidades (assim, aproximadamente metade do que faz um pênalti ser convertido é ainda desconhecido):

Saída do SPSS 6.10

Block 1: Method = Enter (Bloco 1: Método = Enter)

Omnibus Tests of Model Coeficientes (Testes em etapas dos coeficientes do modelo)

		Chi-square (Qui-quadrado)	df (gl)	Sig. (Sig.)
Step 1 (Passo 1)	Step (Passo)	54.977	2	0.000
	Block (Bloco)	54.977	2	0.000
	Model (Modelo)	54.977	2	0.000

Model Summary (Resumo do Modelo)

Step (Passo)	-2 Log likelihood (-2 Verossimilhança-log)	Cox & Snell R Square (R quadrado de Cox e Snell)	Nagelkerke R Square (R quadrado de Nagelkerke.)
1	48.662	0.520	0.694

Hosmer and Lemeshow Test (Teste de Hosmer e Lemeshow)

Step (Passo)	Chi-square (Qui- quadrado)	df (gl)	Sig. (Sig.)
1	7.931	7	0.339

Contingency Table for Hosmer and Lemeshow Test
(Tabela de contingência para o teste de Hosmer e Lemeshow)

		Result of Penalty Kick = Missed Penalty (Resultado da cobrança = Pênalti perdido)		Result of Penalty Kick = Scored Penalty (Resultado da cobrança = Pênalti convertido)		Total
		Observed (Observado)	Expected (Esperado)	Observed (Observado)	Expected (Esperado)	
Step 1 (Passo 1)	1	8	7.904	0	0.096	8
	2	8	7.779	0	0.221	8
	3	8	6.705	0	1.295	8
	4	4	5.438	4	2.562	8
	5	2	3.945	6	4.055	8
	6	2	1.820	6	6.180	8
	7	2	1.004	6	6.996	8
	8	1	0.298	7	7.702	8
	9	0	0.108	11	10.892	11

a O valor de corte é 0,500.

Classificação Table^a (Tabela de classificação)

Observed (Observado)			Predicted (Previstos)		
			Result of Penalty Kick (Resultado da cobrança)		Percentage Correct (Percentual Correto)
			Missed Penalty (Pênalti Perdido)	Scored Penalty (Pênalti Convertido)	
Step 1 (Passo 1)	Result of Penalty Kick (Resultado da cobrança)	Missed Penalty (Pênalti Perdido)	30	5	85.7
		Scored Penalty (Pênalti Convertido)	7	33	82.5
	Overall Percentage (Percentagem Global)				84.0

a O valor de corte é 0,500.

(Continua)

$$R^2 = \frac{\text{qui-quadrado do modelo}}{\text{original} - 2\text{VL}}$$
$$= \frac{54,977}{103,6358}$$
$$= 0,53$$

(6.18)

A Saída 6.12 do SPSS mostra o que acontece ao modelo quando nosso novo preditor é adicionado (**anxious**). Essa parte da saída descreve o bloco 2, que é o modelo descrito no bloco 1 com o novo preditor adicionado. Assim, começamos com o modelo que tínhamos no bloco 1 e depois adicionamos o preditor **anxious** (ansioso) a ele. O efeito de adicionar o novo preditor ao modelo é reduzir o -2VL para 47,416 (uma redução de 1,246 a partir do modelo no bloco 1 como mostrado pelo qui-quadrado do modelo e a estatística de bloco). Essa melhoria não é significativa, ou seja, o acréscimo da variável ansioso (**anxious**) ao modelo não melhorou significativamente a habilidade de prever se um pênalti irá virar ou não um gol. A tabela de classificação nos informa que o modelo está agora classi-

ficando corretamente em 85,33% dos casos. Lembre que no bloco 1,84% foi classificado corretamente, dessa forma, 1,33% a mais dos casos são agora classificados (o que não é grande coisa – examinando a tabela podemos ver que apenas um caso novo foi corretamente classificado).

A seção denominada “Variáveis na Equação” contém agora três preditores e algo interessante aconteceu: **pswq** é ainda um preditor significativo do sucesso em converter penalidades, mas a variável **previous** não é mais significativa para prever o sucesso de um jogador fazer um gol. Além disso, a ansiedade parece não ter uma contribuição significativa para a previsão do sucesso em converter pênaltis.

Por que a experiência anterior não prevê mais o sucesso na marcação do gol nem a ansiedade se a habilidade do modelo de prever o sucesso na conversão de pênaltis aumentou ligeiramente?

O diagrama de classificação (Saída 6.13 do SPSS) é semelhante ao anterior e a contribuição do **pswq** para a previsão do sucesso em marcar o gol praticamente não está alterada. O

Saída do SPSS 6.12

Block 2: Method = Enter (Bloco 2: Método = Enter)

Omnibus Tests of Model Coefficients (Testes em etapas dos coeficientes do modelo)

		Chi-square (Qui-quadrado)	Df (gl)	Sig. (Sig.)
Step 1 (Passo 1)	Step (Passo)	1.246	1	0.264
	Block (Bloco)	1.246	1	0.264
	Model (Modelo)	56.223	3	0.000

Model Summary (Resumo do Modelo)

Step (Passo)	-2 Log likelihood (-2 Verossimilhança-log)	Cox & Snell R Square (R quadrado de Cox e Snell)	Nagelkerke R Square (R quadrado de Nagelkerke.)
1	47.416	0.527	0.704

Hosmer and Lemeshow Test (Teste de Hosmer e Lemeshow)

Step (Passo)	Chi-square (Qui-quadrado)	df (gl)	Sig. (Sig.)
1	9.937	7	0.192

(Continua)

Contingency Table for Hosmer and Lemeshow Test (Tabela de contingência para o teste de Hosmer e Lemeshow)

		<i>Result of Penalty Kick = Missed Penalty</i> (Resultado da cobrança = Pênalti perdido)		<i>Result of Penalty Kick = Scored Penalty</i> (Resultado da cobrança = Pênalti convertido)		<i>Total</i>
		<i>Observed</i> (Observado)	<i>Expected</i> (Esperado)	<i>Observed</i> (Observado)	<i>Expected</i> (Esperado)	
<i>Step 1</i> (Passo 1)	1	8	7.926	0	0.074	8
	2	8	7.769	0	0.231	8
	3	9	7.649	0	1.351	8
	4	4	5.425	4	2.575	8
	5	1	3.210	7	4.790	8
	6	4	1.684	4	6.316	8
	7	1	1.049	7	6.951	8
	8	0	0.222	8	7.778	8
	9	0	0.067	10	9.933	10

a O valor de corte é 0,500.

Classificação Table^a (Tabela de classificação)

			<i>Predicted</i> (Previstos)		
			<i>Result of Penalty Kick</i> (Resultado da cobrança)		<i>Percentage Correct</i> (Percentual correto)
<i>Observed</i> (Observado)			<i>Missed Penalty</i> (Pênalti perdido)	<i>Scored Penalty</i> (Pênalti convertido)	
<i>Step 0</i> (Passo 0)	<i>Result of Penalty Kick</i> (Resultado da cobrança)	<i>Missed Penalty</i> (Pênalti perdido)	30	5	85.7
	<i>Overall Percentage</i> (Per- centagem Global)	<i>Scored Penalty</i> (Pênalti convertido)	6	34	85.0
					85.3

a O valor de corte é 0,500.

Variables in the Equation (Variáveis na equação)

		<i>B</i>	<i>S. E.</i> (Erro Padrão)	<i>Wald</i>	<i>df</i> (gl)	<i>Sig.</i> (Sig.)	<i>Exp(B)</i>	<i>95,0% C.I. for EXP(B)</i> (I.C. de 95% para EXP(B))	
								<i>Lower</i> (Inferior)	<i>Upper</i> (Superior)
<i>Step 1</i> (Passo 1)	<i>PREVIOUS</i>	0.203	0.129	2.454	1	0.117	1.225		1.578
	<i>PSWQ</i>	−0.251	0.084	8.954	1	0.003	0.778	0.950	0.917
	<i>ANXIOUS</i>	0.276	0.253	1.193	1	1.318	3.598	0.660	2.162
	<i>Constant</i> (Constante)	−11.493	11.802	0.948	1	0.000		0.803	

a Variável(is) adicionada(s) no passo 1: ANSIOSO.

que mudou foi a contribuição da experiência anterior. Se examinarmos os valores de *exp b* tanto para a variável **previous** quanto para a **anxious**, fica claro que elas têm um relacionamento positivo potencial para o sucesso em converter pênalti (isto é, se elas aumentam em uma unidade, as chances de conversão melhoram). Contudo, os intervalos de confiança

para esses valores cruzam 1, o que indica que a direção desse relacionamento pode ser instável na população como um todo (isto é, o valor de *exp b* na nossa amostra pode ser totalmente diferente do valor se toda a população for utilizada).

Você pode ficar tentado a utilizar esse modelo final para afirmar que embora a preo-

modelo de regressão. A regressão logística é igualmente suscetível ao efeito da colinearidade, portanto, é essencial testar a colinearidade numa análise de regressão logística. Infelizmente, o SPSS não tem uma opção para fazer um diagnóstico de colinearidade na regressão logística (o que pode criar a ilusão de que não é necessário testá-la!). Contudo, você pode obter estatísticas tais como as de tolerância e FIV simplesmente executando uma análise de regressão linear utilizando a mesma variável de saída e os mesmos previsores. Assim, para o exemplo atual, acesse a caixa de diálogo para a regressão linear (*regression linear*) utilizando o mouse para especificar **Analyze**⇒**Regression**⇒**Linear** (Analisar⇒Regressão⇒Linear). A caixa de diálogo completa é mostrada na Figura 6.9. É desnecessário especificar um monte de opções (estamos utilizando a técnica apenas para testar a colinearidade), mas é essencial que você clique em **Statistics...** (Estatísticas) e depois selecione *Collinearity diagnostics* (diagnósticos de colinearidade) na caixa de diálogo. Uma vez que você selecionou ☒ *Collinearity diagnostics*, desmarque todas as opções padrão, clique em **Continue** (continue) para voltar à caixa de diálogo da regressão linear e clique em **OK** para executar a análise.

Os resultados da análise de regressão linear são mostrados na Saída 6.14 do SPSS. Da primeira tabela podemos ver que os valores da

tolerância são 0,014 para a variável **previous** e **anxious** e 0,575 para a **pswq**. No Capítulo 5, vimos vários critérios para determinar a colinearidade. Menard (1995) sugere que um valor da tolerância menor do que 0,1 provavelmente indica um problema sério de colinearidade. Myers (1990) também sugere que um valor FIV maior do que 10 é motivo de preocupação; nesses dados os valores estão acima de 70 tanto para **anxious** quanto para **previous**. Parece, a partir desses valores, que existe um problema de colinearidade entre as variáveis previsoras. Podemos investigar esse problema mais a fundo examinando os diagnósticos de colinearidade.

A Saída do SPSS 6.14 também mostra uma tabela denominada Diagnósticos de Colinearidade (*Collinearity Diagnostics*). Nessa tabela, são fornecidos os autovalores da matriz produto cruzado, o índice de condição e as proporções das variâncias de cada predictor. Se qualquer um dos autovalores nessa tabela é muito maior do que os demais, a matriz produto cruzado é mal condicionada, o que significa que as soluções dos parâmetros de regressão podem ser bastante afetadas por pequenas mudanças nos previsores ou na saída. Simplificando, esses valores nos dão uma idéia da precisão do nosso modelo: se os autovalores são similares, o modelo derivado tem pouca probabilidade de mudar por pequenas alterações nas variáveis mensuradas. O *índi-*

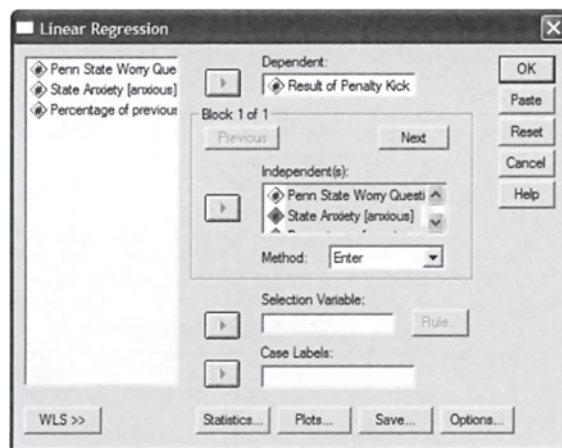


Figura 6.9 Caixa de diálogo para a regressão linear para dados de pênaltis.

Saída 6.14 do SPSS Diagnósticos de Colinearidade para os dados dos pênaltis
Coeficients^a (Coeficientes)

Model (Modelo)		Collinearity Statistics (Estatísticas de Colinearidade)	
		Tolerance (Tolerância)	VIF
1	State Anxiety (Estado da Ansiedade)	0.14	71.764
	Percentage of previous penalties scored (Percentual de pênaltis anteriores convertidos)	0.14	70.749
	Penn State Worry Questionnaire (Questionário de ansiedade da Penn State)	0.575	1.741

a. Variável dependente: Resultado do Pênalti Batido.

Collinearity Diagnostics^a (Diagnósticos de Colinearidade)

Model (Modelo)	Dimension (Dimensão)	Eigenvalue (Autovalor)	Condition Index (Índice de Condição)	Variance Proportions (Proporções da Variância)			
				(Constant) (Constante)	State Anxiety (Estado da Ansiedade)	Percentage of previous penalties scored (Percentual de pênaltis anteriores convertidos)	Penn State Worry Questionnaire (Questionário de ansiedade da Penn State)
1	1	3.434	1.000	0.00	0.00	0.00	0.01
	2	0.492	2.641	0.00	0.00	0.00	0.04
	3	7.274E-02	6.871	0.00	0.01	0.00	0.95
	4	5.195E-04	81.303	1.00	0.99	0.99	0.00

a Variável dependente: Resultado do Pênalti Batido.

ce de condição é outra forma de expressar esses autovalores, e é determinado pela raiz quadrada da razão do maior autovalor pelo autovalor de interesse (assim, para a dimensão com o maior autovalor, o índice de condição será sempre 1). Para esses dados, a dimensão final tem um índice de condição de 81,3, que é enorme se comparado com as demais dimensões. Embora não existam regras sobre o quão grande um índice de condição deva ser para indicar problemas de colinearidade, esse caso mostra claramente que o problema existe.

A etapa final da observação dessa tabela é analisar a proporção das variâncias. A variância de cada coeficiente da regressão pode ser dividida entre os autovalores, e as proporções das variâncias nos informam qual percentual da variância de cada coeficiente da regressão dos previsores é atribuído a cada autovalor. Essas proporções podem ser pensadas como percentuais para facilitar a compreensão. Assim, por exemplo, para o predictor **pswq**, 95%

da variância do coeficiente da regressão está associada com o autovalor número 3, 4% está associada com o autovalor número 2 e 1% é associada ao autovalor número 1. Em termos de colinearidade, estamos procurando por previsores que apresentam altas proporções no mesmo *pequeno* autovalor, porque isso irá indicar que a variância dos seus coeficientes de regressão são dependentes. Desse modo, estamos interessados principalmente nas poucas linhas no final da tabela (que representam pequenos autovalores). Nesse exemplo, 99% da variância nos coeficientes de regressão tanto do predictor **anxiety** quanto do **previous** estão associadas ao autovalor número 4 (o menor de todos os autovalores), o qual indica, claramente, uma dependência entre essas variáveis.

O resultado dessa análise é bem claro: existe uma colinearidade entre o estado da ansiedade e a experiência prévia em bater pênaltis e essa dependência resulta em um modelo viciado. Para ilustrar de onde essa co-

linearidade surge, a Saída 6.15 do SPSS mostra o resultado da correlação de Pearson entre todas as variáveis nessa análise de regressão (você pode executar por si mesmo esse tipo de análise). A partir dessa saída, podemos ver que os previsores **anxious** e **previous** estão altamente correlacionados de forma negativa ($r = -0,99$); de fato, eles estão quase perfeitamente correlacionados. Tanto **previous** quanto **anxious** se correlacionam com o sucesso na conversão dos pênaltis⁷, mas em virtude deles estarem altamente correlacionados entre si,

⁷ Alguns podem questionar se é legítimo calcular a correlação de Pearson com uma variável dicotômica tal como o sucesso na conversão de um pênalti; contudo, isso é simplesmente fazer uma correlação ponto-bisserial, descrita no Capítulo 4.

não fica claro qual prevê o sucesso em bater um pênalti na regressão.

Essa discussão suscita a questão do que fazer quando identificamos uma colinearidade. Bem, não há muito que fazer. Uma solução lógica é retirar uma das variáveis (por exemplo, você pode ficar com o modelo do bloco 1, que ignorou o estado da ansiedade). O problema dessa opção é óbvio: não existe uma maneira de saber qual variável retirar. A conclusão teórica resultante é, dessa forma, sem sentido porque, estatisticamente falando, qualquer uma das variáveis envolvidas na colinearidade pode ser omitida. Resumindo, estatisticamente, não há problema em omitir uma variável em relação à outra. Igualmente, se um preditor é removido, Bowerman e

Saída 6.15 do SPSS

Correlations (Correlações)

		Result of Penalty Kick (Resultado de Cobrança de Pênalti)	State Anxiety (Estado da Ansiedade)	Percentage of previous penalties scored (Percentual de pênaltis anteriores convertidos)	Penn State Worry Questionnaire (Questionário de ansiedade da Penn State)
Result of Penalty Kick (Resultado de Cobrança de Pênalti)	Pearson Correlation (Correlação de Pearson) Sig. (2-tailed) (Sig.) (Bilateral) N	1.000 . 75	-0.668** 0.000 75	0.674** 0.000 75	-0.675** 0.000 75
State Anxiety (Estado da Ansiedade)	Pearson Correlation (Correlação de Pearson) Sig. (2-tailed) (Sig.) (Bilateral) N	-0.668** 0.000 75	1.000 . 75	-0.993** 0.000 75	0.652** 0.000 75
Percentage of previous penalties scored (Percentual de pênaltis anteriores convertidos)	Pearson Correlation (Correlação de Pearson) Sig. (2-tailed) (Sig.) (Bilateral) N	0.674** 0.000 75	-0.993** 0.000 75	1.000 . 75	-0.644** 0.000 75
Penn State Worry Questionnaire (Questionário de ansiedade da Penn State)	Pearson Correlation (Correlação de Pearson) Sig. (2-tailed) (Sig.) (Bilateral) N	-0.675** 0.000 75	0.652** 0.000 75	-0.644** 0.000 75	1.000 . 75

** A correlação é significativa ao nível de 0,01 (bilateral).

O’Connell (1990) afirmam que outro previsor igualmente importante que não tenha tal colinearidade forte deve substituí-lo. Eles também sugerem que se colem mais dados para ver se a multicolinearidade pode ser reduzida. Outra possibilidade quando existem vários previsores envolvidos na multicolinearidade é executar uma análise de fatores nesses previsores e utilizar os escores dos fatores resultantes como um previsor (ver o Capítulo 15). A opção mais segura (embora insatisfatória) é reconhecer a instabilidade do modelo. Assim, se você está relatando uma análise de quais fatores prevêm o sucesso na cobrança de um pênalti, podemos reconhecer que a experiência prévia prevê de forma significativa o sucesso no primeiro modelo, mas expor que essa experiência poderá afetar a cobrança do pênalti pelo aumento do estado de ansiedade. Essa afirmação será altamente especulativa porque a correlação entre as variáveis **anxious** e **previous** nada indica sobre a direção da causalidade, mas ela valida a inexplicável conexão entre os dois previsores. Estou certo de que muitos de vocês podem julgar a falta de soluções para a colinearidade insatisfatória – infelizmente, a estatística é frustrante às vezes!

6.9 O QUE PODE DAR ERRADO ④

O SPSS resolve problemas de regressão logística por um procedimento iterativo; ele supõe uma solução provável, testa para ver quão bem essa solução adere aos dados e então repete o processo para tentar melhorar o ajuste. Ele pára quando a suposição feita estiver perto o suficiente da verdadeira solução. Algumas vezes, em vez de ir rapidamente para a solução correta, você irá notar que nada acontece: o SPSS começa a se mover muito devagar. Se você não consegue encontrar uma solução correta, às vezes ele desiste oferecendo a você (sem qualquer desculpa) um resultado completamente errado. Normalmente isso é revelado por um erro padrão absurdamente grande. Duas situações podem provocar esse tipo de ocorrência.

6.9.1 Informação incompleta dos previsores ④

Imagine que você está tentando prever a ocorrência de câncer de pulmão a partir do hábito de fumar (vício que, acredita-se, aumenta o risco de câncer) e comer ou não tomates (que, acredita-se, reduz o risco de câncer). Você coleta dados de pessoas fumantes e não fumantes e de pessoas que comem e que não comem tomates; contudo, isso não é suficiente, a menos que você colete dados de todas as combinações de fumar e comer tomates. Imagine que você conseguiu os seguintes dados:

Você fuma?	Você come tomates?	Você tem câncer?
Sim	Não	Sim
Sim	Sim	Sim
Não	Não	Sim
Não	Sim	?????

Observar apenas as três primeiras possibilidades não o prepara para a saída da quarta. Não é possível saber se a última pessoa tem ou não câncer com base nos dados coletados. Dessa forma, o SPSS terá problemas a menos que você tenha coletado dados de todas as combinações das suas variáveis. Isso deve ser verificado antes de executar a análise utilizando uma tabulação cruzada; descrevo como fazer isso no Capítulo 16.

Isso se aplica não apenas para as variáveis categóricas, mas também para as contínuas. Suponha que você quer investigar fatores relacionados à felicidade. Isso poderá incluir idade, gênero, orientação sexual, crenças religiosas, níveis de ansiedade e mesmo se a pessoa é destra. Você entrevista 1000 pessoas, registra as características e se elas são ou não felizes (“sim” ou “não”). Embora a amostra de 1000 pareça bastante grande, é provável que ela inclua uma mulher budista, lésbica, canhota e altamente ansiosa de 80 anos? Se você achasse tal pessoa e ela fosse feliz, você poderia concluir que qualquer uma com as mesmas características é feliz? Seria, obviamente, necessário ter muito mais pessoas com essas

características para confirmar que essa combinação de valores proporciona felicidade.

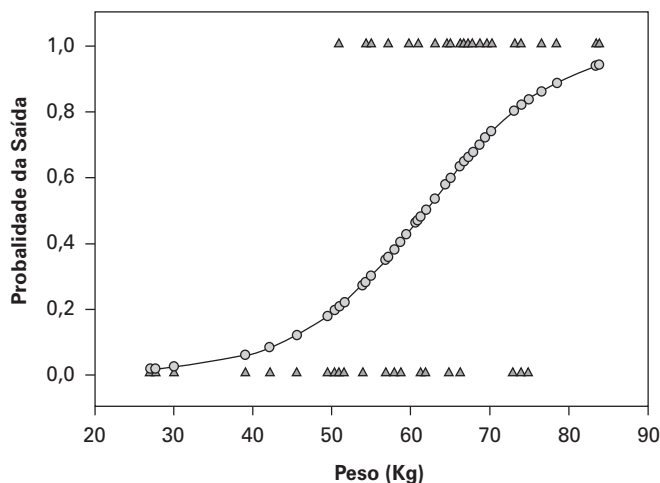
Como regra geral, sempre que amostras são divididas em categorias e uma ou mais combinações estiverem vazias teremos problemas. Isso será representado por coeficientes que apresentam erros padrão anormalmente grandes. Pesquisadores conscienciosos produzem e verificam tabulações cruzadas multiformes de todas as variáveis categóricas independentes. Os preguiçosos, mas cautelosos, não irão se incomodar com tabulações cruzadas, porém, irão olhar cuidadosamente os erros padrão. Aqueles que não se preocupam com nada disso terão problemas.

6.9.2 Separação completa ④

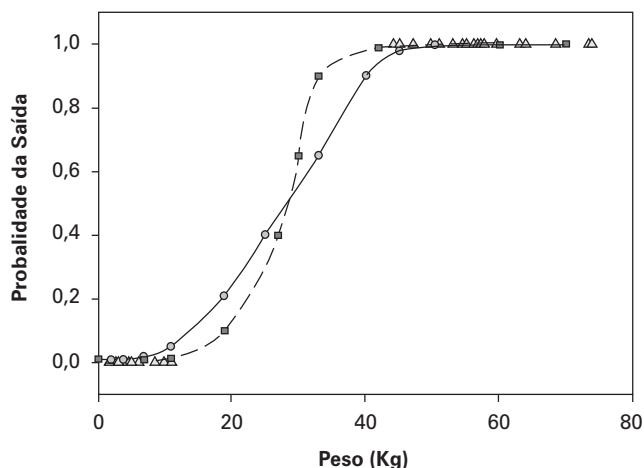
Uma segunda situação em que a regressão logística malogra pode lhe surpreender: quando a variável de saída pode ser perfeitamente prevista por uma variável ou uma combinação de variáveis! Isso é conhecido como **separação completa**.

Vamos ilustrar com um exemplo: imagine que você colocou uma almofada de pressão embaixo do seu tapete da porta de entrada e conectou a almofada ao sistema de segurança, de modo que você possa detectar ladrões se eles tentarem entrar à noite. Contudo, os filhos adolescentes (que você terá se for velho e rico o suficiente para possuir um sistema de segurança e almofadas de pressão) e seus amigos podem chegar no meio da noite. Quando eles pisarem no tapete, você quer determinar a probabilidade de que a pessoa é um ladrão e não um adolescente. Para tanto, você pode pesar alguns ladrões e adolescentes e utilizar a regressão logística para prever a saída (ladrão ou ado-

lescente) a partir do peso. O gráfico irá mostrar uma linha de triângulos em 0 (os pontos de dados dos adolescentes que você mediu) e uma linha de triângulos em 1 (os pontos de dados dos ladrões que foram pesados). Note que essas linhas de triângulos se sobrepõem (alguns adolescentes são tão pesados quanto os ladrões). Vimos que numa regressão logística, o SPSS tenta prever a probabilidade de uma saída dado um valor do preditor. Nesse caso, com pesos baixos



a probabilidade ajustada segue a linha inferior do diagrama e com pesos altos ela segue a linha superior. Em valores intermediários, ela tenta seguir a probabilidade como ela muda.



Imagine que nós temos a mesma almofada de pressão, se mudou e foi para a universidade. Agora estamos interessados em distinguir ladrões de gatos de estimação com base no peso. Novamente, podemos pesar alguns gatos e alguns ladrões. Dessa vez o gráfico ainda tem uma linha de triângulos em 0 (os pesos dos gatos) e uma linha 1 (os dos ladrões), mas agora as linhas dos triângulos não se sobrepõem, pois não existe um ladrão que pese o mesmo que um gato. Isso é conhecido como uma separação perfeita: a saída (gato ou ladrão) pode ser prevista perfeitamente a partir dos pesos (qualquer coisa abaixo de 15 kg é um gato e qualquer coisa acima de 40 kg é um ladrão). Se tentarmos calcular as probabilidades de uma saída dado certo peso, estaremos encrencados. Quando o peso é baixo, a probabilidade é zero, e quando é alto, a probabilidade é 1, mas o que acontece no meio? Não temos dados entre 15 e 40 em que basear essas probabilidades. A figura mostra duas curvas possíveis que podem ser ajustadas a esses dados, sendo que uma é bem mais íngreme que a outra. Qualquer uma dessas curvas é válida com base nos dados que temos. A falta de dados fará com que o SPSS não tenha certeza de qual íngreme deve ser a inclinação intermediária e ele tentará trazer o centro o mais próximo possível da vertical e estimará a mudança em direção ao infinito (com grandes erros padrão).

Esse problema às vezes ocorre quando muitas variáveis são ajustadas a poucos casos. Em muitos casos, a única solução satisfatória é coletar mais dados, mas algumas vezes uma resposta é encontrada adotando um modelo mais simples.

6.10 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Aprendi muito escrevendo este capítulo (como ignorar meu cérebro quando ele pensa “Eu sei o que seria ótimo colocar no meu livro, o capítulo de regressão logística!”). Pelo menos, espero que eu tenha conseguido disfarçar a minha incompetência pelo menos na

maior parte do tempo. Começamos o capítulo entendendo por que não podemos utilizar a regressão linear (Capítulo 5) quando temos uma variável de saída dicotômica. Estudamos então um pouco de teoria sobre a regressão logística analisando a equação de regressão e seu significado. Depois, avaliamos o modelo e falamos sobre a estatística verossimilhança-log e o teste qui quadrado associado. Abordei os diferentes métodos de obter valores equivalentes ao R^2 na regressão (Hosmer e Lemeshow, Cox e Snell e Nagelkerke). Descobrimos também a estatística de Wald e o Exp(B). O restante do capítulo apresentou dois exemplos utilizando o SPSS para executar a regressão logística. Com isso, espero que você tenha compreendido de forma geral como conduzir e executar uma regressão logística.

6.11 TERMOS-CHAVE QUE DESCOBRIMOS

- $-2VL$
- Distribuição qui-quadrado
- Separação completa
- R^2_{CS} de Cox e Snell
- Exp(B)
- R^2_L de Hosmer e Lemeshow
- Regressão Logística
- Verossimilhança-log
- Estimação de máxima verossimilhança
- R^2_N de Nagelkerke
- Chances
- Estatística do escore eficiente de Roa
- Estatística de Wald

6.12 TAREFAS DO ALEX ESPERTO



As respostas dessas tarefas estão no *site* www.artmed.com.br no arquivo **Answers(Chapter6).pdf**.

- Pesquisas recentes mostraram que os professores estão entre os trabalhadores mais estressados. Uma pesquisadora quer saber quais os fatores que provocam estresse em um professor. Ela entrevistou 467

professores e utilizou vários questionários que mediram: **burnout** (colapso) (teve ou não), **perceived control** (controle percebido) (escore alto = controle percebido baixo), **coping style** (estilo de enfrentar) (escore alto = alta habilidade de enfrentar o estresse), **stress from teaching** (estresse de ensinar) (escore alto = lecionar estressa muito a pessoa), **stress from research** (estresse de pesquisar) (escore alto = pesquisar provoca muito estresse na pessoa) e **stress from providing pastoral care** (estresse de aconselhar) (escore alto = aconselhamento provoca muito estresse na pessoa). A variável de saída de interesse é **burnout** (colapso), e o modelo de estresse de Cooper, Sloan e Williams (1988) indica que o controle percebido (*perceived control*) e o estilo de enfrentar (*coping style*) são previsores importantes dessa variável. Os previsores restantes foram medidos para ver a sua contribuição sobre diferentes aspectos do trabalho do professor e seu colapso nervoso. Você pode ajudar a pesquisadora realizando uma regressão logística a fim de ver quais fatores preveem o colapso nervoso? Os dados estão no arquivo **Burnout.sav**. ③

- Um psicólogo clínico está interessado em pesquisar sobre o HIV e quer saber os fatores que influenciam o uso do preservativo com uma nova parceira (relacionamentos com menos de um mês de duração). A medida de saída é se o preservativo foi ou não-utilizado (**use** = preservativo utilizado = 1, **not used** = preservativo não-utilizado = 0). As variáveis previsoras foram principalmente resultados obtidos da utilização da EAP (Escala de Atitudes em relação ao Preservativo)* de Sacco, Levine, Reed e Thompson (*Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1991). **Gender** (Gênero – gênero da pessoa); **Safety** (Segurança –

relacionamento seguro medido em uma escala de cinco pontos indicando o grau em que a pessoa vê a segurança do relacionamento em relação a doenças sexualmente transmissíveis); **Sexexp** (experiência sexual, medida em uma escala de 10 pontos indicando o grau em que experiências anteriores influenciam a atitude em relação ao uso de preservativos); **previous** (uma medida fora da escala EAP, que indica se o preservativo foi ou não utilizado pelo casal em encontros anteriores, 1 = preservativo utilizado, 0 = preservativo não utilizado, 2 = não houve encontro anterior com essa parceira); **selfcon** (autocontrole medido em uma escala de nove pontos, indicando o autocontrole da pessoa quando chega o momento de utilizar o preservativo; isto é, eles se deixaram levar pelo momento ou se controlaram?); **Perceive** (risco percebido, medido em uma escala de seis pontos, indicando o grau de preocupação da pessoa em relação ao sexo desprotegido). Pesquisas anteriores (Sacco, Rickman, Thompson, Levine e Reed, em *AIDS Education and Prevention*, 1993), mostraram que **gênero, relacionamento seguro e risco percebido** preveem o uso de preservativo. Execute uma análise apropriada para verificar esses resultados anteriores e testar se autocontrole, utilização anterior e experiência sexual podem prever qualquer variância restante na utilização de preservativo. (1) Interprete todas as partes importantes da saída do SPSS. (2) Quão confiável é o modelo final? (3) Quais são as probabilidades que os participantes 12, 53 e 75 irão utilizar preservativos? (4) Uma mulher utilizou preservativo em seu encontro anterior com um novo parceiro, e apresenta 2 em todas as variáveis exceto em risco percebido (na qual ela tem um escore de 6). Utilize o modelo para estimar a probabilidade de que ela irá utilizar preservativo em seu próximo encontro. Os dados estão no arquivo **condom.sav**. ③

* N. de T.: CAS (Condom Attitude Scale).

6.13 LEITURAS COMPLEMENTARES

HUTCHESON, G., SOFRONIOU, N. *The multivariate social scientist*. London: Sage, 1999. Capítulo 4.

MENARD, S. *Applied logistic regression analysis. Sage university paper series on quantitative applications in the social sciences*. Thousand

Oaks (CA): Sage. 1995, pp. 07-106. Esse é um texto avançado, mas bom. Infelizmente, poucos textos básicos incluem a regressão logística, portanto, você terá que confiar no que eu escrevi!

MILES, J., SHEVLIN, M. *Applying regression and correlation: a guide for students and researchers*. London: Sage, 2001. O Capítulo 6 apresenta uma bela introdução à regressão logística.

COMPARANDO DUAS MÉDIAS

7.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

Em vez de olhar os relacionamentos entre variáveis, os pesquisadores às vezes estão interessados em verificar as diferenças entre grupos de pessoas. Em pesquisas experimentais em geral queremos manipular o que acontece às pessoas para que possamos fazer inferências causais. Por exemplo, se tomarmos dois grupos de pessoas e aleatoriamente designarmos um grupo a um programa de pílulas dietéticas e o outro grupo a um programa de pílulas de açúcar (que eles acham que irá auxiliá-los na perda de peso), se as pessoas que tomam as pílulas dietéticas perderem mais peso do que as que tomam pílulas de açúcar podemos inferir que as pílulas dietéticas causaram a perda de peso. Isso é uma ferramenta poderosa de pesquisa porque vai além da mera observação de variá-

veis e da procura por relacionamentos (como na correlação e regressão).¹ Este capítulo é o primeiro de muitos que aborda esse tipo de linha de pesquisa e começamos com o cenário mais simples: quando temos dois grupos ou, para ser mais específico, quando queremos comparar duas médias. Como veremos, existem duas maneiras de coletar dados: podemos expor diversas pessoas a diferentes manipulações experimentais (*entre grupos* ou *delineamento independente*) ou tomar um único grupo de pessoas e expô-lo a diferentes manipulações experimentais em pontos diferentes no tempo (um delineamento de *medidas repetidas*). Primeiro veremos como fazer um gráfico desses diferentes tipos de dados e, então, veremos os procedimentos estatísticos para análise de delineamentos independentes (o *teste t independente*). Depois, seguimos para o delineamento de medidas repetidas (o *teste t dependente*).

¹ As pessoas às vezes ficam confusas e pensam que certos procedimentos estatísticos permitem inferências causais e outros não. Isso não é verdade; é o fato de que em experimentos nós manipulamos uma variável sistematicamente para ver seu efeito em outra que permite a inferência causal. Em outras formas de pesquisa, meramente obser-

vamos mudanças em variáveis sem qualquer intervenção do investigador e, assim, não podemos dizer qual variável causa a mudança na outra, mas apenas que elas mudam de uma maneira específica. Como vocês irão ver os procedimentos estatísticos para analisar os diferentes tipos de dados são, na verdade, matematicamente idênticos!

7.2 REVISÃO DA PESQUISA EXPERIMENTAL ①

Geralmente, nas ciências sociais não estamos apenas interessados em ver quais as variáveis que covariam ou prevêm uma saída. Em vez disso, queremos olhar para o efeito de uma variável em outra pela alteração sistemática de alguns aspectos daquela variável. Assim, em vez de coletar dados que ocorrem naturalmente como na correlação e regressão, manipulamos uma variável para observar seus efeitos em outra. Como um caso simples, podemos querer ver qual o efeito que um incentivo positivo tem no aprendizado sobre estatística. Posso, portanto, dividir aleatoriamente alguns estudantes em dois grupos diferentes:

- **Grupo 1 (estímulo positivo):** Durante as aulas, parabeno todos os estudantes nesse grupo pelo seu esforço e sucesso. Mesmo quando as pessoas fazem algo errado, eu as apóio e digo coisas como “você quase acertou, está progredindo muito bem” e dou um chocolate para o estudante.
- **Grupo 2 (estímulo negativo):** Esse grupo recebe uma aula padrão; assim, pratico abuso verbal mesmo quando os alunos fornecem a resposta correta. Não valorizo suas contribuições, sou condescendente e desdenho tudo o que dizem. Chamo os estudantes de estúpidos, ignorantes e afirmo que eles não deveriam estar fazendo esse curso.

Qual é a diferença entre pesquisa experimental e correlacional?



A variável que manipulei é o método de ensino (estímulo positivo *versus* estímulo negativo). Essa variável é conhecida como *variável independente* (VI) e essa situação têm dois níveis por-

que ela foi manipulada de duas maneiras (isto é, o estímulo foi dividido em dois tipos: positivo e negativo). Uma vez conduzida essa manipulação, preciso ter algum tipo de saída que estou interessado em medir. Nesse caso, é a habilida-

de estatística, e eu posso medir essa variável usando os resultados da prova do final do ano. Essa variável de saída é conhecida como *variável dependente* ou VD porque assumimos que esses escores irão depender do tipo do método de ensino utilizado (a *variável independente*).

7.2.1 Os dois métodos de coletar dados ①

Podemos escolher entre dois métodos de coleta de dados. O primeiro é manipular a variável independente usando participantes diferentes. Esse é o método descrito acima, em que diferentes grupos de pessoas fazem parte de cada condição experimental (*entre grupos*, *entre participantes* ou *delineamento independente*). O segundo método é manipular a variável independente usando os mesmos participantes. Simplificando, esse método significa que damos a um grupo de estudantes estímulo positivo por algumas semanas e testamos suas habilidades estatísticas e, depois, damos a esse mesmo grupo estímulo negativo por algumas semanas antes de testá-los novamente (*delineamento dentro participantes* ou *de medidas repetidas*). A maneira na qual os dados são coletados determina o tipo de teste que será utilizado para analisar os dados.

7.2.2 Dois tipos de variação ①

Imagine que você usou o delineamento de medidas repetidas num experimento que tem duas condições. Portanto, os mesmos participantes participam em ambas as condições, ou seja, medimos o comportamento dos participantes na condição 1 e na condição 2. Por exemplo, suponha que estamos tentando ver se conseguimos treinar chimpanzés para administrar a economia. Em uma fase do treino, eles sentam na frente de um computador e clicam teclas que mudam vários parâmetros da economia; uma vez que esses parâmetros mudaram, uma figura aparece na tela indicando o crescimento econômico resultante desses parâmetros. Como chimpanzés não sabem ler, essa realimentação não tem sentido. A segunda fase de treino é a mesma, exceto que, se

o crescimento da economia for bom, eles ganham uma banana (se o crescimento é ruim, eles não ganham) – essa realimentação é valiosa para o chimpanzé.

Vamos parar um pouco e pensar no que teria acontecido se **não** tivéssemos introduzido uma manipulação experimental (isto é, não houvesse bananas na segunda fase de treinamento, assim, as condições 1 e 2 seriam idênticas). Se não há manipulação experimental, esperamos que o comportamento de um chimpanzé seja o mesmo em ambas as condições. Esperamos isso porque fatores externos como idade, gênero, QI, motivação e excitação serão os mesmos em ambas as condições (o gênero do chimpanzé etc. não mudará da condição 1 para a condição 2). Se o desempenho da medição é confiável (isto é, testa quão bem eles administraram a economia) e a variável ou a característica que estamos medindo (nesse caso, a habilidade de administrar a economia) permanece estável, então o desempenho do participante na condição 1 deverá estar altamente relacionado ao seu desempenho na condição 2. Assim, chimpanzés que tiveram um alto desempenho na condição 1 terão também um alto desempenho na condição 2, e aqueles que tiveram baixo desempenho na condição 1 terão baixo desempenho na condição 2. Entretanto, o desempenho não será *idêntico*; haverá pequenas diferenças no desempenho criadas por fatores desconhecidos. Essa variação no desempenho é conhecida como *variação não-sistemática*.

Se introduzirmos uma manipulação experimental (isto é, fornecermos bananas como *feedback* em uma das seções do treinamento), diferenciamos o tratamento dos participantes da condição 1 e da condição 2. Assim, a *única* diferença entre as condições 1 e 2 é a manipulação que o investigador fez (nesse caso, os chimpanzés receberam bananas como uma recompensa positiva em uma condição, mas não na outra). Portanto, qualquer diferença entre as médias das duas condições é, provavelmente, devido à manipulação do experimento. Assim, se os chimpanzés desempenharam melhor numa fase do que na outra, isso *deve* ter ocorrido porque bananas foram usadas para

fornecer um reforço ou realimentação em uma fase de treinamento e não na outra. As diferenças no desempenho criadas por uma manipulação experimental específica são conhecidas como *variação sistemática*.

Agora, vamos pensar no que acontece quando usamos participantes diferentes – um delineamento independente. Nesse delineamento ainda teremos duas condições, mas, dessa vez, participantes diferentes participam em cada condição. Voltando para o nosso exemplo, um grupo de participantes recebe treinamento sem realimentação e o segundo grupo de diferentes chimpanzés recebe reforço para o seu desempenho via bananas². Imagine, novamente, que não temos uma manipulação experimental. Se nada fizessemos com os grupos, ainda encontraríamos alguma variação de comportamento entre eles, porque diferentes chimpanzés irão variar na sua habilidade, motivação, QI e outros fatores. Resumindo, o tipo de fatores que eram mantidos constantes no delineamento de medidas repetidas está livre para variar no delineamento independente. Assim, a variação não-sistemática será maior do que para o delineamento de medidas repetidas. Se novamente introduzimos uma manipulação (isto é, bananas), mais uma vez teremos variação adicional criada por essa manipulação. Como tal, tanto no delineamento de medidas repetidas quanto no delineamento independente há sempre duas fontes de variação:

- **Variação sistemática:** Essa variação ocorre porque o investigador faz algo a todos os participantes em uma condição, mas não na outra.
- **Variação não-sistemática:** Essa variação resulta de fatores aleatórios que existem entre as condições experimentais (tais como diferenças naturais em habilidades).

² Quando eu digo “via”, não significa que as bananas têm pequenas bocas e falaram “muito bem camarada, a economia cresceu desta vez” na linguagem de chimpanzés. O que eu quero dizer é que quando eles acertam eles recebem uma banana como recompensa pela resposta correta.

O papel da estatística é descobrir quanta variação existe no desempenho e, então, decidir quanto disso é sistemático e quanto não é sistemático.

Em um delineamento de medidas repetidas, as diferenças entre duas condições podem ser causadas somente por dois fatores: (1) a manipulação que foi realizada nos participantes ou (2) qualquer outro fator que possa afetar a maneira que uma pessoa executa tarefas repetidamente. Esse último fator é menor comparado à influência da manipulação experimental. Em um delineamento independente, as diferenças entre duas condições também podem ser causadas por uma de duas coisas: (1) a manipulação que foi conduzida nos participantes ou (2) as diferenças entre características das pessoas alocadas a cada um dos grupos. Esse último fator em geral cria uma considerável variação aleatória dentro de cada condição e entre elas. Portanto, o efeito da nossa manipulação experimental provavelmente será mais visível no delineamento de medidas repetidas do que no delineamento entre grupos, já que no primeiro a variação não-sistemática pode ser causada somente por diferenças na maneira com que alguém se comporta em oportunidades distintas. Nos delineamentos entre grupos, diferenças entre habilidades inatas contribuem para a variação não-sistemática. Portanto, essa variação de erro será quase sempre maior do que se o mesmo participante fosse utilizado. Quando observamos o efeito da nossa manipulação experimental, é sempre contra um contexto de “ruído” causado por diferenças aleatórias e incontrolláveis entre nossas condições. Em um delineamento de medidas repetidas, esse “ruído” é mínimo. Desse modo, é mais provável que o efeito do experimento apareça. Isso significa que o delineamento de medidas repetidas é melhor para detectar efeitos que genuinamente existem do que delineamentos independentes.

7.2.3 Aleatorização ①

Tanto nos delineamentos de medidas repetidas quanto nos delineamentos independentes, é importante tentar manter a variação não-sistemática no menor nível possível. Dessa forma,

conseguimos uma medida mais sensível da manipulação experimental. Geralmente os cientistas usam a *aleatorização* dos participantes para atingir esse objetivo. Muitos testes estatísticos (por exemplo, o teste *t*) identificam as fontes sistemáticas e não-sistemáticas da variação e depois as comparam. Essa comparação permite ver se o experimento gerou consideravelmente mais variação do que conseguiríamos se tivéssemos testado os participantes sem a manipulação experimental. A aleatorização é importante porque ela elimina a maioria das outras fontes de variação sistemática, o que nos permite ter certeza de que qualquer variação sistemática entre condições experimentais é devido à manipulação da variável independente.

Vejam os delineamentos de medidas repetidas primeiro. Quando as mesmas pessoas participam em mais de uma condição experimental, isso obviamente viola a noção de atribuir aleatoriamente pessoas aos grupos. Embora elas sejam ingênuas durante a primeira condição experimental, na segunda condição as pessoas têm experiência prévia do que é esperado delas. No mínimo, elas estarão familiarizadas com a medida dependente (por exemplo, a tarefa que elas estão desenvolvendo). As duas fontes mais importantes da variação sistemática nesse tipo de delineamento são:

- **Efeitos práticos:** Os participantes podem ter um desempenho diferente na segunda condição por causa da familiaridade com a situação experimental e/ou com as medidas usadas.
- **Efeitos de tédio:** Os participantes podem ter um desempenho diferente na segunda condição porque eles estão cansados ou entediados depois de completar a primeira condição.

Embora seja impossível eliminar completamente esses efeitos, podemos assegurar que eles não produzam variação sistemática entre nossas condições. Fazemos isso *contrabalançando* a ordem em que uma pessoa participa de uma condição. Podemos decidir aleatoriamente se um participante faz a condição 1 antes da condição 2 ou a 2 antes da 1.

Retornando ao exemplo do método de ensino no começo deste capítulo, se os mesmos participantes foram usados em ambas as condições, podemos achar que a habilidade estatística foi mais alta após a condição de estímulo negativo. Entretanto, cada estudante experienciou o estímulo negativo após o positivo e, assim, foi para a condição de estímulo negativo já tendo um conhecimento melhor de estatística do que quando eles começaram a condição de estímulo positivo. Assim, a melhoria aparente após o estímulo negativo não ocorre devido à manipulação experimental (isto é, não é por causa do estímulo negativo), mas sim porque os participantes assistiram a mais aulas de estatística quando eles terminaram a condição de estímulo negativo. Para assegurar que o número de aulas de estatística não introduza um viés no resultado, podemos distribuir a ordem das condições de modo aleatório; assim, metade dos estudantes poderia ter o estímulo negativo primeiro e a outra metade, o positivo primeiro.

Um argumento similar pode ser aplicado nos delineamentos independentes. Sabemos que pessoas diferentes participam em diferentes condições experimentais e que essas pessoas irão diferir em muitos sentidos (QI, tempo de atenção, etc.). Embora saibamos que esses fatores (conhecidos como *variáveis de confusão*) contribuam para a variação entre condições, precisamos ter certeza de que essas variáveis contribuam com a variação não-sistemática e *não* com a variação sistemática. Para tanto, é preciso distribuir aleatoriamente os participantes a uma condição experimental particular. Isso deve assegurar que as variáveis de confusão estejam distribuídas igualmente nas condições.

Um bom exemplo são os efeitos do álcool na personalidade. Você pode dar a um grupo de pessoas cinco canecas de cerveja e manter um segundo grupo sóbrio e, então, contar em quantas brigas cada pessoa entrou. O efeito que o álcool tem sobre as pessoas varia muito devido aos diferentes níveis de tolerância: pessoas abstinidas podem ficar muito embriagadas com uma pequena quantidade enquanto alcoolistas precisam consumir grandes quantidades para que o álcool provoque algum

efeito. Se você distribuir um grupo de abstêmios numa condição que consumiu álcool, você pode não encontrar diferença entre eles e o grupo sóbrio (porque os participantes abstêmios ficarão inconscientes após o primeiro copo e, assim, não poderão se envolver em briga alguma). Assim, a experiência prévia da pessoa com álcool irá criar uma variação sistemática que não pode ser dissociada do efeito da manipulação experimental. A melhor maneira de reduzir isso é distribuir os participantes às condições de modo aleatório.

7.3 ATRIBUINDO DADOS E APRESENTANDO MÉDIAS COM DIAGRAMAS DE BARRAS DE ERROS

Como tentei enfatizar no Capítulo 3, é sempre recomendável observar os seus dados com gráficos antes de começar a análise. Os gráficos nos auxiliam a entender o que está acontecendo e que tipo de resultados

esperamos obter. Já vimos como criar diagramas de barras de médias na Seção 3.7. Diagramas de barras são frequentemente apresentados em artigos acadêmicos e são uma boa maneira de resumir dados, mas eles fornecem uma visão muito limitada dos dados (por exemplo, os diagramas de barras não indicam o número de escores que contribuíram para cada média nem nos dizem quanta variação havia nos escores). Os diagramas de barra de erros, abordados na Seção 3.7, são uma maneira melhor de examinar os dados. Um diagrama de **barra de erros** mostra não somente a média, mas também um intervalo de 95% confiança (veja a Seção 1.6.2) para a média de cada condição experimental.

Você já deve estar familiarizado com os intervalos de confiança, já que eles foram vistos nos Capítulos 1, 5 e 6. Vamos retornar ao exemplo dos espermatozoides da Seção 1.6.2, no qual tentávamos medir a média do volume de espermatozoides (em milhões) produzida pela codorna japonesa. Expliquei que se to-



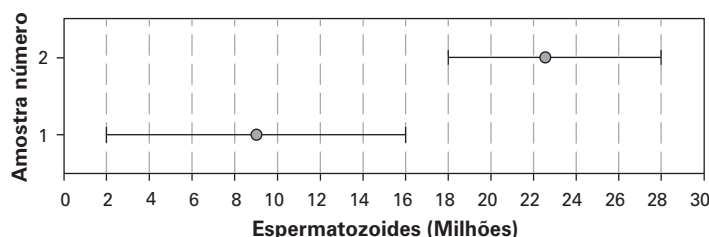
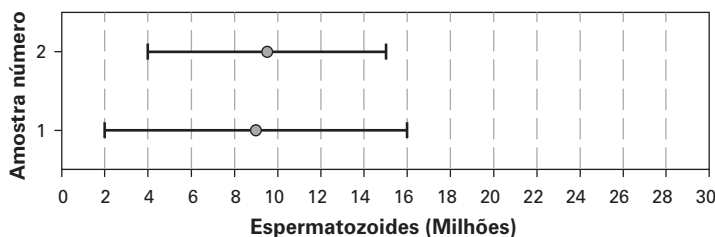
marmos 100 amostras de uma população, calcularmos a média de cada uma e construirmos o intervalo de confiança de 95% para cada, isso significa que 95 dos 100 intervalos de confiança conteriam o valor real da média (o valor da população). Assim, se um intervalo de confiança tem um limite inferior de 2 e um limite superior de 16, isso significa que estamos 95% confiantes que esse intervalo contém a média da população (isto é, a quantidade real de espermatozoides produzido pela codorna). Portanto, um diagrama de barra de erros mostra os limites entre os quais pensamos que o valor real da média se encontra.

Em breve você irá descobrir que o teste *t* funciona com base no princípio de que se duas amostras são tiradas da mesma população, então elas devem ter médias bem similares. Sabemos do Capítulo 1 que é possível que quaisquer duas amostras tenham médias ligeiramente diferentes (e o erro padrão nos informará até que ponto as médias podem ser diferentes). O intervalo de confiança nos diz os limites entre os quais a média da população provavelmente estará incluída (na verdade, o tamanho do intervalo de confiança dependerá do tamanho do erro padrão). No exemplo acima, sabemos que a média da população está provavelmente entre 2 e 16 milhões de espermatozoides. E se pegarmos uma segunda

amostra de codornas e encontrarmos um intervalo de confiança variando de 4 a 15? Esse intervalo se sobrepõe em boa parte ao da primeira amostra:

O fato de que os intervalos de confiança se sobrepõem dessa maneira nos diz que essas médias poderiam, plausivelmente, vir da mesma população: nos dois casos estamos 95% confiantes de que os intervalos contêm o valor real da média e ambos os intervalos se sobrepõem consideravelmente, assim, eles contêm muitos valores similares. E se o intervalo de confiança para a nossa segunda amostra variasse de 18 a 28? Se compararmos isso com nossa primeira amostra, temos:

Agora esses intervalos de confiança não se sobrepõem. Assim, um intervalo de confiança (o qual estamos 95% certos que contém o valor real da média) nos diz que a média da população é algo entre 2 e 16 milhões, e o outro intervalo (o qual também estamos 95% certos que contém o valor real da média) nos informa que a média da população está entre 18 e 28. Isso sugere que ou nossas amostras vêm de diferentes populações ou ambas as amostras vêm da mesma população, mas um dos intervalos não deve conter a média da população. Se utilizarmos intervalos de 95% de confiança, saberemos que a segunda possibilidade é improvável (isso acontece somente 5



vezes em 100 ou 5% das vezes), assim, a primeira explicação é a mais provável.

OK, vocês devem estar pensando: “e se as amostras vêm de populações diferentes?”. Bem, se pegamos duas amostras aleatórias de pessoas e as testamos em alguma medida (por exemplo, medo de livros de estatística), esperamos que essas pessoas pertençam à mesma população. Se as médias das amostras são tão diferentes para sugerir que, de fato, elas vêm de diferentes populações, como isso pode acontecer? A resposta é que nossa manipulação experimental induziu uma diferença entre as amostras.

Repetindo, quando uma manipulação experimental é bem-sucedida, esperamos que nossas amostras venham de diferentes populações. Se a manipulação não foi um sucesso, esperamos que nossas amostras venham da mesma população (isto é, as médias das amostras devem ser similares). O intervalo de 95% de confiança nos informa algo sobre o provável valor da média da população. Se tirarmos amostras de duas populações, esperamos que os intervalos de confiança sejam diferentes (para termos certeza de que as amostras são de diferentes populações, não pode haver a sobreposição dos dois intervalos de confiança). Se tirarmos duas amostras da mesma população, esperamos, se nossa medida é confiável, que os intervalos de confiança sejam muito similares (isto é, eles devem se sobrepor completamente). Você pode estar se perguntando onde esse desvio para a teoria dos testes de hipóteses está nos levando; bem, tem a ver com o que eu acabei de dizer que se nossa manipulação experimental tiver sucesso, os intervalos de confiança dos grupos-experimentais não devem se sobrepor. Se as manipulações não tiverem sucesso, os intervalos de confiança irão se sobrepor. Em termos de diagrama de barra de erros, veremos que **se as barras no nosso diagrama de barra de erros não se sobrepõem, isso indica uma diferença significativa entre grupos**. Também vale a pena mencionar, nesse momento, que diagramas de barras do SPSS são adequados somente para dados normalmente distribuídos.

7.3.1 Diagramas de barras de erros para delineamentos entre grupos ①

Quando os dados são coletados usando participantes diferentes em cada grupo, precisamos entrar com os dados usando uma variável codificadora (veja a Seção 2.4.4). Assim, o editor de dados terá duas colunas de dados. A primeira coluna é uma variável codificada ou de código (chamada de **grupo**), a qual, quando temos somente dois grupos, irá obviamente ter dois códigos (por conveniência, sugiro 0 = grupo 1 e 1 = grupo 2). A segunda irá ter valores para a variável dependente.

Ao longo deste capítulo, uso o mesmo conjunto de dados, não porque sou muito preguiçoso para pensar em diferentes conjuntos de dados, mas porque ele me permite ilustrar vários pontos. O exemplo é se a aracnofobia (medo de aranhas) é específico a aranhas reais ou se figuras de aranhas podem evocar níveis semelhantes de ansiedade. No total, foram usados 24 aracnofóbicos. Foi pedido a 12 que brincassem com uma tarântula gigante e cabeluda, com grandes presas e um olhar maligno nos seus oito olhos. A ansiedade desse grupo foi medida. Aos 12 restantes foi apenas mostrado figuras da mesma tarântula grande e cabeluda e, novamente, sua ansiedade foi medida. Os dados estão no arquivo **spiderBG.sav**, mas para quem quiser praticar a entrada de dados, eles estão também presentes na Tabela 7.1. Lembre que cada linha no editor de dados representa um dado diferente de cada participante. Portanto, você precisa de uma coluna para designar o grupo a que eles pertencem e uma segunda coluna para representar sua ansiedade. Os dados da Tabela 7.1 mostram somente os códigos do grupo e não o rótulo correspondente. Quando você entrar os dados no SPSS, lembre de informar ao computador que o código 0 representa o grupo para o qual foi mostrada a figura e o código 1 representa o grupo que viu a aranha verdadeira (veja a Seção 2.4.4).

Depois que você entrou com os dados (ou acessou o arquivo **spiderBG.sav**), acesse a caixa de diálogo do *diagrama da barra de erros* clicando em **Graphs⇒Error Bar...** (Diagramas⇒Barra de Erros...). A caixa de

Tabela 7.1 Dados do spiderGB.sav

Participante	Grupo	Ansiedade
1	0	30
2	0	35
3	0	45
4	0	40
5	0	50
6	0	35
7	0	55
8	0	25
9	0	30
10	0	45
11	0	40
12	0	50
13	1	40
14	1	35
15	1	50
16	1	55
17	1	65
18	1	55
19	1	50
20	1	35
21	1	30
22	1	50
23	1	60
24	1	39

diálogo inicial é mostrada na Figura 7.1. Há duas escolhas do gráfico do diagrama da barra de erros. A primeira é um diagrama de barra de erro simples (para traçar níveis de uma única variável independente) e a segunda é um diagrama do gráfico de erros agrupados. O gráfico agrupado pode ser usado quando uma segunda variável independente entre grupos foi medida (tal como gênero). Nós nos limitaremos a usar um diagrama de barra de erro simples. A próxima escolha é se os dados no diagrama resumem grupos de casos ou variáveis isoladas. Há uma regra simples aqui: selecione *Summaries for groups of cases* (resumo para grupos de casos) quando os dados forem coletados usando diferentes participantes (como é o caso) e selecione *Summaries of separate variables* (resu-

mo para variáveis isoladas) quando os dados forem coletados usando os mesmos participantes. Quando você tiver selecionado as opções apropriadas, clique em **Define** (Definir).

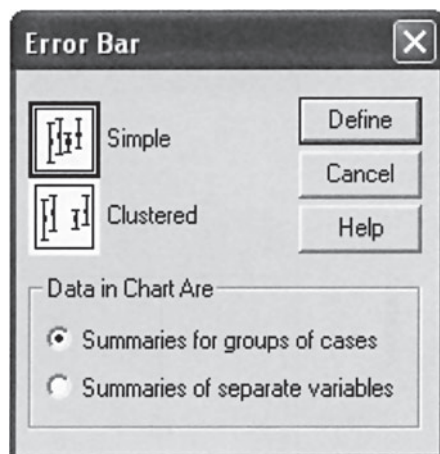


Figura 7.1 Caixa de diálogo inicial para diagramas de barras de erro.

Quando você clicar em **Define** (Definir), uma nova caixa de diálogo aparece (Figura 7.2), que permite especificar que variável iremos apresentar. Nesse exemplo, temos apenas duas variáveis: **group** (grupo) e **anxiety** (ansiedade). Assim, usando o mouse, selecione *Anxiety (anxiety)* da lista de variáveis e insira-a no espaço rotulado de *Variable* (variável) clicando em **►**. Depois, destaque *Condition (group)* na lista de variáveis e transfira-a para o espaço rotulado de *Category Axis* (eixo das categorias) clicando em **►**. Com isso você pode desenhar várias coisas usando esse tipo de gráfico, e a opção padrão é apresentar um intervalo de 95% de confiança. Este é o tipo de gráfico mais útil então as opções padrão podem ser mantidas.

O diagrama da barra de erros resultante é mostrado na Figura 7.3 e, como você pode ver, ele parece como dois Is. No meio de cada uma das duas barras há um quadrado que representa a média de cada grupo. A barra vertical mostra o intervalo de confiança em volta daquela média. Assim, a partir do gráfico desses dados pode-

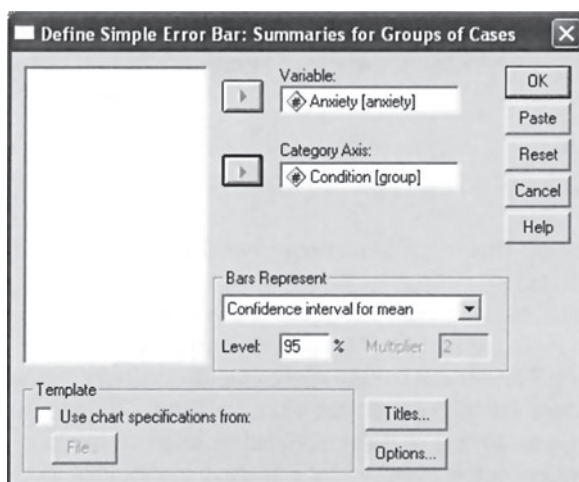


Figura 7.2 Caixa de diálogo principal para diagrama de barras de erro para resumos de grupos de casos.

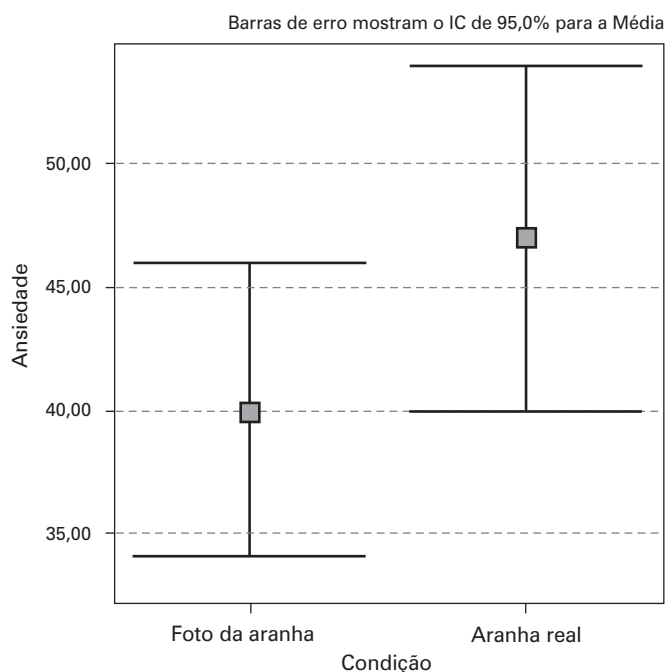


Figura 7.3 Diagrama da barra de erros de **spiderBG.sav**.

mos ver que quando a figura de uma aranha foi usada, o nível médio de ansiedade foi 40, e que provavelmente a média da população deve ficar entre 34 e 46. Entretanto, quando uma aranha

verdadeira foi usada, o nível médio de ansiedade foi 47, e provavelmente a média da população deve ficar entre 40 e 54. Mais importante, as barras de erro se sobrepõem consideravelmente,

indicando que é improvável que essas amostras sejam de populações diferentes (portanto, a manipulação experimental não teve sucesso). Para ver se essa última afirmação é verdadeira, você terá de esperar até a Seção 7.6.3.

7.3.2 Diagramas de barras de erros para delineamentos de medidas repetidas ②

Se repetirmos o estudo recém-descrito usando os mesmos participantes em cada condição, podemos produzir um diagrama de barras de erro idêntico ao mostrado na Figura 7.3. O problema com a criação de um diagrama de barras de erro de dados de medidas repetidas é que o SPSS trata os dados como se diferentes grupos de participantes tivessem sido usados. Para provar que não estou mentindo, os dados para esse estudo estão incluídos no arquivo chamado de **spiderRM.sav**. Nesse arquivo, os valores da ansiedade são idênticos aos dados entre grupos. Entretanto, os dados estão organizados como se os mesmos participantes tivessem sido usados em cada condição (ou seja, cada participante foi exposto a uma figura de uma aranha e sua ansiedade foi medida, e em outra ocasião os mesmos participantes foram expostos a aranhas verdadeiras e sua ansiedade foi medida novamente). Os dados estão organizados de modo diferente agora porque os mesmos participantes foram usados. No SPSS, cada linha do editor de dados representa um único participante, assim, com esse desenho os dados são organizados em duas colunas (uma representando a condição da **figura da aranha** e uma representando a condição **aranha verdadeira**). Os dados estão expostos na Tabela 7.2, e eu recomendo que você tente inserir esses dados em um novo editor de dados em vez de acessar o arquivo em disco.

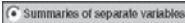


Para traçar o gráfico da barra de erros desses dados, selecione a caixa de diálogo da barra de erros clicando em **Graphs⇒Error Bar...** (Gráficos⇒Barra de Erros), e na caixa de diálogo (veja a Figura 7.1), clique em  (Resumos de variáveis separadas) e, depois, em  (Definir). Esse


Tabela 7.2 Dados do **spiderRM.sav**

Sujeito	Foto (Escore da Ansiedade)	Real (Escore da Ansiedade)
1	30	40
2	35	35
3	45	50
4	40	55
5	50	65
6	35	55
7	55	50
8	25	35
9	30	30
10	45	50
11	40	60
12	50	39

processo mostrará a caixa de diálogo principal. Uma vez que a caixa de diálogo for ativada, você deve selecionar as duas variáveis de interesse **picture** (figura) e **real** (verdadeira) e clicar em . O gráfico resultante deve ser idêntico ao que você obteve para os dados entre participantes. Isso é um problema porque eu perdi muito tempo lhe dizendo como o delineamento de medidas repetidas elimina muito da variância não-sistemática dos dados e, traçando os dados das medidas repetidas dessa maneira, nós ignoramos os componentes das medidas repetidas dos dados. Loftus & Masson (1994) sugerem que uma maneira de resolver esse problema é eliminar a variabilidade entre participantes normalizando a média dos participantes (isso significa assegurar que todos os participantes tenham a mesma média entre as condições). Para normalizar as médias, precisamos usar a função *compute* (calcular) do SPSS e executar vários passos (tudo isso foi apenas um ardid para você usar a função *calculate*). Essa função já foi vista na Seção 3.3.4.



Na Seção 3.3.4, descobrimos que o comando *compute* (calcular) permite executar várias funções em colunas de dados (isto é, adicionar ou multiplicar colunas). Antes de continuarmos, retorne a essa seção e refresque sua memória sobre as maravilhas dessa função!

7.3.2.1 Passo 1: Calcular a média para cada participante ②

Agora que você revisou o comando *compute* (calcular), podemos usá-lo para produzir diagramas de barra de erro dentre participantes mais precisos. Para começar, precisamos calcular a ansiedade média de cada participante e, assim, usamos a função *mean* (média). Acesse a caixa de diálogo *compute* (calcular) utilizando o caminho de menu **Transform⇒Compute...** (Transformar⇒Calcular...). Digite a palavra **mean** (média) no quadro *Target Variable* (Variável Alvo) e navegue na lista de funções até encontrar uma chamada de *MEAN* (*numexpr, numexpr, ...*). Destaque essa função e transfira-a para a área de comando clicando em . Quando ela for transferida irá aparecer na área de comando como “*MEAN*(?, ?)”, e os pontos de interrogação devem ser substituídos pelos nomes de variáveis (que podem ser digitados ou transferidos da lista de variáveis). Assim, substitua o primeiro ponto de interrogação pela variável **picture** (figura) e o segundo pela variável **real** (verdadeira). A caixa de diálogo completa deve ficar como a da Figura 7.4.

7.3.2.2 Passo 2: Calcular a média geral ②

A média geral é a média de todos os escores (independentemente de qual condição ele se origina), assim, para os dados atuais esse valor será a média de todos os 24 escores. Podemos calcular isso manualmente (isto é, somando to-

dos os escores e dividindo o resultado por 24); entretanto, uma maneira mais fácil é usar as médias que acabamos de calcular. Essas médias representam o escore médio de cada participante e, desse modo, se tomarmos a média desses escores médios teremos a média de todos os participantes (isto é, a média geral) – ufa, tinha muitas médias nessa frase! OK, para fazer isso podemos usar um pequeno e útil dispositivo chamado de comando *descriptives* (descritivas) (você pode também usar as funções *Explore* (Explorar) ou *Frequencies* (Frequências) que vimos no Capítulo 3, mas como eu já falei sobre elas, tentaremos algo diferente). Acesse a função *descriptives* (descritivas) utilizando o menu **Analyze⇒Descriptive Statistics⇒Descriptives...** (Analisar⇒Estatística Descritiva⇒Descritivas...). Uma caixa de diálogo semelhante a da Figura 7.5 deverá aparecer. O comando *descriptives* (descritivas) é usado para conseguir estatísticas descritivas básicas para variáveis, e clicando  (opções), uma segunda caixa de diálogo é ativada. Selecione a variável **mean** (média) da lista e transfira-a para a caixa rotulada de *Variable(s)* (variável) clicando em . Depois, use a caixa de diálogo *options* (opções) para especificar somente a média (você pode deixar os demais valores como estão porque estamos interessados apenas na média). Se você executar essa análise, a saída irá fornecer algumas estatísticas descritivas autoexplicativas para cada uma das três variáveis (presumindo que você tenha

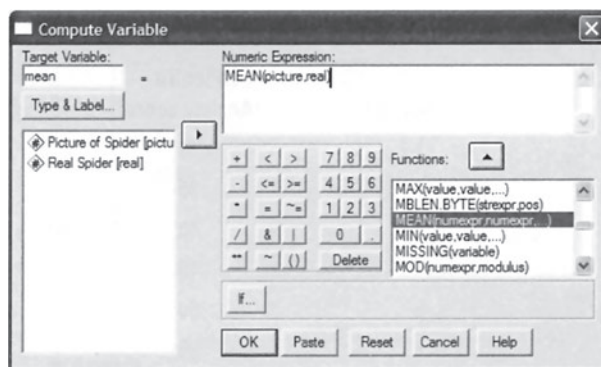


Figura 7.4 Usando a função *compute* (calcular) para determinar a média de duas colunas.

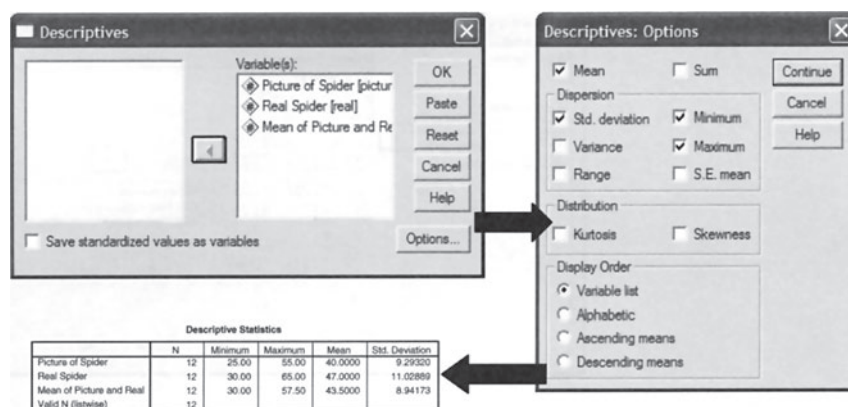


Figura 7.5 Caixas de diálogo e saídas para o procedimento estatísticas descritivas.

selecionado as três). Você deve observar que temos a média para a condição “figura da aranha” e a média para a condição “aranha real”, mas é na variável final que estamos realmente interessados: a média das condições “figura da aranha” e “aranha real”. A média dessa variável é a média geral, e você pode ver na tabela resumo que seu valor é 43,50. Nós iremos utilizar a média geral em cálculos posteriores.

7.3.2.3 Passo 3: Calcular o fator de ajustamento ②

Se você olhar a variável rotulada de **mean** (média), deverá notar que os valores para cada participante são diferentes, ou seja, algumas pessoas tiveram mais ansiedade do que outras ao longo das condições. O fato de que os escores da média dos participantes diferem representa diferenças individuais entre pessoas diferentes (isso representa o fato de que alguns participantes têm mais medo de aranhas do que outros). Essas diferenças na ansiedade natural a aranhas contaminam os gráficos de barras de erro, por isso, se não ajustarmos os valores que traçamos, teremos o mesmo gráfico que foi obtido quando o delineamento entre participantes foi usado. Loftus e Masson (1994) argumentam que para eliminar essa contaminação, devemos nivelar as médias entre os participantes (isto é, ajustar os escores em cada condição para que quando calcu-

larmos a média entre as condições ela seja a mesma para todos os participantes). Para fazer isso, precisamos calcular um fator de ajuste subtraindo o escore médio de cada participante da média geral. A função *compute* (calcular) pode fazer esses cálculos para nós. Ative a caixa de diálogo *compute* (calcular), dê um nome para a variável-alvo (sugiro **ajustar**) e use o comando (43.5 mean). Esse comando irá pegar a média geral (43,5) e subtrair dela o nível médio de ansiedade de cada participante (veja a Figura 7.6).

Esse processo cria uma nova variável no editor de dados chamada de **adjust** (ajuste). Os escores na coluna **adjust** representam as diferenças médias da ansiedade de cada participante e o nível médio de ansiedade entre todos os participantes. Note que alguns dos valores são positivos e representam os participantes menos ansiosos do que a média. Os participantes mais ansiosos do que a média têm escores ajustados negativos. Podemos utilizar agora esses valores ajustados para eliminar as diferenças de ansiedade entre os participantes.

7.3.2.4 Passo 4: Criar os valores ajustados para cada variável ②

Até agora, calculamos a diferença entre o escore médio de cada participante e o escore médio de todos os participantes (a média geral).

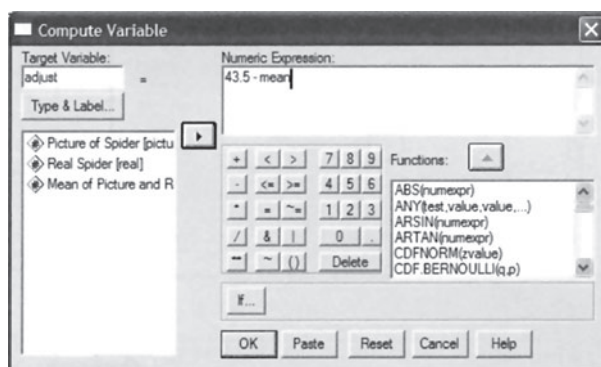


Figura 7.6 Calculando o fator de ajuste.

Essa diferença pode ser usada para ajustar os escores existentes de cada participante. Primeiro, precisamos ajustar os escores na condição **picture of spider** (figura da aranha). Mais uma vez, podemos usar o comando *compute* (calcular) para fazer o ajuste. Ative a caixa de diálogo *compute* da mesma forma que antes e, então, dê o nome para a nova variável de **picture2** (você pode, depois, clicar em *Type & Label...* (Tipo e Rótulo) e dar a essa variável um nome tal como “valores ajustados para a condição figura”). Iremos acrescentar o escore de cada participante na condição **picture of spider** para o seu valor ajustado. Selecione a variável **picture** (figura) e a transfira para a área de comando clicando em *→*. Depois, clique em *+*, selecione a variá-

vel **adjust** (ajuste) e a transfira para a área de comando clicando em *→*. A caixa de diálogo completa é mostrada na Figura 7.7. Faça o mesmo com a variável **real spider**: crie uma variável chamada de **real2** que contém os valores da **real** acrescidos ao valor da coluna **adjust**.

As variáveis **real2** e **picture2** (verdadeira2 e figura2) representam a ansiedade experimentada em cada condição ajustada de maneira a eliminar as diferenças entre os participantes. Se você não acredita em mim, use o comando *compute* (calcular) para criar a variável **mean2** (média2) que é a média entre **real2** e **picture2** (verdadeira2 e figura2) (como fizemos na Seção 7.3.2.1). Você deverá verificar que o valor nessa coluna é o mesmo para cada

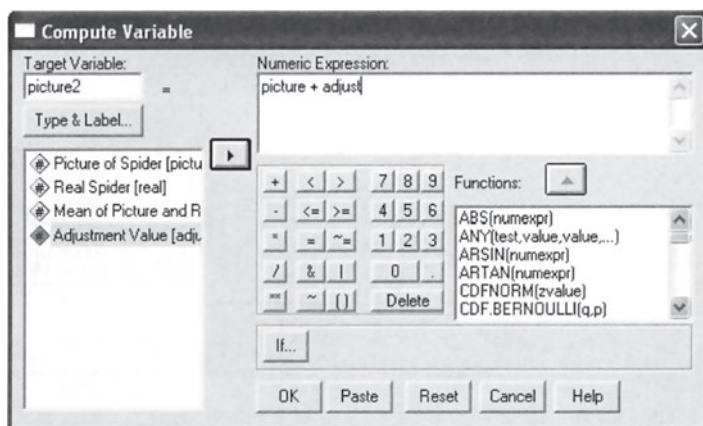


Figura 7.7 Ajustando os valores da condição **picture of spider**.

participante demonstrando que a variabilidade entre participantes nas médias desapareceu: o valor será 43,50 – a média geral.

7.3.2.5 Traçar o diagrama de barras de erros ②

Traçar o gráfico propriamente dito é semelhante ao cenário entre grupos. Primeiro, acesse a caixa de diálogo principal por meio dos menus **Graphs⇒ErrorBar...** (Gráficos⇒Barra de Erros...). Depois, nessa caixa de diálogo (veja a Figura 7.1) clique em **Summaries of separate variables** (Resumo para variáveis separadas) e em **Define** (Definir). Esse processo irá criar a caixa de diálogo principal. Uma vez que a caixa de diálogo for ativada, selecione as duas variáveis de interesse (queremos ajustar valores, assim, precisamos selecionar **real2** e **picture2** (verdadeira2 e figura2)) e, então, clique em **OK** (veja a Figura 7.8).

O gráfico de barras de erro resultante é apresentado na Figura 7.9. Você irá notar que o gráfico de barras de erro dos valores ajustados é um pouco diferente do gráfico de barras de erro dos valores não-ajustados (retorne à Figura 7.3). A primeira coisa a ser notada é que as barras de erro não se sobrepõem. Anteriormente falei que quando as barras de erro não se sobrepõem podemos estar confiantes de que nossas amostras não vieram da mesma popula-

ção (e, assim, nossa manipulação experimental foi um sucesso). Portanto, a versão do gráfico de medidas repetidas desses dados indica que aranhas verdadeiras podem evocar maiores graus de ansiedade (porque a média é alta) do que apenas uma a figura de uma aranha. Esse exemplo ilustra que conclusões diferentes podem ser tiradas se os dados foram coletados usando os mesmos ou diferentes participantes (falarei mais sobre isso na Seção 7.7).

7.4 TESTANDO DIFERENÇAS ENTRE MÉDIAS: O TESTE T ①

A forma mais simples de fazer um experimento é manipular uma só variável de duas maneiras apenas e medir somente uma saída. A manipulação da variável independente em geral envolve ter uma condição experimental e um grupo-controle (veja Field e Hole, 2003). Alguns exemplos de tipos de projetos são:

- O filme *Pânico 2* é mais assustador do que o *Pânico* original? Podemos medir os bati-

Qual é a diferença entre os testes t dependente e independente?

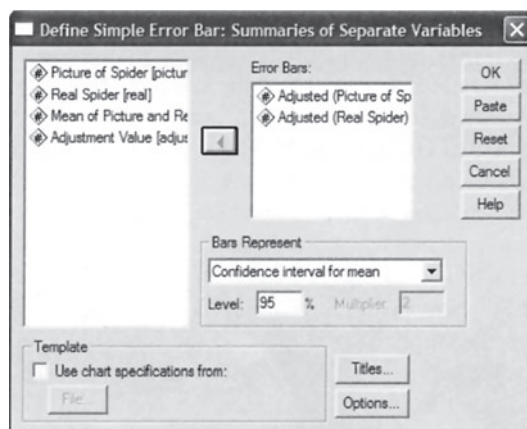


Figura 7.8 Caixa de diálogo para traçar um gráfico de barras de erro de uma variável de medidas repetidas.

te pelo acaso, grandes diferenças entre as médias amostrais não devem ocorrer com frequência. Sob o que é conhecido como *hipótese nula*, presumimos que a manipulação experimental não tenha efeito nos participantes: portanto, esperamos que as médias das amostras sejam similares.

- Comparamos a diferença entre as médias das amostras que coletamos com a diferença entre as médias das amostras que esperamos obter somente pelo acaso. Usamos o erro padrão (veja a Seção 1.6) como uma medida da variabilidade entre as médias das amostras. Se o erro padrão é pequeno, esperamos que a maioria das amostras tenham médias similares. Quando o erro padrão é grande, esperamos obter grandes diferenças nas médias das amostras somente pelo acaso. Se a diferença entre as amostras que coletamos for maior do que esperamos no erro padrão, podemos presumir uma de duas coisas:

- a) Que as médias das amostras na nossa população flutuam muito somente por acaso e temos, por acaso, coletado duas amostras que são atípicas da população de onde foram retiradas.
- b) As duas amostras vêm de populações diferentes, mas são típicas das suas respectivas populações originais. Nesse cenário, a diferença entre amostras representa uma diferença genuína entre as amostras (e, assim, a hipótese nula é incorreta).

- À medida que a diferença observada entre as médias das amostras aumenta mais confiante ficamos de que a segunda explicação seja a correta (isto é, que a hipótese nula deva ser rejeitada). Se a hipótese nula é incorreta, aceitamos o que é denominado de *hipótese experimental* (também chamada de *hipótese alternativa* ou algo semelhante), isto é, que as médias das duas amostras diferem por causa da manipulação diferenciada que foi imposta a cada uma das amostras.

Resumindo, calculamos o teste t usando a equação (7.1), mas a forma exata que essa

equação tomará vai depender de como os dados foram coletados (isto é, se os mesmos ou diferentes participantes foram usados em cada condição experimental).

$$t = \frac{\begin{array}{c} \text{diferença observada} \\ \text{entre as} \\ \text{médias} \\ \text{amostrais} \end{array} - \begin{array}{c} \text{diferença esperada} \\ \text{entre as médias} \\ \text{populacionais} \\ \text{(se a hipótese nula} \\ \text{for verdadeira)} \end{array}}{\begin{array}{c} \text{estimativa do erro padrão} \\ \text{da diferença entre as} \\ \text{médias de duas amostras} \end{array}} \quad (7.1)$$

7.4.2 Suposições do teste t ①

Tanto o teste t independente como o *dependente* são testes paramétricos baseados na distribuição normal (veja o Capítulo 3). Portanto, é assumido que:

- Os dados são de populações normalmente distribuídas.
- Os dados são medidos pelo menos em um nível intervalar.

O teste t independente, que é usado para testar diferentes grupos de pessoas, também assume que:

- As variâncias populacionais são iguais (*homogeneidade da variância*).
- Os escores são independentes (porque eles vêm de diferentes pessoas).

Essas hipóteses foram explicadas em detalhes no Capítulo 3, no qual enfatizei a necessidade de verificar essas suposições antes de realizar seu teste estatístico. Não vou falar sobre isso novamente, mas se você ignorou meu conselho e não verificou essas suposições, então precisa fazê-lo agora! O SPSS incorpora, também, procedimentos do teste t (por exemplo, o teste de Levene – veja a Seção 3.6 – pode ser feito ao mesmo tempo em que o teste t). Vamos agora analisar cada um dos testes t .

7.5 O TESTE t DEPENDENTE ①

Agora que sabemos como fazer gráficos de médias, podemos continuar a analisar se diferenças entre as médias dos grupos são esta-

tisticamente significativas. Se ficarmos com os nossos dados de medidas repetidas, podemos olhar para o teste t dependente, ou teste t de amostras emparelhadas. O teste t dependente é fácil de calcular. Usamos a versão numérica da equação (7.1):

$$t = \frac{\bar{D} - \mu_D}{S_D / \sqrt{N}} \quad (7.2)$$

A equação (7.2) compara a diferença média entre nossas amostras (\bar{D}) com a diferença que devemos esperar encontrar entre as médias populacionais (μ_D) e, então, leva em conta o erro padrão das diferenças (S_D/\sqrt{N}). Se a hipótese nula é verdadeira, esperamos que não haja diferenças entre as médias populacionais (isto é, $\mu_D = 0$).

7.5.1 Distribuições amostrais e o erro padrão ①

Na equação (7.1), me referi ao denominador da fração anterior como o erro padrão das diferenças. O erro padrão foi apresentado na Seção 1.6, e é simplesmente o desvio padrão da *distribuição amostral*. Dê uma olhada nessa seção agora para refrescar sua memória sobre distribuições amostrais e erro padrão. As distribuições amostrais têm muitas propriedades importantes. Para começar, se a população é distribuída normalmente, então o mesmo ocorre com a distribuição amostral; na verdade, se as amostras contêm mais do que 50 escores, a distribuição amostral será sempre distribuída normalmente. A média da distribuição amostral é igual à média da população – assim, se você calculou a média de todas as médias das amostras, o valor que você obtém deve ser o mesmo da média da população. Essa propriedade faz sentido porque se uma amostra é representativa da população, você deve esperar que sua média seja igual à da população. Entretanto, algumas vezes amostras não são representativas e sua média difere da média da população. Em geral, entretanto, a média da amostra estará bem próxima da média da população e raramente a média da amostra será substancialmente diferente daquela da população. Finalmente, outra

propriedade da distribuição amostral é que seu desvio padrão é igual ao desvio padrão da população dividido pela raiz quadrada do número de observações na amostra. Como mencionei anteriormente, esse desvio padrão é conhecido como o erro padrão.

Podemos estender essa ideia olhando para as *diferenças* entre as médias das amostras. Se você fosse tirar vários pares de amostras da população e calcular suas médias, poderia, também, calcular a diferença entre suas médias. Mencionei anteriormente que em geral as médias das amostras serão muito similares à média da população, portanto, a maioria das amostras terá médias muito similares. Desse modo, na maior parte do tempo, a diferença entre as médias das amostras de uma mesma população será zero ou próxima de zero. Entretanto, algumas vezes uma ou ambas as amostras podem ter um média muito afastada da média da população e, assim, é possível obter uma grande diferença entre as médias das amostras somente por acaso. Entretanto, isso acontece com menos frequência.

Na verdade, se traçar essas diferenças entre as médias das amostras como um histograma você terá, novamente, uma distribuição amostral com todas as propriedades previamente descritas. O desvio padrão dessa distribuição amostral é chamado de *erro padrão das diferenças*. Um erro padrão pequeno indica que a maioria dos pares das amostras de uma população terá médias muito parecidas (isto é, a diferença entre as médias da amostra será frequentemente muito pequena). Um erro padrão grande indica que as médias das amostras podem se desviar muito da média da população e, assim, diferenças entre pares de amostras podem ser bem grandes somente por acaso.

7.5.2 A equação do teste t dependente explicada ①

Em um experimento, o escore de uma pessoa na condição 1 será diferente do seu escore na condição 2, e essa diferença pode ser muito grande ou muito pequena. Se calculássemos as diferenças entre o escore de cada pessoa em cada condição e somássemos essas diferenças,



teríamos o total das diferenças. Se, depois, dividíssemos essa diferença pelo número de participantes, teríamos a diferença média (ou seja, quanto, em média, o escore de uma pessoa diferiu na condição 1

comparado com a condição 2). Essa diferença média é (\bar{D}) na equação (7.2) e é um indicador da variação sistemática nos dados (isto é, ele representa o efeito experimental). Se tivéssemos tirado duas amostras aleatórias da população (e nada feito com essas amostras), as médias poderiam ser diferentes somente por acaso. Como já vimos, o grau no qual duas médias da amostra provavelmente diferem é determinado pelo tamanho do erro padrão das diferenças. Precisamos ter certeza de que a diferença observada entre nossas amostras é devido à manipulação experimental (e não resultado do acaso). Saber somente a diferença média não é útil porque ela depende da escala da medida, portanto, padronizamos o valor. Uma maneira de padronizar a soma da diferença dos grupos seria dividi-la pelo desvio padrão da amostra dessas diferenças (veja a Seção 1.4.1). O desvio padrão representa a média dos desvios a partir da média, assim, o desvio padrão das diferenças entre condições representa o desvio médio da diferença das médias. Como tal, o desvio padrão é uma medida de quanta variação existe entre a diferença dos escores dos participantes. Desse modo, o desvio padrão dessas diferenças representa a variação *não-sistemática* no experimento.

Esclarecendo, imagine que um alienígena (quando eu estava paralisado de medo de uma aranha) me clonou 12 vezes (que os céus não permitam que existam 12 iguais a mim no planeta!). Todos os meus clones seriam idênticos a mim e se comportariam do mesmo modo que eu. Portanto, todos nós estaríamos amedrontados com a figura da aranha e teríamos um escore de ansiedade de 40. Ao ver aranhas verdadeiras, todos molharíamos nossa roupa de baixo e teríamos escores de ansiedade de 50. Lembre-se de que somos clones, portanto,

temos um comportamento idêntico. Se você calcular a diferença entre os escores da nossa ansiedade para a figura e a aranha verdadeira, a diferença será de 10 pontos para cada um de nós. Se você calcular o desvio padrão dessas diferenças o resultado será 0 (porque todos nós temos os mesmos escores, não há variação). Como nossos participantes são os mesmos, o desvio padrão é 0; isto é, não há variação não-sistemática. Em outras palavras, todas as diferenças em ansiedade resultam da observação de aranhas verdadeiras em vez de figuras de aranhas e nenhuma das diferenças pode ser explicada por outros fatores (como diferenças individuais).

Embora dividir pelo desvio padrão seja útil para padronizar a diferença média entre condições, de fato, estamos interessados em saber como a diferença entre as médias das amostras se compara com a que esperaríamos obter se não tivéssemos feito uma manipulação experimental. Podemos usar as propriedades da distribuição amostral: em vez de dividir a diferença média entre condições pelo desvio padrão das diferenças, podemos dividi-la pelo erro padrão das diferenças. Dividir pelo erro padrão não somente padroniza a diferença média entre condições, mas também informa como a diferença entre as duas médias das amostras se compara em magnitude ao que se esperaria somente por acaso. Se o erro padrão é grande, grandes diferenças entre amostras são mais comuns (porque a distribuição entre as diferenças é mais espalhada). Já se o erro padrão é pequeno, grandes diferenças entre as médias das amostras são incomuns (porque a distribuição está mais concentrada em torno do zero). Portanto, se a diferença média entre nossas amostras é grande e o erro padrão das diferenças é pequeno, podemos estar confiantes de que a diferença que observamos na nossa amostra não é um resultado casual, ou seja, ela deve ter sido provocada pela manipulação experimental.

Em um mundo perfeito, calcularíamos o erro padrão retirando todos os pares possíveis de amostras da população, descobrindo as diferenças entre suas médias e determinando o desvio padrão dessas diferenças. Entretanto,

na realidade isso é impossível. Portanto, estimamos o erro padrão do desvio padrão das diferenças obtido da amostra (S_D) e do tamanho da amostra (N). Na Seção 1.6, vimos que o erro padrão é simplesmente o desvio padrão dividido pela raiz quadrada do tamanho da amostra; da mesma forma, o erro padrão das diferenças $\sigma_{\bar{D}}$ é o desvio padrão das diferenças dividido pela raiz quadrada do tamanho da amostra:

$$\sigma_{\bar{D}} = \frac{S_D}{\sqrt{N}}$$

Se o erro padrão das diferenças é uma medida da variação não-sistemática dos dados e a soma das diferenças dos escores representa a variação sistemática, deveria ficar claro que **a estatística t é simplesmente a razão da variação sistemática no experimento para a variação não-sistemática**. Se a manipulação experimental criar qualquer tipo de efeito, a variação sistemática será bem maior do que a variação não-sistemática (o t será, no mínimo, maior do que 1). Se a manipulação experimental não for bem-sucedida, a variação causada pelas diferenças individuais será bem maior do que a causada pelo experimento (assim, o t será menor do que 1). Podemos comparar o

valor obtido para o t com o valor máximo que esperamos conseguir somente por acaso numa distribuição t com os mesmos graus de liberdade (esses valores podem ser encontrados no Apêndice A.2); se o valor que obtivermos exceder esse valor crítico, isso reflete um efeito da nossa variável independente.

7.5.3 O teste t dependente utilizando o SPSS ①

Nos nossos dados sobre aranhas (**spiderRM.sav**), temos 12 aracnofóbicos que foram expostos a uma figura de aranha **picture** (figura) e, numa outra ocasião, foram expostos a uma aranha verdadeira **real** (verdadeira). A sua ansiedade foi medida em cada condição (metade dos participantes foi exposta à figura antes da aranha verdadeira enquanto a outra metade foi exposta à aranha verdadeira primeiro). Já descrevi como os dados foram organizados e, assim, podemos seguir adiante e fazer o teste propriamente dito. Primeiro precisamos acessar a caixa de diálogo principal usando o menu **Analyze**⇒**Compare Means**⇒**Paired-Samples T Test...** (Analisar⇒Comparar Médias⇒Teste t para Amostras Emparelhadas...) (Figura 7.10). Uma vez ativada a caixa de diálogo, selecione duas variáveis da lista (clique

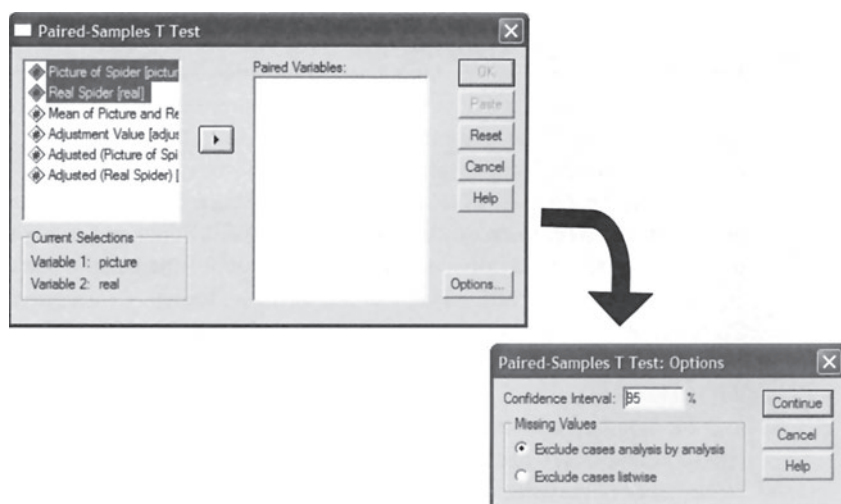





Figura 7.10 Caixa de diálogo principal para o teste t pareado.

na primeira variável com o *mouse*, depois na segunda e note que os nomes das variáveis aparecem na caixa denominada de *Current Selections* (Seleções atuais)) e as transfira para a caixa denominada *Paired Variables* (Variáveis Pareadas) clicando em . Se você quiser fazer vários testes *t*, pode selecionar outro par de variáveis, transferi-las para a lista de variáveis, depois selecionar outro par, e assim por diante. Nesse caso, queremos executar somente um teste. Se você clicar em  (opções), aparece outra caixa de diálogo que oferece a opção de mudar a amplitude do intervalo de confiança que é calculado. A configuração padrão é para um intervalo de confiança de 95% e isso é adequado. Entretanto, se você quisesse ser mais rigoroso sobre sua análise, poderia ter escolhido um intervalo de confiança de 99%, mas assim você correria um risco maior de não conseguir detectar um efeito genuíno (um erro do Tipo II). Para executar a análise, clique em .

7.5.4 Saídas do teste *t* dependente ①

A saída resultante produz três tabelas. A saída do SPSS 7.1 mostra uma tabela de estatísticas resumo para as duas condições experimentais. Para cada condição, sabemos a média, o número de participantes (*N*) e o desvio padrão da amostra. A coluna final informa o erro padrão (veja a Seção 7.5.1), que é o desvio padrão da amostra dividido pela raiz quadrada do tamanho da amostra ($EP = s/\sqrt{N}$), assim, para a condição da figura $EP = 9,2932/\sqrt{12} = 9,2932/3,4641 = 2,68$.

A Saída do SPSS 7.1 também mostra a correlação de Pearson entre as duas condições. Quando medidas repetidas são usadas, é possível que as condições experimentais se correlacionem (porque os dados em cada condição vêm das mesmas pessoas e, assim, pode haver consistência em suas respostas). O SPSS dá o valor do *r* de Pearson e a significância bilateral (veja o Capítulo 4). Para esses dados, as condições experimentais produzem um coeficiente de correlação bem grande ($r = 0,545$), mas não estão correlacionadas de forma significativa porque $p > 0,05$.

A Saída do SPSS 7.2 mostra a tabela mais importante: aquela que nos diz se a diferença entre as médias das duas condições foi grande o suficiente para *não* ser um resultado ao acaso. Primeiro a tabela fornece a diferença da média entre escores (esse valor $-\bar{D}$ na equação (7.2) – é a diferença entre os escores médios de cada condição: $40 - 47 = -7$). A tabela também apresenta o desvio padrão das diferenças entre as médias e, mais importante, o erro padrão das diferenças entre os escores dos participantes em cada condição (veja a Seção 7.5.1). A estatística teste *t* é calculada dividindo-se a média das diferenças pelo erro padrão das diferenças (veja a equação (7.2): $t = -7/2,8311 = -2,47$). O tamanho do *t* é comparado contra valores conhecidos com base nos graus de liberdade. Quando os mesmos participantes foram usados, os graus de liberdade são o tamanho da amostra menos 1 ($gl = N - 1 = 11$). O SPSS usa os graus de liberdade para calcular a probabilidade exata de que um valor de *t* tão grande quanto o obtido possa ocorrer.

Saída do SPSS 7.1

Paired Sample Statistics
(Estatísticas Amostrais Pareadas)

		Mean (Média)	N	Std. Deviation (Desvio Padrão)	Std. Error Mean (Erro Padrão da Média)
Pair 1 (Par 1)	Picture of Spider (Figura da Aranha)	40.0000	12	9.2932	2.6827
	Real Spider (Aranha Verdadeira)	47.0000	12	11.0289	3.1838

Paired Sample Correlations
(Correlações Amostrais Pareadas)

		N	Correlation (Correlação)	Sig.
Pair 1 (Par 1)	Picture of Spider (Figura da Aranha) Real Spider (Aranha Verdadeira)	12	0.545	0.067

Saída do SPSS 7.2

Paired Samples Test (Testes Amostrais Pareados)

		Paired Differences (Diferenças pareadas)					t	df (gl)	Sig. (2-tailed) (Sig. bilateral)
		Mean (Média)	Std. Deviation (Desvio Padrão)	Std. Error Mean (Erro Padrão da Média)	95% confidence interval of the differences (Intervalo de confiança de 95% das diferenças)				
					Lower (Inferior)	Upper (Superior)			
Pair 1 (Par 1)	Picture of Spider (Figura da Ara- nha) Real Spider (Ara- nha Verdadeira)	-7.0000	9.8072	2.8311	-13.2312	-0.7688	-2.473	11	0.031

rer por acaso. Essa probabilidade está na coluna chamada de Sig. (Significância). Por padrão, o SPSS fornece a probabilidade bilateral, que é a probabilidade quando nenhuma previsão foi feita sobre a direção das diferenças entre os grupos. Se uma previsão específica foi feita (por exemplo, podemos prever que a ansiedade será maior quando uma aranha verdadeira é usada), então a probabilidade unilateral deve ser informada, e esse valor é obtido dividindo-se a probabilidade bilateral por 2. A probabilidade bilateral para os dados da aranha é muito baixa ($p = 0,031$); ela nos diz que há somente uma chance de 3,1% de que um valor de t desse tamanho possa ter ocorrido devido ao acaso. Como cientistas sociais, aceitamos como estatisticamente significativo tudo o que tem menos de 5% de chance de ocorrer por acaso. O fato de o valor de t ser um número negativo nos diz que a condição 1 (a condição **picture of spider** (figura da aranha)) tem uma média menor do que a segunda (a condição **real spider** (aranha verdadeira)) e, assim, a aranha verdadeira provoca uma ansiedade maior do que a da figura. Quando apresentamos um teste t , sempre incluímos os graus de liberdade (nesse caso, 11), o valor da estatística t e o nível no qual esse valor é significativo. Portanto, podemos concluir que a exposição à aranha verdadeira causou mais ansiedade em aracnofóbicos do que a exposição a uma figura ($t(11) = -2,47$, $p < 0,05$). Esse resultado foi previsto pelo gráfico de barra de erros apresentado na Figura 7.9.

A última informação que essa saída nos fornece é um intervalo de confiança de 95% para a diferença média. Imagine que tomamos 100 amostras de uma população de diferentes escores, calculamos suas médias (\bar{D}) e um intervalo de confiança para essa média. Em 95 das amostras, o intervalo de confiança construído contém o valor real da diferença média. O intervalo de confiança fornece os limites dentro dos quais é mais provável que a diferença real média esteja³. Assim, assumindo que o intervalo de confiança da amostra é um dos 95 dos 100 que contém o valor da população, podemos dizer que a diferença média verdadeira está entre $-13,23$ e $-0,77$. A importância desse intervalo é que ele não contém o zero (isto é, ambos os limites são negativos); isso nos diz que o valor verdadeiro da diferença média dificilmente será zero. Se comparássemos pares de amostras aleatórias de uma população, esperaríamos que a maioria das diferenças entre as médias das

³ Vimos na Seção 1.6.2 que esses intervalos representam o valor de dois (1,96, precisamente) erros padrão para cada lado da média da distribuição amostral. Para esses dados em que a diferença média foi -7 e o erro padrão foi de 2,8311, esses limites serão de $-7 \pm (1,96 \times 2,8311)$. Contudo, em virtude de estarmos utilizando a distribuição t e não a distribuição normal usaremos um valor crítico da distribuição t para calcular os intervalos de confiança. Esse valor é (com 11 graus de liberdade como nesse exemplo) 2,201 (bilateral), que fornece um intervalo de $-7 \pm (2,201 \times 2,8311)$.

amostras fosse zero. Esse intervalo nos diz que, com base nas nossas duas amostras, é pouco provável que o valor verdadeiro entre as médias seja zero. Portanto, podemos estar confiantes de que nossas duas amostras não representam amostras aleatórias da mesma população. Em vez disso, elas representam amostras de populações diferentes induzidas pela manipulação experimental.

7.5.5 Calculando o tamanho de efeito ②

Embora nosso t estatístico seja estatisticamente significativo, isso não quer dizer que nosso efeito seja importante em termos práticos. Para descobrir se o efeito é importante, precisamos usar o que sabemos sobre o tamanho de efeito (veja a Seção 1.8.4). Ficarei com o tamanho de efeito r porque ele é amplamente entendido, frequentemente utilizado e sim, admito, eu realmente gosto dele! Converter um valor t em um valor r é muito fácil; usamos a seguinte equação (de Rosenthal, 1991, p.19 e Rosnow e Rosenthal, 2005, p.328):

$$r = \sqrt{\frac{t^2}{t^2 + gl}}$$

Sabemos os valores de t e do gl da saída do SPSS, assim, podemos calcular r da seguinte maneira:

$$\begin{aligned} r &= \sqrt{\frac{2,473^2}{-2,473^2 + 11}} \\ &= \sqrt{\frac{6,116}{17,116}} \\ &= 0,60 \end{aligned}$$

Se você lembrar os nossos pontos de referência para os tamanhos de efeito, isso representa um efeito muito grande (ele está acima de 0,5, o limite para um efeito grande). Portanto, além de estatisticamente significativo, esse efeito também é grande; desse modo, representa uma descoberta significativa.

7.5.6 Relatando o teste t dependente ①

Existe uma maneira padrão para relatar qualquer estatística teste: você declara a descoberta com a qual o teste se relaciona, e depois relata a estatística teste, seus graus de liberdade e o valor de probabilidade dessa estatística teste. A Associação Americana de Psicologia, entre outras associações, recomenda que uma estimativa do tamanho de efeito seja rotineiramente relatada. Mesmo que os tamanhos de efeito sejam usados esporadicamente, quero que você tenha bons hábitos; portanto, vamos começar a pensar em tamanhos de efeito agora. Nesse exemplo, a saída do SPSS indica que o valor do t era $-2,47$, que os graus de liberdade nos quais isso foi baseado era 11 e que era significativo em $p = 0,031$. Também é possível ver as médias para cada grupo. Podemos escrever isso como:

- ✓ Na média, os participantes experienciaram maior ansiedade com aranhas verdadeiras ($M = 47,00$, $EP = 3,18$) do que com figuras de aranhas ($M = 40,00$, $EP = 2,68$, $t(11) = -2,43$, $p < 0,05$, $r = 0,60$).

Observe como relatamos as médias em cada grupo (e erros padrão) no formato padrão. Para a estatística teste, note que usamos um t em itálico a fim de representar o fato de que calculamos uma estatística t , e entre parênteses colocamos os graus de liberdade e apresentamos o valor da estatística teste. A probabilidade pode ser expressa de várias maneiras: geralmente, relatamos pesquisas em um nível padrão de significância (como 0,05), conforme fizemos aqui, mas, às vezes, podemos relatar a significância exata. Finalmente, note que relatei o tamanho de efeito no final – dificilmente você verá isso em artigos publicados, mas deve relatá-lo mesmo assim!

Evite escrever de forma vaga, como:

- ✗ As pessoas estavam mais assustadas com as aranhas verdadeiras ($t = -2,47$)

Mais assustadas em comparação a quê? Onde estão os gl ? O resultado foi estatisticamente significativo? O efeito era importante (qual é o tamanho de efeito?)

Dica da Samanta Ferrinho



- O teste t dependente compara duas médias quando essas médias vieram do mesmo grupo de pessoas; por exemplo, se você tem duas condições experimentais e usou os mesmos participantes em cada condição.
- Veja a coluna rotulada de *Sig.* Se o valor for menor do que 0,05, as médias dos dois grupos são significativamente diferentes.
- Veja os valores das médias para informar como os grupos diferem.
- O SPSS fornece somente a significância bilateral; se você quiser a significância unilateral, basta dividir o valor por 2.
- Relate a estatística t , os graus de liberdade e o valor da significância. Relate, também, as médias e os seus erros padrão correspondentes (ou faça um gráfico de barras de erro).
- Se você for corajoso, calcule o tamanho de efeito e relate isso, também!

7.6 O TESTE t INDEPENDENTE ①

7.6.1 A equação do teste t independente explicada ①

O teste t independente é usado em situações nas quais existem duas condições experimentais e participantes diferentes foram usados em cada uma. Duas equações diferentes podem ser utilizadas para calcular a estatística t dependendo se as amostras contêm ou não um número igual de participantes. Assim como com o teste t dependente, podemos calcular a estatística t usando uma versão numérica da equação (7.1). Com o teste t dependente, poderíamos calcular as diferenças entre os pares de escores porque eles vinham dos mesmos participantes e, assim, diferenças individuais entre condições eram eliminadas. Dessa maneira, a diferença no escore refletia somente o efeito da manipulação experimental. Já quando participantes diferentes representam condições diferentes, os pares dos escores irão diferir não somente por causa da manipulação experimental, mas também devido a outras fontes de variação (como diferenças individuais entre a motivação dos participantes, QI, etc.). Se não podemos investigar as diferenças entre as condições em uma base *por participante* (comparando pares de escores como fizemos com o teste t dependente), devemos fazer comparações em bases *por condição* (olhando para o efeito geral na condição – veja a equação (7.3)):

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\text{Estimativa do erro padrão}} \quad (7.3)$$

Em vez de observar as diferenças entre pares de escores, agora olhamos para as diferenças entre as médias gerais das duas amostras e as comparamos com as diferenças que esperamos conseguir entre as médias das duas populações de onde as amostras vieram. Se a hipótese nula é verdadeira, as amostras foram retiradas da mesma população. Portanto, sob a hipótese nula $\mu_1 = \mu_2$ e $\mu_1 - \mu_2 = 0$. Assim, a equação fica:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{Estimativa do erro padrão}} \quad (7.4)$$

No teste t dependente, dividimos a diferença média entre pares de escores pelo erro padrão dessas diferenças. No teste t independente, estamos olhando para as diferenças entre grupos, assim, precisamos dividir pelo desvio padrão das diferenças entre os grupos. Também podemos aplicar a lógica das distribuições amostrais nessa situação. Imagine que pegamos vários pares de amostras – cada par contendo uma amostra de cada uma das duas diferentes populações e comparamos as médias dessas amostras. Do que aprendemos sobre distribuições amostrais, sabemos que a maioria das amostras de uma população terá médias muito parecidas. Portanto, se pegarmos vários pares de amostras (de popu-

lações diferentes), as diferenças entre as médias das amostras serão semelhantes entre os pares. Entretanto, geralmente a diferença entre as médias de um par de amostras irá se desviar por um pequeno valor, e, muito ocasionalmente, irá desviar por um valor maior. Se pudéssemos traçar uma distribuição amostral das diferenças entre cada par de médias amostrais que pudéssemos obter de duas populações, descobriríamos que ela teria uma distribuição normal com uma média igual à diferença entre as médias das populações ($\mu_1 - \mu_2$). A distribuição amostral nos informaria o quanto podemos esperar que as médias de duas (ou mais) amostras difiram. Como anteriormente, o desvio padrão da distribuição amostral (o erro padrão) indica quão variáveis serão as diferenças entre as médias das amostras apenas como consequência do acaso. Se o desvio padrão é alto, grandes diferenças entre as médias das amostras podem ocorrer por acaso; se é pequeno, somente pequenas diferenças entre as médias das amostras são esperadas. Portanto, faz sentido usar o erro padrão da distribuição amostral para avaliar se a diferença entre as duas médias das amostras é estatisticamente significativa ou simplesmente fruto do acaso. Especificamente, dividimos as diferenças entre as médias das amostras pelo desvio padrão da distribuição amostral.

Mas como obtemos o desvio padrão da distribuição amostral das diferenças entre médias amostrais? Bem, utilizamos a *propriedade da soma das variâncias*, a qual afirma que a variância da diferença entre duas variáveis independentes é igual à soma das suas variâncias (veja Howell, 2002, Capítulo 7). Essa afirmação significa que a variância da distribuição amostral é igual à soma das variâncias das duas populações de onde as amostras foram coletadas. Vimos anteriormente que o erro padrão é o desvio padrão da distribuição amostral de uma população. Podemos usar os desvios padrão amostrais para calcular o erro padrão da distribuição amostral de cada população:

$$\begin{array}{l} \text{EP da distribuição} \\ \text{amostral da primeira} \\ \text{população} \end{array} = \frac{s_1}{\sqrt{N_1}}$$

$$\begin{array}{l} \text{EP da distribuição} \\ \text{amostral da segunda} \\ \text{população} \end{array} = \frac{s_2}{\sqrt{N_2}}$$

Portanto, lembrando que a variância é simplesmente o desvio padrão ao quadrado, podemos calcular a variância de cada distribuição amostral:

$$\begin{array}{l} \text{Variância da distribuição} \\ \text{amostral da primeira} \\ \text{população} \end{array} = \left(\frac{s_1}{\sqrt{N_1}} \right)^2 = \frac{s_1^2}{N_1}$$

$$\begin{array}{l} \text{Variância da distribuição} \\ \text{amostral da segunda} \\ \text{população} \end{array} = \left(\frac{s_2}{\sqrt{N_2}} \right)^2 = \frac{s_2^2}{N_2}$$

A lei da soma das variâncias significa que para encontrar a variância da distribuição amostral das diferenças simplesmente somamos as variâncias das distribuições amostrais das duas populações:

$$\begin{array}{l} \text{Variância da distribuição} \\ \text{amostral das diferenças} \end{array} = \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}$$

Para encontrar o erro padrão da distribuição amostral das diferenças, tiramos a raiz quadrada da variância (porque a variância é o desvio padrão ao quadrado):

$$\begin{array}{l} \text{EP da distribuição} \\ \text{amostral das diferenças} \end{array} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

Portanto, a equação (7.4) se torna a equação (7.5):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \quad (7.5)$$

A equação (7.5) é verdadeira somente quando os tamanhos da amostra são iguais. Em geral, nas ciências sociais não é possível coletar amostras de mesmo tamanho (porque, por exemplo, nem todas as pessoas completam um experimento). Quando queremos comparar dois grupos que contêm números diferentes de participantes, a equação (7.5) não é apropriada. Em vez disso, usamos a va-

riância combinada, que leva em conta as diferenças no tamanho das amostras *ponderando* a variância de cada amostra. Intuitivamente, você pode perceber que as amostras grandes são melhores do que as pequenas porque elas se aproximam mais da população; portanto, ponderamos as variâncias pelo tamanho da amostra na qual ela é baseada (na verdade ponderamos pelos graus de liberdade, que são o tamanho da amostra menos 1). Portanto, uma estimativa da variância combinada é:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Isso é simplesmente uma média ponderada na qual cada variância é multiplicada (ponderada) pelos seus graus de liberdade e, depois, dividida pela soma das ponderações (ou soma dos dois graus de liberdade). A média da variância ponderada resultante é, então, substituída na equação do teste *t*:


$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Assim como o teste *t* dependente, podemos comparar o valor obtido de *t* contra o valor máximo que esperamos conseguir por acaso numa distribuição *t* com os mesmos graus de liberdade (esses valores podem ser encontrados no Apêndice A.2); se o valor que obtemos exceder esse valor crítico, isso reflete um efeito na nossa variável independente.

A derivação da estatística *t* serve apenas para fornecer um conhecimento conceitual do que estamos fazendo quando executamos um teste *t* no SPSS. Portanto, se você não sabe o que eu estou falando, não se preocupe (apenas pense no meu gato: ele tem que ouvir essa baboseira o tempo todo!), pois o SPSS sabe como fazê-lo e isso basta!

7.6.2 O teste *t* independente utilizando o SPSS ①

Equações são entediantes, e o SPSS foi inventado justamente para diminuir nosso con-

tato com elas. Usando nossos dados da aranha novamente (**spiderBG.sav**), temos 12 aracnofóbicos que foram expostos a uma figura de aranha e 12 outros aracnofóbicos expostos a uma tarântula verdadeira (os grupos foram codificados usando a variável **group** (grupo)). Sua ansiedade foi medida em cada condição **anxiety** (ansiedade). Já descrevi como os dados foram organizados (veja a Seção 7.3.1), assim, podemos seguir direto para o teste propriamente dito. Primeiro, precisamos acessar a caixa de diálogo principal usando o menu **Analyze⇒Compare Means⇒Independent Samples T Test...** (Analisar⇒Comparar médias⇒Teste *t* para amostras independentes) (veja a Figura 7.11). Uma vez que a caixa de diálogo for ativada, selecione a variável dependente da lista (clique em **anxiety** (ansiedade)) e transfira-a para a caixa rotulada de **Test Variable(s)** (Variáveis do teste) clicando em . Se você quiser executar testes *t* em diversas variáveis dependentes, pode selecionar outras variáveis dependentes e transferi-las para a lista de variáveis. Entretanto, não é recomendável executar muitos testes (veja o Capítulo 8).

A seguir, precisamos selecionar uma variável independente (a variável de grupo). Nesse caso, temos que selecionar **group** (grupo) e transferi-lo para a caixa rotulada de **Grouping Variable** (Variável de agrupamento). Quando sua variável de grupo for selecionada, o botão **Define Groups...** será ativado; clique nele para ativar a caixa de diálogo **define groups** (definir grupos). O SPSS precisa saber quais os códigos numéricos que você atribuiu para os seus dois grupos e há um espaço para você digitar esses códigos. Nesse exemplo, codificamos o grupo que mostra a figura da aranha como 0 e o grupo que mostra a aranha verdadeira como 1, portanto, esses são os códigos que precisamos informar. Como alternativa você pode especificar um **Ponto de Corte** (**Cut point**); nesse caso, o SPSS irá atribuir todos os casos maiores ou iguais àquele valor para um grupo e todos os valores abaixo do ponto de corte a um segundo grupo. Esse recurso é útil se você estiver testando grupos diferentes de participantes com base em

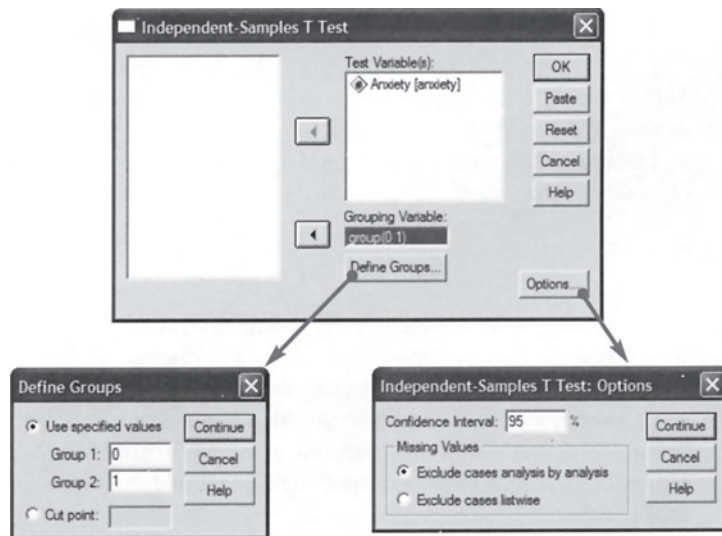


Figura 7.11 Caixas de diálogo para o teste *t* de duas médias independentes.

algo como a divisão pela mediana. Você deve classificar as pessoas como aracnofóbicas ou não, medindo seu escore num questionário sobre fobia de aranhas e calculando a mediana. Qualquer pessoa com um escore acima da mediana é classificado como com fobia, e aqueles abaixo da mediana, como sem fobia. Em vez de registrar todos os seus participantes e criar uma variável código, você simplesmente digita o valor da mediana na caixa rotulada de **Cut Point** (Ponto de Corte). Quando você tiver definido os grupos, clique em **Continue** (continuar) para retornar à caixa de diálogo principal. Se você clicar em **Options...** (Opções), outra caixa de diálogo aparece e fornece as mesmas opções oferecidas para o teste *t* independente. Para executar a análise, clique em **OK**.

7.6.3 Saídas do teste *t* independente ①

A saída para o teste *t* independente contém somente duas tabelas. A primeira tabela (a saída do SPSS 7.3) fornece um resumo estatístico para as duas condições experimentais. Dessa tabela podemos ver que ambos os grupos tinham 12 participantes (coluna rotulada de *N*). O grupo que viu a figura da aranha teve uma ansiedade média de 40 com um desvio padrão de 9,29. E mais, o erro padrão daquele grupo (o desvio padrão da distribuição amostral) é de 2,68 ($SE = 9,2932/\sqrt{12} = 9,2932/3,46431 = 2,68$). Além disso, a tabela indica que o nível médio de ansiedade dos participantes aos quais foi mostrada uma aranha verdadeira era de 47, com um desvio padrão de 11,03 e um erro padrão de 3,18 ($SE = 11,03/\sqrt{12} = 11,03/3,4641 = 3,18$).

Saída do SPSS 7.3

Group Statistics (Estatísticas dos grupos)

<i>Condition</i> (Condição)		<i>N</i>	<i>Mean</i> (Média)	<i>Std. Error</i> (Erro Padrão)	<i>Std. Error Mean</i> (Erro Padrão da Média)
<i>Anxiety</i> (Ansiedade)	<i>Picture of Spider</i> (Figura da Aranha)	12	40.0000	9.2932	2.6827
	<i>Real Spider</i> (Aranha Verdadeira)	12	47.0000	11.0289	3.1838

A segunda tabela da saída (Saída do SPSS 7.4) contém a estatística teste principal. Note que há duas filas contendo valores para a estatística teste: um fila é rotulada de *Variâncias iguais assumidas*, e a outra é rotulada de *Variâncias iguais não assumidas*. No Capítulo 3, vimos que os testes paramétricos supõem que as variâncias em grupos-experimentais são aproximadamente iguais. Bem, na verdade existem ajustes que podem ser feitos em situações nas quais as variâncias não são iguais. As filas da tabela relatam se essa suposição foi quebrada ou não. Como sabemos se essa suposição foi quebrada ou não?

Podemos apenas olhar aos valores da variância e ver se eles são similares (por exemplo, sabemos que os desvios padrão dos dois grupos são 9,29 e 11,03; se elevarmos ao quadrado esses valores teremos variâncias de 86,30 e 121,66). Entretanto, essa medida seria muito subjetiva: “Oh, veja, a variância no grupo 1 é somente 3000 vezes maior do que a variância no grupo 2: isso é praticamente igual”. Felizmente, existe um teste que pode ser feito para ver

se as variâncias são diferentes o suficiente para causar preocupação. O teste de Levene (veja a Seção 3.6) é similar ao teste *t*, pois ele testa a hipótese de que as variâncias nos dois grupos são iguais (isto é, a diferença entre as variâncias é zero). Portanto, se o teste de Levene é significativo em $p < 0,05$, podemos concluir que a hipótese nula está incorreta e que as variâncias são significativamente diferentes – desse modo, a suposição de homogeneidade das variâncias foi violada. Se, entretanto, o teste de Levene for não-significativo (isto é, $p > 0,05$), precisamos aceitar a hipótese nula de que a diferença entre as variâncias é zero – as variâncias são praticamente iguais e a suposição é convincente. Para esses dados, o teste de Levene não é significativo (porque $p = 0,386$, que é maior do que 0,05); assim, devemos ler a estatística teste na linha chamada de *Variâncias iguais assumida*. Se o teste de Levene tivesse sido significativo, teríamos lido a estatística teste na linha chamada de *Variâncias iguais não-assumidas*.

Determinado que a suposição de homogeneidade das variâncias está satisfeita, podemos

Saída do SPSS 7.4

Independent Samples Test (Testes Amostrais Independentes)

		Levene's test for Equality of Variances (Teste de Levene de Homogeneidade de Variâncias)		t-test for Equality of Means (Teste t de Igualdade de Médias)						
		F	Sig.	t	df (gl)	Sig. (2-tailed) Sig. Bilateral	Mean Difference (Diferença das Médias)	Std. Error Difference (Erro Padrão da Diferença)	95% Confidence Interval of the Mean (Intervalo de Confiança de 95% para a Média)	
									Lower (Inferior)	Upper (Superior)
Anxiety (Ansiedade)	Equal variances assumed (Supondo Variâncias Homogêneas)	0.782	0.386	-1.681	22	0.107	7.0000	4.1633	-15.6342	1.6342
	Equal Variances Not Assumed (Supondo Variâncias Não-Homogêneas)			-1.681	21.385	0.107	7.0000	4.1633	-15.6486	1.6486

executar o teste t propriamente dito. Sabemos a diferença das médias ($\bar{X}_1 - \bar{X}_2 = 40 - 47 = -7$) e o erro padrão da distribuição amostral das diferenças é calculado utilizando-se o denominador da equação (7.5):

$$\begin{aligned}\sqrt{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)} &= \sqrt{\left(\frac{9,29^2}{12} + \frac{11,03^2}{12}\right)} \\ &= \sqrt{(7,19 + 10,14)} \\ &= \sqrt{17,33} \\ &= 4,16\end{aligned}$$

A estatística t é calculada dividindo a diferença média pelo erro padrão da distribuição amostral das diferenças ($t = -7/4,16 = -1,68$). O valor de t é, então, avaliado em relação ao valor de t que você pode esperar conseguir ao acaso quando você tem certos graus de liberdade. Para o teste t independente, os graus de liberdade são calculados somando os dois tamanhos amostrais e, depois, subtraindo o número de amostras utilizado, nesse caso duas ($df = N_1 + N_2 - 2 = 12 + 12 - 2 = 22$). O SPSS produz o valor exato da significância de t e queremos saber se esse valor é menor ou maior do que 0,05. Nesse caso, o valor bilateral de p é 0,107, o que é maior do que 0,05; portanto, não houve diferença significativa entre as médias dessas duas amostras. Nos termos do experimento, podemos inferir que aracnofóbicos ficaram igualmente ansiosos tanto com a figura da aranha quanto com sua presença verdadeira.

A probabilidade bilateral é utilizada quando uma previsão específica sobre a direção do nosso efeito não foi feita (veja a Seção 1.8.2). Por exemplo, se não soubermos se a aranha verdadeira irá provocar mais ou menos ansiedade, devemos usar o teste bilateral. Entretanto, geralmente em pesquisas podemos fazer previsões específicas sobre qual grupo tem a média maior. Nesse exemplo, provavelmente teríamos previsto que a aranha verdadeira provocaria mais ansiedade do que a figura e que, conseqüentemente, a média do grupo verdadeira (*real*) seria maior do que a média do grupo figura (*pictu-re*). Nesse caso, podemos usar o teste unilateral (para mais detalhes sobre esse assunto, veja a

Seção 1.8.2, ou Rowntree, 1981, Capítulo 7). Alguns estudantes ficam confusos com o fato de que o SPSS produz somente a significância bilateral e não tem uma opção para produzir a significância unilateral. O motivo é simples: não há necessidade de uma opção porque a probabilidade unilateral pode ser averiguada dividindo o valor da significância bilateral por 2. Nesse caso, a probabilidade bilateral foi 0,107, portanto, a probabilidade unilateral é 0,054 (0,107/2). A probabilidade unilateral é maior do que 0,05 (embora por uma pequena margem), desse modo, concluímos que os aracnofóbicos expostos à aranha verdadeira estavam tão ansiosos quanto os aracnofóbicos expostos à figura da mesma aranha. Esse resultado foi previsto pelo gráfico de barra de erros da Figura 7.3.

7.6.4 Calculando o tamanho de efeito ①

Para descobrir se o nosso efeito é importante, podemos usar a mesma equação da Seção 7.5.5 a fim de converter a estatística t em um valor de r . Sabemos o valor de t e do gl da saída do SPSS, podemos calcular r da seguinte maneira:

$$\begin{aligned}r &= \frac{-1,681^2}{\sqrt{-1,681^2 + 22}} \\ &= \frac{2,826}{\sqrt{24,826}} \\ &= 0,34\end{aligned}$$

Lembrando dos nossos pontos de referência para os tamanhos de efeito, esse resultado representa um efeito médio (ele está por volta de 0,3, o limite para um efeito médio). Portanto, mesmo que o efeito não tenha sido significativo, ele ainda representa um efeito substancial. Você também pode notar que o efeito encolheu, o que parece um pouco estranho dado que usamos exatamente os mesmos dados (mas veja a Seção 7.7)!

7.6.5 Relatando o teste t independente ①

As regras que relatei para o teste t dependente se aplicam quase que totalmente para o teste t independente. A saída do SPSS nos in-

Dica da Samanta Ferrinho



- O teste t independente compara duas médias quando elas vieram de diferentes grupos de pessoas; por exemplo, se você tem duas condições experimentais e usou participantes diferentes em cada condição.
- Veja a coluna rotulada de *Teste de Levene para a Igualdade de Variâncias*. Se o valor de *Sig* é menor do que 0,05, a hipótese de homogeneidade da variância foi violada e você deve olhar somente para a linha na tabela rotulada de *Variâncias iguais não presumidas*. Se o valor de *Sig* do teste de Levene é maior do que 0,05, você deve olhar para a linha na tabela rotulada de *Variâncias iguais presumidas*.
- Olhe para a coluna rotulada de *Sig*: se o valor é menor do que 0,05, as médias dos dois grupos são significativamente diferentes.
- Olhe para os valores das médias para saber como os grupos diferem.
- O SPSS fornece somente o valor de significância bilateral; se você quiser a significância unilateral, basta dividir o valor por 2.
- Relate a estatística t , os graus de liberdade e o valor de significância. Indique, também, as médias e seus erros padrão correspondentes (ou trace um gráfico de barra de erros).
- Se você for corajoso, calcule o tamanho de efeito e relate-o também!

forma que o valor de t era $-1,68$, que os graus de liberdade nos quais isso foi baseado eram 22, e que não era significativo para $p < 0,05$. Também conseguimos ver as médias para cada grupo. Podemos escrever isso como:

- ✓ Em média, os participantes tiveram mais ansiedade com a aranha verdadeira ($M = 47,00$, $EP = 3,18$) do que com a figura de uma aranha ($M = 40,00$, $SE = 2,68$). Essa diferença não foi significativa $t(22) = -1,68$, $p > 0,05$; entretanto, ela representou um tamanho de efeito médio $r = 0,34$.

Observe que relatamos as médias em cada grupo (e os erros padrão) como anteriormente. Para a estatística teste, tudo é muito parecido com antes, exceto que precisei relatar que p era maior do que ($>$) 0,05 em vez de menor do que ($<$). Finalmente, note que comentei sobre o tamanho de efeito no final.

7.7 ENTRE GRUPOS OU MEDIDAS REPETIDAS? ①

Os dois exemplos deste capítulo são interessantes porque ilustram a diferença entre os dados coletados usando os mesmos partici-

pantes e usando participantes diferentes. Eles utilizam os mesmos escores em cada condição. Os dois exemplos deste capítulo utilizam os mesmos escores nas duas condições. Quando os dados foram analisados como se tivessem vindo dos mesmos participantes, o resultado foi uma diferença significativa entre as médias, mas quando eles foram analisados como se tivessem vindo de participantes diferentes, não houve diferença significativa entre as médias dos grupos. Isso pode parecer uma descoberta desconcertante – afinal, os números eram idênticos em ambos os exemplos. O que isso ilustra é o *poder* relativo do delineamento das medidas repetidas. Quando os mesmos participantes são usados através das condições, a variância não-sistemática (geralmente chamada de variância do erro) é reduzida drasticamente, facilitando a detecção de qualquer variância sistemática. Muitos podem pensar que a maneira de coletar é irrelevante, mas espero ter provado que ela pode fazer a diferença entre detectar uma diferença ou não. De fato, pesquisadores têm conduzido estudos usando os mesmos participantes em condições experimentais, repetido o estudo utilizando participantes diferentes em condições expe-

rimentais e, então, usado o método de coleta de dados como uma variável independente na análise. Em geral, eles têm encontrado que o método de coleta de dados interage de forma significativa com os resultados encontrados (veja Erlebacher, 1977).

7.8 O TESTE t COMO UM MODELO LINEAR GENERALIZADO ②

Pode parecer estranho eu ter escolhido representar o tamanho de efeito para os meus testes t usando r , o coeficiente de correlação. Você pode estar pensando: “mas correlações mostram relacionamentos, não diferenças entre médias”. Eu costumava pensar assim também, até ler um artigo fantástico de Cohen (1968), o qual me fez entender o que eu estava perdendo: o complexo e espinhoso mundo da estatística subitamente ficou muito mais fácil e simples. Bem, nem tão fácil e simples assim, mas era um bom artigo! O que vou dizer poderá não fazer sentido algum para você ou poderá ajudá-lo a perceber que todos os procedimentos estatísticos são basicamente os mesmos: eles são versões mais ou menos elaboradas do coeficiente de correlação!

No Capítulo 5, vimos que o teste t foi usado para testar se o coeficiente de regressão de uma previsão era igual a 0. O projeto experimental para o qual o teste t dependente é usado pode ser conceitualizado como uma equação de regressão (afinal, existe uma variável independente – preditor – e uma variável dependente – saída). Se quisermos prever nossa saída, podemos usar a equação geral que mencionei em vários locais:

Se quisermos usar um modelo linear, vimos que essa equação geral se torna a equação (5.2), na qual o modelo é definido pela inclinação e intersecção da reta.

$$\text{Saída}_i = (\text{Modelo}_i) + \text{erro}_i$$

A equação (7.6) é muito similar, em que A_i é a variável dependente (saída), b_0 é o intercepto, b_1 é a ponderação do preditor e G_i é a variável independente (preditor). Inclui a mesma equação, mas com algumas das letras

trocadas em relação ao que elas representam no experimento da aranha (assim, $A = \text{ansiedade}$, $G = \text{grupo}$). Quando executamos um experimento com duas condições, a variável independente tem somente dois valores (grupo 1 e grupo 2). Esses grupos podem ser codificados de várias maneiras (no exemplo da aranha, codificamos o grupo 1 com o valor de 0 e o grupo 2 com o valor de 1). Essa variável codificada é chamada de *dummy* (variável auxiliar) e os valores dessa variável representam grupos de pessoas. Os leitores mais corajosos encontraram esta codificação na Seção 5.10:

$$A_i = b_0 + b_1 G_i + \varepsilon_i$$

$$\text{Ansiedade}_i = b_0 + b_1 \text{Grupo}_i + \varepsilon_i \quad (7.6)$$

Usando o exemplo da aranha, sabemos que a média da **ansiedade** do grupo da figura é de 40 e que a variável **grupo** é igual a 0 para essa condição. Veja o que acontece quando a variável **grupo** é igual a 0 (a condição figura); a equação (7.6) se torna (se ignorarmos o termo residual):

$$40 = b_0 + (b_1 \times 0)$$

$$b_0 = 40$$

Portanto, b_0 (o intercepto) é igual à média do grupo da figura (isto é, à média do grupo codificado de 0). Agora, vamos ver o que acontece quando a variável **grupo** é igual a 1. Essa condição, a da aranha verdadeira, teve uma média da **ansiedade** ($\bar{X}_{\text{Verdadeira}}$) de 47. Lembrando que acabamos de descobrir que b_0 é igual à média do grupo da figura (\bar{X}_{Figura}), a equação (7.6) se torna:

$$47 = b_0 + (b_1 \times 1)$$

$$47 = 40 + b_1$$

$$b_1 = 47 - 40$$

$$b_1 = \bar{X}_{\text{Verdadeira}} - \bar{X}_{\text{Figura}}$$

Portanto, b_1 representa a diferença entre as médias dos grupos. Como tal, podemos representar um experimento de dois grupos como uma equação de regressão na qual o coeficiente da variável independente (b_1) é igual à diferença entre as médias dos grupo e o intercepto (b_0) é igual à média do grupo codificado como 0. Na regressão, o teste t é usado para averi-

guar se o coeficiente de regressão (b_1) é igual a 0, e quando conduzimos um teste t com dados em grupos, testamos se a diferença entre as médias dos grupos é igual a 0.

Para provar que não estou inventando isso, use os dados em *spiderBG.sav* e execute uma regressão linear simples utilizando **grupo** como preditor e **ansiedade** como saída. **Grupo** é codificado usando esquema de código de 0 e 1 e, assim, representa a *variável auxiliar* descrita acima. A saída do SPSS resultante deve conter a tabela do resumo da regressão mostrada na Saída 7.5 do SPSS. Primeiro note o valor da constante (b_0): é 40, o mesmo da média da categoria base (o grupo da figura). Agora observe que o valor do coeficiente de regressão b_1 é 7, que é a diferença entre as médias dos dois grupos ($47 - 40 = 7$). Finalmente, se a estatística t , que testa se b_1 é significativamente diferente de 0, é a mesma para o teste t independente (veja a Saída 7.4 do SPSS), portanto, é o valor da significância.⁴

Esta seção demonstrou que experimentos podem ser representados em termos de modelos lineares; esse conceito é essencial para o entendimento dos capítulos seguintes sobre modelos lineares generalizados.

⁴ Na verdade, o valor da estatística t é o mesmo, mas é positivo em vez de negativo. Lembre-se da discussão da correlação ponto bisserial na Seção 4.5.6: quando você correlaciona uma variável dicotômica, a direção do coeficiente de correlação depende inteiramente de que casos são designados a quais grupos. Portanto, a direção da estatística t aqui é influenciada pelo grupo que selecionamos como a categoria base (a categoria codificada de 0).

7.9 O QUE FAZER SE OS DADOS NÃO FOREM NORMALMENTE DISTRIBUÍDOS? ②

Neste capítulo, vimos que é possível fazer ajustes no teste t quando a hipótese de homogeneidade da variância é violada; mas e quando você tiver dados não normalmente distribuídos? Uma possibilidade é tentar corrigir a distribuição usando uma transformação (veja o Capítulo 3), mas nem sempre funciona. Outra solução útil é usar um *teste não-paramétrico*. Esse tipo de teste tem menos hipóteses do que seus equivalentes paramétricos e, assim, é útil quando seus dados violam as hipóteses dos dados paramétricos descritos na Seção 3.2.1. Alguns desses testes estão descritos no Capítulo 13. O equivalente não-paramétrico do *teste t dependente* é chamado de *teste de postos com sinais de Wilcoxon* (Seção 13.3) e o teste t independente tem dois equivalentes não-paramétricos (ambos muito similares) chamados de *teste da soma dos postos de Wilcoxon* e *teste de Mann-Whitney* (Seção 13.2). Recomendo a leitura dessas seções antes de continuar.

7.10 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Abordamos muitas coisas neste capítulo; o que começou como “um capítulo breve sobre comparação de médias” acabou imenso. Essas tendências perfeccionistas são uma cruz que tenho que carregar!

Começamos o capítulo analisando as diferentes maneiras como os dados podem ser coletados e por que pode ser interessante com-

Saída 7.5 do SPSS Análise de Regressão entre grupos para os dados das aranhas

Coefficients^a (Coeficientes)

Model (Modelo)		Unstandardized Coefficients (Coeficientes Não-Padronizados)		Standardized Coefficients (Coeficientes Padronizados)	t	Sig.
		B	Std. Error (Erro Padrão)	Beta		
1	Constant (Constante)	40.000	2.944		13.587	0.000
	Condition (Condição)	7.000	4.163	0.337	1.681	0.107

a. Variável dependente: ansiedade.

parar duas médias. Abordamos alguns assuntos interessantes, como aleatorização. Falei sobre gráficos de barras de erros, primeiro para situações em que os dados são coletados usando participantes diferentes e depois quando os mesmos participantes são utilizados. Aprendemos sobre as características conceituais gerais do teste t , um teste paramétrico usado para testar diferenças entre duas médias, e depois vimos os aspectos específicos do teste t dependente (usado quando sua condição envolve as mesmas pessoas). Expliquei como ele é calculado, como fazê-lo no SPSS e como interpretar os resultados. Descobrimos praticamente o mesmo para o teste t independente (usado quando suas condições envolvem pessoas diferentes). Depois, falei sobre como uma situação com duas condições pode ser conceitualizada como um modelo linear geral. Finalmente, mencionei os equivalentes não-paramétricos do teste t : o teste da soma dos postos de Mann-Whitney e Wilcoxon (utilizados quando queremos comparar duas condições em que pessoas diferentes participaram) e o teste dos postos com sinais de Wilcoxon (usado quando você quer comparar duas condições em que as mesmas pessoas participaram). Eles estão descritos no Capítulo 13. Agora, com licença, que depois disso tudo eu mereço uma xícara de chá Darjeeling* e uma barra enorme de chocolate.

7.11 TERMOS-CHAVE QUE DESCOBRIMOS

- Delineamento entre grupos
- Delineamento entre participantes
- Efeitos de tédio
- Variáveis de confusão
- Contrabalanceamento
- Teste t dependente
- Variável dependente
- Gráfico de barras de erros
- Média geral

- Delineamento independente
- Teste t independente
- Variável independente
- Efeitos práticos
- Aleatorização
- Delineamento de medidas repetidas
- Erro padrão das diferenças
- Lei da soma das variâncias
- Delineamento dentro participantes

7.12 TAREFAS DO ALEX ESPERTO



Os dois cenários seguintes foram tirados de Field e Hold (2003). Em cada caso, analise os dados no SPSS. Respostas detalhadas para essas tarefas podem ser encontradas em Field e Hold ou resumidos no *site* www.artmed.com.br no arquivo **Answers (Chapter 7).pdf**.

- **Tarefa 1:** Eu simplesmente detesto livros de “psicologia pop”. Além de banir Freud de todas as livrarias, é minha ambição livrar o mundo desse desperdício de árvores. Esses livros não apenas dão má reputação à psicologia declarando o óbvio e ainda cobrando por isso, como também são menos agradáveis de olhar do que as árvores derrubadas para produzi-los. (O que também pode ser dito a respeito das bobagens que eu produzo em nome da educação, mas não vamos entrar nesse assunto agora!) Enfim, como parte do meu plano de abolir a psicologia popular do universo, fiz um pequeno experimento. Peguei dois grupos de pessoas que estavam em relacionamentos e aleatoriamente os designei a uma de duas condições. Um grupo leu o famoso livro de psicologia popular *Women are from bras and men are from penis*, e o outro grupo leu *Marie Claire*. Testei apenas 10 pessoas em cada um desses grupos e a variável dependente foi uma medida objetiva da sua felicidade com seus relacionamentos depois de ler o livro. Não fiz previsão específica alguma sobre qual material de leitura iria aumentar a felicidade nos relacionamentos. Os

* N. de T.: Chá preto muito apreciado na Inglaterra, proveniente da região de Darjeeling, Índia.

dados estão no arquivo **Penis.sav**; analise-os com o teste t apropriado. ①

- **Tarefa 2:** Imagine que a Twaddle & Sons, editora de *Women are from bras and men are from penis*, está chateada com a minha crítica de que seu livro é tão útil quanto um guarda-chuva de papel. Eles decidiram me encarregar da tarefa de delinear um experimento próprio no qual os participantes leram seu livro e um dos meus livros (Field e Hole) em períodos de tempo diferentes. A felicidade nos relacionamentos foi medida após a leitura de cada livro. Para maximizar as chances de encontrar uma diferença, eles usaram uma amostra de 500 participantes em que todos os participantes tomaram parte em ambas as condições (eles leram os dois livros). A ordem de leitura dos livros foi equilibrada, e houve um espaço de seis meses entre a leitura de um livro e o outro. Eles previram que sua contribuição maravilhosa à psicologia popular iria levar a uma felicidade maior nos relaciona-

mentos do que a leitura de um livro chato e entediante sobre experimentos. Os dados estão em **Field&Hole.sav**; analise-os usando o teste t apropriado. ①

7.13 LEITURAS COMPLEMENTARES

FIELD, A. P., HOLE, G. *How to design and report experiments*. London: Sage, 2003. Em minha opinião completamente imparcial, esse é um livro útil para adquirir mais conhecimento sobre métodos experimentais.

ROSNOW, R. L., ROSENTHAL, R. *Beginning behavioral research: a conceptual primer*. Englewood Cliffs (NJ): Pearson/Prentice Hall, 2005. 5th ed.

ROWNTREE, D. *Statistics without tears: a primer for non-mathematicians*. London: Penguin, 1981. Os Capítulos 4, 5, 6 e 7 contêm uma ótima e abrangente introdução às ideias de distribuições amostrais, teste de hipótese e inferência estatística.

WRIGHT, D. B. *First steps in statistics*. London: Sage, 2002. Contém um capítulo claro e excelente sobre o teste t (Capítulo 6).

8

COMPARANDO VÁRIAS MÉDIAS: ANOVA (MLG 1)

8.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

O capítulo anterior mostrou que podemos analisar as diferenças entre duas condições ou grupos de pessoas. Os testes explorados naquele capítulo são recursos úteis para a pesquisa em ciências sociais, contudo, eles estão limitados a situações em que existem apenas dois níveis da variável independente (isto é, dois grupos-experimentais). É comum a execução de experimentos em que existem três, quatro ou mesmo cinco níveis da variável independente; nesses casos, os testes vistos no Capítulo 7 não são apropriados. No lugar deles, uma técnica denominada **análise de variância** (ou **ANOVA** para os íntimos) deve ser utilizada. A vantagem da ANOVA é que ela pode ser utilizada para analisar situações nas quais existem diversas variáveis independentes. Nessas situações, a ANOVA informa como essas variáveis independentes interagem umas com as outras e que efeitos essas interações apresentam sobre a variável dependente. Este capítulo começa explicando a teoria da ANOVA quando diferentes participantes são usados (**ANOVA independente**). Depois iremos aprender a executar uma análise de variância no SPSS e a interpretar os resultados.

8.2 A TEORIA POR TRÁS DA ANOVA ②

8.2.1 Taxas de erros infladas ②

Antes de explicar como a ANOVA funciona, vale a pena mencionar por que não executamos testes t para comparar todas as combinações de grupos que estão sendo testados. Imagine uma situação em que existem três condições experimentais e estamos interessados nas diferenças entre esses três grupos. Se executássemos testes t para cada par de grupos, seria necessário executar três testes: um para comparar os grupos 1 e 2, outro para comparar os grupos 1 e 3 e um para comparar os grupos 2 e 3. Se cada um desses testes t utilizasse uma significância de 5%, a probabilidade de falsamente rejeitar a hipótese nula (conhecido como erro do Tipo I) para cada teste seria de somente 5%. Dessa forma, a probabilidade de não cometer o erro do Tipo I seria 0,95 (95%) para cada teste. Se assumirmos que cada teste é independente (ou seja, podemos multiplicar as probabilidades), a probabilidade total de não cometermos erro do Tipo I seria $(0,95)^3 = 0,857$, porque a probabilidade de não cometermos o erro do Tipo I em cada teste seria 0,95 e existem três testes. Dado

Por que não
realizar vários
testes t ?



que a probabilidade de não cometer o erro do Tipo I é 0,857, é possível calcular a probabilidade de cometer pelo menos um erro do Tipo I subtraindo esse resultado de 1 (lembre que a probabilidade máxima de algum evento ocorrer é 1). Assim, a probabilidade de pelo menos um erro do Tipo I ser cometido é $1 - 0,857 = 0,143$ ou 14,3%. Dessa forma, nesse conjunto de testes, a probabilidade de cometer erro do Tipo I aumentou de 5% para 14,3%, um valor bem maior que o critério aceito por cientistas sociais. Essa taxa de erro envolvendo um conjunto de testes conduzidos nos mesmos dados experimentais é conhecida como **erro de conjunto** ou **erro de experimento**. Um experimento com três condições é um delineamento relativamente simples, assim, o efeito de realizar vários testes não é severo. Se aumentássemos o número de condições experimentais de três para cinco (apenas acrescentando mais dois grupos), o número de testes t necessário aumentaria para 10.¹ O valor do erro de conjunto poderia ser calculado utilizando a equação geral (8.1), na qual n é o número de testes executados com os dados. Com 10 testes executados, o valor do erro de conjunto seria $0,40$ ($1 - 0,95^{10} = 0,40$), o que significa que teríamos uma probabilidade de 40% de ter cometido pelo menos um erro do Tipo I. Por essa razão, utilizamos a ANOVA em vez de realizar vários testes t :

$$\text{Erro de conjunto} = 1 - (0,95)^n \quad (8.1)$$

¹ Essas comparações são: 1 vs. 2, 1 vs. 3, vs. 4, 1 vs. 5, 2 vs. 3, 2 vs. 4, 2 vs. 5, 3 vs. 4, 3 vs. 5 e 4 vs. 5. O número de testes necessário é determinado utilizando a seguinte equação:

$$\text{Número de comparações } C = \frac{k!}{2(k-2)!}$$

onde k é o número de condições experimentais. O símbolo $!$ é denominado fatorial e significa que você deve multiplicar o valor que vem antes do símbolo por todos os números inteiros entre ele (inclusive) e zero (exclusive) (assim, $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$). Dessa forma, com cinco condições encontraremos que:

$$C = \frac{5!}{2(5-2)!} = \frac{120}{2 \times 6} = 10$$

Quando executamos testes t , testamos a hipótese de que as duas populações de onde as amostras foram retiradas apresentam a mesma média. Da mesma maneira, a ANOVA nos informa se três ou mais médias populacionais são iguais, assim, ela testa a hipótese de que as médias de todas as condições são iguais.

Uma ANOVA produz uma estatística F ou razão F que é semelhante a estatística t pois ela compara a variância sistemática nos dados com a variância não-sistemática. Contudo a ANOVA é um teste abrangente (*omnibus*), o que significa que ela testa um efeito experimental de forma global: assim, existem coisas que a ANOVA não nos informa. Embora a ANOVA nos informe se a manipulação experimental teve sucesso, ela não fornece informações específicas sobre quais grupos foram afetados. Supondo que um experimento tenha sido realizado com três grupos diferentes, a razão F simplesmente nos informa que as médias das três populações de onde os grupos foram retirados são diferentes (isto é, $m_1 = m_2 = m_3$ não é verdadeiro). Contudo, as médias podem diferir de várias maneiras. A primeira possibilidade é que as três médias sejam significativamente diferentes ($m_1 \neq m_2 \neq m_3$). Uma segunda possibilidade é que as médias das populações 1 e 2 não sejam diferentes mas tenham uma diferença significativa em relação à terceira população ($m_1 = m_2 \neq m_3$). Outra possibilidade é que as populações 2 e 3 não difiram mas sejam significativamente diferente da média da população 1 ($m_1 \neq m_2 = m_3$). Finalmente os grupos 1 e 3 podem indicar médias semelhantes mas o grupo 2 indica diferença significativa das populações 1 e 3 ($m_1 = m_3 \neq m_2$). Assim, a razão F informa apenas que a manipulação experimental tem algum efeito, mas não nos informa especificamente qual é esse efeito.

O que uma ANOVA diz?



8.2.2 ANOVA como regressão ②

Muitos cientistas sociais não se dão conta de que a ANOVA e a análise de regressão são conceitualmente o mesmo procedimento.

O motivo é principalmente histórico, pois dois ramos distintos da metodologia se desenvolveram nas ciências sociais: a pesquisa correlacional e a experimental. Pesquisadores interessados em experimentos controlados adotaram a ANOVA como técnica estatística básica, ao passo que os interessados em determinar relacionamentos reais adotaram a regressão múltipla. Como os cientistas são pessoas maduras, inteligentes e racionais, nenhum dos grupos tentou difamar o outro defendendo que a sua escolha metodológica foi melhor. Com essa divisão, a metodologia tornou-se uma divergência entre os métodos estatísticos adotados pelos dois campos opostos (Crombach, 1957, documenta essa divisão em um ótimo artigo). Essa divisão vingou por muitas décadas, tanto que ainda hoje os estudantes aprendem regressão e ANOVA em contextos bem diferentes, e muitos livros didáticos ensinam ANOVA de uma forma inteiramente distinta da regressão.

Embora muitas pessoas mais inteligentes do que eu tenham tentado restabelecer essa conexão entre ANOVA e análise de regressão (como o grande Jacob Cohen, 1968), estou empolgado em escrever o meu – breve e tolo – esclarecimento (e colocar a bola em jogo nas Seções 5.10 e 7.8). Existem várias boas razões para pensar que a ANOVA deve ser ensinada no contexto da regressão. Primeiro, ela fornece um contexto familiar: desperdicei várias árvores explicando a regressão, então por que não utilizar essa base de conhecimento para explicar o novo conceito (tornando-o mais fácil de compreender). Segundo, o método tradicional de ensinar a ANOVA (conhecido como método da razão das variâncias) é bom para delineamentos simples, mas problemático em situações mais complexas (como na análise de covariância). O modelo de regressão se estende de uma forma lógica para esses delineamentos mais complexos sem que seja necessário ficar atolado na matemática. Finalmente, o método das razões das variâncias é extremamente não-manejável em circunstâncias incomuns, como quando você tem amostras

de tamanhos diferentes.² O método da regressão torna essas situações muito mais simples. Embora essas razões sejam boas o suficiente, a verdade é que o SPSS também abandonou o método da razão das variâncias e passou a utilizar apenas o modelo de regressão (conhecido como modelo linear generalizado ou MLG). O resultado final das duas abordagens para a ANOVA é o mesmo e muitos textos já detalham a abordagem da razão das variâncias (recomendo o Howell, 2002). Assim faz sentido utilizar a abordagem da regressão.

A ANOVA é uma forma de comparar a razão entre a variância sistemática e a não sistemática em um estudo experimental. A razão dessas variâncias é conhecida como estatística *F* ou razão *F*. Quem leu o Capítulo 5 deve lembrar a razão *F* (veja a Seção 5.2.3) porque ela foi utilizada para avaliar quão bem um modelo de regressão pode prever uma saída comparada ao erro dentro do próprio modelo. Se você não leu o Capítulo 5 (provavelmente não!), dê uma olhada antes de ir adiante (deve demorar apenas duas semanas para lê-lo). A razão *F* na ANOVA é exatamente a mesma da regressão, exceto que um modelo de regressão contém somente precursores categóricos (isto é, variáveis de agrupamento). Desse modo, assim como o teste *t* pode ser representado por uma equação de regressão linear (veja a Seção 7.8), a ANOVA pode ser representada por uma equação de regressão múltipla na qual o número de precursores é um a menos do que o número de categorias da variável independente.

Vejamos um exemplo. Havia muita controvérsia sobre o medicamento Viagra quando escrevi a primeira edição deste livro. Agora existe bem menos polêmica, mas ela foi substituída por milhares de *spams* sobre o assunto (pelos quais eu serei grato, sem dúvida, daqui a 20 anos); assim, irei continuar com esse exemplo. O Viagra é um estimulante sexual (para tratar a impotência) utilizado com a crença de que ele fará do sujeito um amante melhor (estranha-

² Dito isso, vale tentar obter amostras de mesmo tamanho nas diferentes situações, pois delineamentos desbalanceados provocam complicações estatísticas (veja a Seção 8.2.9).

mente, vários jornalistas estão usando o produto em nome do “jornalismo investigativo”... hummmm!). Suponha que vamos testar essa crença em três grupos de participantes e administrar a um grupo um placebo (como uma pílula de açúcar), a outro grupo uma dose baixa de Viagra e ao terceiro grupo uma dose alta. A variável dependente é uma medida objetiva da libido (apenas relatarei que ela foi mensurada ao longo de uma semana – o resto deixo para a imaginação de cada um). Os dados estão no arquivo **Viagra.sav** (descrito em detalhes posteriormente nesse capítulo) e são apresentados na Tabela 8.1.

Se quisermos prever os níveis da libido a partir dos diferentes níveis de Viagra, poderemos utilizar a famosa equação geral:

$$\text{Saída}_i = (\text{Modelo}_i) + \text{erro}_i$$

Se você quiser utilizar um modelo linear, vimos na Seção 7.8 que quando existem somente dois grupos, podemos substituir o “modelo” nessa equação com uma equação de regressão linear como uma *dummy* (variável auxiliar) para descrever dois grupos-experimentais. Essa variável auxiliar é uma variável categórica com dois códigos numéricos (0 para um grupo e 1 para o outro). Com três grupos, no entanto, podemos estender essa ideia e utilizar um modelo e regressão múltipla com duas variáveis auxiliares. De fato, como regra geral, podemos estender um modelo para qualquer número de grupos e o número de variáveis auxiliares necessárias será um a menos que o número de categorias

da variável independente. No caso de dois grupos, uma categoria foi considerada como base (lembre que na Seção 7.8 escolhemos a **condição foto** como uma base) e essa categoria foi codificado como 0. Quando existem três categorias também precisamos de uma categoria base, e deve ser escolhida a condição com a qual se pretende comparar os demais grupos. Em geral, essa categoria será o grupo-controle. Nos experimentos sociais bem-conduzidos quase sempre há um grupo de participantes que agiu como grupo-base para as demais categorias. Esse grupo atua como uma referência ou categoria base, embora o grupo que é escolhido dependa das hipóteses que se pretende testar. Em delineamentos não-balanceados (os grupos não são do mesmo tamanho), é importante que a categoria base contenha muitos casos para assegurar que as estimativas dos coeficientes de regressão sejam confiáveis. No exemplo do Viagra, podemos usar o grupo-placebo como categoria base porque esse é o grupo-controle. Estamos interessados em comparar tanto o grupo com altas doses quanto o com doses baixas com o grupo que não tomou Viagra. Se o grupo-placebo é a categoria base, as duas variáveis auxiliares que precisamos criar representam as duas outras condições, assim, devemos ter uma variável auxiliar denominada Alta e outra Baixa. A equação resultante é apresentada em (8.2):

$$\text{Libido}_i = b_0 + b_2\text{Alta}_i + b_1\text{Baixa}_i + \varepsilon_i \quad (8.2)$$

Na equação (8.2), a libido de uma pessoa pode ser prevista a partir do conhecimento do código do seu grupo (isto é, o código das variáveis auxiliares Alta e Baixa) e pelo intercepto (b_0) do modelo. As variáveis auxiliares na equação (8.2) podem ser codificadas de várias formas, mas a mais simples é utilizar uma técnica semelhante a do teste t. A categoria base é sempre codificada como 0. Se um participante recebeu uma alta dose de Viagra, ele é codificado como 1 para a variável auxiliar e 0 para todas as outras variáveis. Se um participante recebeu uma dose baixa de Viagra, ele será codificado como o valor 1 para a variável auxiliar baixa e como 0 para as demais (esse é o mesmo esquema utilizado na Seção 5.10). Utilizando

Tabela 8.1 Dados em **Viagra.sav**

	Placebo	Dose baixa	Dose alta
	3	5	7
	2	2	4
	1	4	5
	1	2	3
	4	3	6
\bar{X}	2,20	3,20	5,00
s	1,30	1,30	1,58
s^2	1,70	1,70	2,50
Média global = 3,467 Desvio padrão global = 1,767			
Variância global = 3,124			

Tabela 8.2 Variáveis auxiliares para os três grupos do projeto experimental

Grupo	Variável auxiliar 1 alta y	Variável auxiliar 2 baixa y
Placebo	0	0
Dose baixa de Viagra	0	1
Dose alta de Viagra	1	0

esse esquema de codificação, podemos expressar cada grupo combinando os códigos das duas variáveis auxiliares (veja a Tabela 8.2).

Grupo-placebo: vamos examinar o modelo para o grupo-placebo. No grupo-placebo, as variáveis auxiliares dose alta e dose baixa são codificadas como 0. Além disso, se ignorarmos o termo erro (ε_i), a equação de regressão se torna:

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0$$

$$\bar{X}_{\text{Placebo}} = b_0$$

Essa é uma situação em que os grupos de dose alta e baixa foram ambos excluídos (eles foram codificados como 0). Queremos prever o nível de libido quando ambas as doses de Viagra forem ignoradas, desse modo, o valor previsto será a média do grupo-placebo (porque esse é o único grupo incluído no modelo). Assim, o intercepto do modelo de regressão, b_0 , é sempre a média da categoria base (nesse caso, a média do grupo-placebo).

Grupo de alta dosagem: se examinarmos o grupo de alta dosagem, a variável auxiliar alta será codificada como 1 e a variável auxiliar baixa, como 0. Se substituirmos os valores desses códigos na equação (8.2), o modelo fica:

$$\text{Libido}_i = b_0 + (b_2 \times 1) + (b_1 \times 0)$$

$$\text{Libido}_i = b_0 + b_2$$

Já sabemos que b_0 é a média do grupo-placebo. Se estivermos somente interessados no grupo de alta dosagem, então o modelo deve prever que o valor da libido para um dado participante iguale a média do grupo de alta dosagem. Dada essa informação, a equação torna-se:

$$\text{Libido}_i = b_0 + b_2$$

$$\bar{X}_{\text{Alta}} = \bar{X}_{\text{Placebo}} + b_2$$

$$b_2 = \bar{X}_{\text{Alta}} - \bar{X}_{\text{Placebo}}$$

Assim, b_2 representa a diferença entre as médias dos grupos de alta dosagem e o grupo-placebo.

Grupo de baixa dosagem: finalmente, se olharmos para o modelo quando uma dose baixa de Viagra for administrada, a variável auxiliar será codificada como 1 (e o grupo de alta dosagem como 0). Assim, a equação de regressão se torna:

$$\text{Libido}_i = b_0 + (b_2 \times 0) + (b_1 \times 1)$$

$$\text{Libido}_i = b_0 + b_1$$

Sabemos que o intercepto é igual a média da categoria-base e que, para o grupo de baixa dosagem, o valor previsto deve ser a média da libido para o de baixa dosagem. Dessa forma, o modelo pode ser reduzido para:

$$\text{Libido}_i = b_0 + b_1$$

$$\bar{X}_{\text{Alta}} = \bar{X}_{\text{Placebo}} + b_1$$

$$b_1 = \bar{X}_{\text{Baixa}} - \bar{X}_{\text{Placebo}}$$

Assim, b_1 representa a diferença entre as médias dos grupos de baixa dosagem e o grupo-placebo. Essa forma de codificação com variáveis auxiliares é a mais simples, mas, como veremos mais tarde, existem outras formas em que as variáveis podem ser codificadas para testar hipóteses específicas. Esses esquemas de codificação alternativos são denominados *contrastes* (veja a Seção 8.2.10.2). O objetivo dos contrastes é codificar as variáveis auxiliares de tal forma que os valores-b representem diferenças entre os grupos que estamos interessados em testar.

Para ilustrar exatamente o que está acontecendo, criei um arquivo denominado **dummy.sav** na pasta do Chapter 8 disponível no *site* www.artmed.com.br. Esse arquivo contém os dados do Viagra, mas com duas variáveis auxiliares adicionais (**dummy1** e **dummy2**) que especificam a que grupo cada dado pertence (como na Tabela 8.2). Acesse esse arquivo

e execute uma análise de regressão múltipla utilizando **libido** como saída e **dummy1** e **dummy2** como variáveis predictoras. Se você não sabe como executar uma regressão, leia o Capítulo 5 novamente (observe que os capítulos estão ordenados por um motivo!). A análise resultante é mostrada na Saída 8.1 do SPSS. É recomendável relembrar as médias dos grupos a partir da Tabela 8.1. Note que a constante é igual à média da categoria-base (o grupo-placebo). O coeficiente de regressão para a primeira variável auxiliar (b_2) é igual à diferença entre as médias dos grupos de alta dosagem e o grupo-placebo ($5,0 - 2,2 = 2,8$). Finalmente, o coeficiente de regressão para a segunda variável auxiliar (b_1) é igual à diferença entre as médias dos grupos de baixa dosagem e o grupo-placebo ($3,2 - 2,2 = 1$). Essa análise demonstra como o modelo de regressão representa a situação dos três grupos. Podemos ver a partir dos valores da significância dos testes t que a diferença entre o grupo de alta dosagem e o grupo-placebo (b_2) é significativa porque $p < 0,05$. A diferença entre o grupo de baixa dosagem e o grupo-placebo não é, entretanto, significativa ($p = 0,282$).

Um experimento para quatro grupos pode ser descrito estendendo o cenário dos três grupos. Mencionei anteriormente que você sempre irá precisar de uma variável auxiliar a menos do que o número de grupos no experimento: assim, esse modelo requer três ou mais variáveis. Como antes, você precisa especificar uma categoria que será a categoria base (o grupo-controle). Essa categoria base

deve ser codificada como 0 para todas as variáveis auxiliares. As três condições restantes terão um código de 1 para a variável auxiliar que descreve a condição e o código de 0 para as outras duas variáveis auxiliares. A Tabela 8.3 ilustra como codificar esse esquema.

8.2.3 A lógica da razão F ②

No Capítulo 5, aprendemos um pouco sobre a razão F e como calculá-la. Recapitulando, aprendemos que a razão F é utilizada para testar o ajustamento global de um modelo de regressão a um conjunto de dados. Recém expliquei que a ANOVA pode ser representada como uma equação de regressão, e isso deve ajudá-lo a entender o que a razão F informa sobre os dados. A Figura 8.1 mostra os dados do Viagra na forma gráfica (incluindo as médias dos grupos, a média geral e a diferença entre cada caso e a média do grupo). Nesse exemplo, existem três grupos; dessa forma, queremos testar a hipótese de que as médias dos três grupos são diferentes (assim, a hipótese nula

Tabela 8.3 Variáveis auxiliares para um projeto de quatro grupos

Grupo	Variável auxiliar 1	Variável auxiliar 2	Variável auxiliar 3
Grupo 1	1	0	0
Grupo 2	0	1	0
Grupo 3	0	0	1
Grupo 4 (base)	0	0	0

Saída do SPSS 8.1

Coefficients^a (Coeficientes)

Model (Modelo)		Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
		B	Std. Error (Erro padrão)	Beta		
1	(Constant) (Constante)	2.200	0.627		3.508	0.004
	Dummy Variable 1 (Variável auxiliar 1)	2.800	0.887	0.773	3.157	0.008
	Dummy Variable 2 (Variável auxiliar 2)	1.000	0.887	0.276	1.127	0.282

a. Dependent Variable: Libido (Variável dependente: Libido)

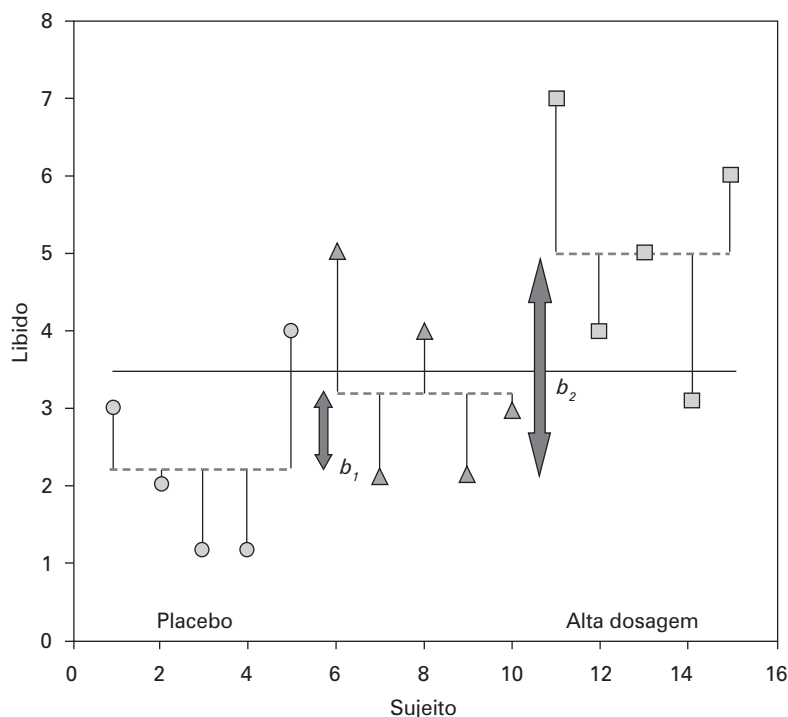


Figura 8.1 Dados do Viagra no formato gráfico. As linhas pontilhadas representam a libido média de cada grupo. As diferentes formas representam a libido de participantes individuais (formas diferentes indicam grupos-experimentais diferentes). A linha cheia horizontal é a libido média de todos os participantes.

é que as médias dos grupos são iguais). Se as médias dos grupos são todas iguais, não esperamos que o grupo-placebo seja diferente dos grupos de alta e baixa dosagem e não esperamos que o grupo de baixa dosagem difira do grupo de alta dosagem. Portanto, no diagrama, as três linhas pontilhadas estarão na mesma posição horizontal (a posição exata será a média geral). É possível ver a partir do diagrama que as médias dos grupos são de fato diferentes porque as linhas pontilhadas (as médias dos grupos) estão em posições horizontais distintas. Acabamos de descobrir que no modelo de regressão, b_2 representa a diferença entre as médias dos grupos-placebo e alta dosagem e que b_1 representa a diferença nas médias dos grupos-placebo e o de baixa dosagem. Essas duas distâncias estão representadas na Figura 8.1 por setas verticais. Se a hipótese nula é verdadeira e todos os grupos têm a mesma média,

esses coeficientes b devem ser 0 (porque se as médias dos grupos são iguais, as diferenças entre elas serão iguais a zero).

A lógica da ANOVA deriva do que entendemos sobre regressão:

- O modelo mais simples que podemos ajustar a um conjunto de dados é a média geral (a média da variável de saída). Quando esse modelo básico é ajustado, haverá uma grande quantidade de erro entre o modelo e os dados observados.
- Uma de muitas opções de linhas retas pode ser escolhida para modelar os dados coletados. Se esse modelo adere bem aos dados, ele deve ser melhor do que utilizar a média geral.
- O intercepto e um ou mais dos coeficientes de regressão podem descrever a linha escolhida.

- Os coeficientes de regressão determinam a inclinação da linha; dessa forma, quanto maior os coeficientes, maior o desvio entre a linha e a média geral.
- Na ANOVA, os coeficientes de regressão são definidos pelas diferenças entre médias de grupos.
- Quanto maior as diferenças entre médias dos grupos, maior a diferença entre a linha de regressão e a média geral.
- Se as diferenças entre as médias dos grupos são grandes o suficiente, a linha resultante será um modelo melhor do que a média geral.
- Além disso, se as médias dos grupos são significativamente diferentes, o modelo de regressão deve se ajustar melhor do que a média geral.

Podemos ver se a última afirmação é verdadeira comparando a melhoria no ajuste devido ao uso do modelo (em vez de utilizar a média geral) com o erro que permanece. Outra forma de dizer isso é que quando a média geral é utilizada como um modelo, existirá certa quantidade de variação entre os dados e a média geral. Quando o modelo é ajustado ele irá explicar parte dessa variação, mas parte ainda

permanecerá não-explicada. A estatística F é a razão entre a variação explicada e a não-explicada. Veja novamente a Seção 5.2.3 para refrescar a memória sobre esses conceitos antes de prosseguir. Isso pode parecer complicado, mas, de fato, quase tudo se resume a variações de uma equação simples (veja o Quadro 8.1).

8.2.4 Soma dos quadrados Total (SS_T) ②

Para encontrar a variação total dentro dos nossos dados calculamos as diferenças entre cada valor observado e a média geral (média de todos os valores). Elevamos então essas diferenças ao quadrado e adicionamos todas para obter a soma total dos quadrados (SS_T):

$$SS_T = \sum (x_i - \bar{x}_{\text{geral}})^2 \quad (8.4)$$

Vimos também na Seção 1.4.1 que a variância e a soma dos quadrados estão relacionadas de forma que a variância $s^2 = SS_T / (N - 1)$, onde N é o número de observações. Dessa forma, podemos calcular a soma dos quadrados total da variância de todas as observações (a **variância geral**) pela modificação da relação ($SS_T = s^2(N - 1)$). A variância ge-

Quadro 8.1

Você pode achar surpreendente que a ANOVA se resume a uma equação (bem, mais ou menos) ②

A cada estágio da ANOVA, estamos avaliando variações (ou desvios) de um determinado modelo (seja ele o modelo mais básico ou o mais sofisticado). Vimos na Seção 1.4.1 que “o quanto” um modelo se desvia dos dados observados pode ser expresso, em geral, na forma da equação (8.3). Assim, na ANOVA, como na regressão, utilizamos a equação (8.3) para calcular a aderência ao modelo mais básico e depois a aderência ao melhor modelo (a linha da melhor aderência). Se o melhor modelo é mais adequado, ele deve se ajustar aos dados significativamente melhor do que o modelo básico:

$$\text{desvios} = \sum (\text{observado} - \text{modelo})^2 \quad (8.3)$$

O ponto interessante é que todas as somas dos quadrados na ANOVA são variações dessa equação básica. O que muda é o que utilizamos como modelo e os dados que são ajustados. Olhe as várias seções das somas dos quadrados e compare as equações resultantes com a equação (8.3); espero que você consiga ver que todas são variações dessa equação básica!

ral é a variação entre todos os escores, independentemente das condições experimentais a que esses escores foram submetidos. Além disso, na Figura 8.1 ela será a soma das distâncias ao quadrado entre cada valor e a linha cheia horizontal. A variância geral para os dados do Viagra é apresentada na Tabela 8.1 e se contarmos o número de observações encontraremos que são 15 no total. Portanto, SS_T é determinado da seguinte forma:

$$\begin{aligned} SS_T &= s_{\text{geral}}^2(n - 1) \\ &= 3,124(15 - 1) \\ &= 3,124 \times 14 \\ &= 43,74 \end{aligned}$$

Antes de prosseguirmos, é importante entender o que são graus de liberdade; assim, leia o Quadro 8.2.

8.2.5 Soma dos Quadrados do Modelo (SS_M) ②

Por enquanto, sabemos que a quantidade de variação entre os dados é 43,74 unidades. Agora precisamos saber quanto o modelo de regressão pode explicar dessa variação. Na

ANOVA, o modelo é baseado nas diferenças entre as médias dos grupos, desse modo, a soma dos quadrados do modelo nos informa quanto da variação total pode ser explicado pelo fato de que dados diferentes provêm de grupos diferentes. Resumindo, o modelo representa o efeito da manipulação experimental.

Na Seção 5.2.3, vimos que a soma dos quadrados do modelo é calculada tomando-se as diferenças entre os valores previstos pelo modelo e a média geral (veja a Figura 5.3). Na ANOVA, os valores previstos pelo modelo são as médias dos grupos (portanto, na Figura 8.1 as linhas pontilhadas representam os valores da libido previstos pelo modelo). Para cada participante o valor previsto pelo modelo é a média do grupo a qual pertence o participante. No exemplo do Viagra, o valor previsto para os cinco participantes no grupo-placebo é 2,2, para os cinco participantes do grupo de baixa dosagem esse valor é 3,2 e para os cinco participantes do grupo de alta dosagem, 5,0. A soma dos quadrados do modelo requer que calculemos as diferenças entre o valor previsto de cada participante e a média geral. Essas diferenças são então elevadas ao quadrado e adicionadas (por razões que já devem estar claras

Quadro 8.2

Graus de liberdade ②

O conceito de graus de liberdade (gl) é difícil de explicar. Já o mencionei várias vezes ao longo do livro sem, no entanto, defini-lo. Muitos livros didáticos não explicam bem a ideia (embora Howell, 2002, p. 56, faça um bom trabalho) e este livro não será uma exceção à regra! Vou começar com uma analogia. Imagine que você seja o treinador de um time de rúgbi e tenha uma planilha do time com 15 células vazias que correspondem às posições dos jogadores no campo. Existe uma formação padrão nesse esporte: cada time tem 15 posições específicas que devem permanecer constantes durante o jogo. Quando o primeiro jogador chega existem 15 posições livres em que ele pode ser colocado. Você coloca o nome desse jogador em uma célula da planilha e o aloca em uma posição (por exemplo, linha média); uma posição no campo agora está ocupada. Quando o jogador seguinte chega, existem 14 posições livres e você ainda tem a liberdade de escolher em que posição o jogador será colocado. À medida que mais jogadores são alocados, chegará um ponto em que 14 posições já estarão preenchidas e restará um jogador. Não há liberdade de escolher a posição desse jogador, porque só existe uma possível. Portanto, existem 14 graus de liberdade, isto é, você tem possibilidade de escolher onde 14 jogadores irão jogar, mas para um jogador não há escolha. O grau de liberdade é um a menos que o número de jogadores.

(*Continua*)

Quadro 8.2 (Continuação)

Em termos estatísticos, os graus de liberdade estão relacionados ao número de observações que são livres para variar. Se retirarmos uma amostra de quatro observações de uma população para estimarmos o seu desvio padrão, precisaremos utilizar a média da amostra como uma estimativa da média da população. Assim, manteremos um valor constante. Digamos que a média da amostra seja 10; vamos assumir que a média da população seja esse valor e vamos mantê-lo constante. Com esse valor fixado, os quatro valores da amostra podem variar? A resposta é não porque para manter a média constante somente três valores podem variar. Por exemplo, se os valores da amostra forem 8, 9, 11 e 12 (média = 10) e trocarmos três desses valores por 7, 15 e 8, o último valor deve ser 10 para que a média não seja alterada. Portanto, se mantivermos um parâmetro constante, os graus de liberdade devem ser o tamanho da amostra menos um. Esse fato explica por que, quando utilizamos uma amostra para estimar o desvio padrão da população (como fizemos na Seção 1.4.1), devemos dividir a soma dos quadrados por $N - 1$ em vez de por N apenas. Para o SS_T , utilizamos a totalidade da amostra para determinar a soma dos quadrados, assim o total de graus de liberdade (GL_T) será o tamanho da amostra menos um ($N - 1$). Para os dados do Viagra, esse valor é 14.

para você a essa altura). Sabemos que o valor previsto para os participantes em um grupo específico é a média desse grupo. Portanto, a forma mais simples de calcular a SS_M é:

1. Calcular as diferenças entre a média de cada grupo e a média geral.
2. Elevar cada uma dessas diferenças ao quadrado.
3. Multiplicar cada resultado pelo número de participantes dentro de cada grupo (n_k).
4. Somar os valores de cada um dos grupos.

A expressão resultante desse processo é mostrada na equação (8.5):

$$SS_M = \sum n_k (\bar{x}_k - \bar{x}_{\text{geral}})^2 \quad (8.5)$$

Utilizando as médias dos dados do Viagra, podemos calcular a SS_M da seguinte forma:

$$\begin{aligned} SS_M &= 5(2,200 - 3,467)^2 + 5(3,200 - 3,467)^2 \\ &\quad + 5(5,000 - 3,467)^2 \\ &= 5(-1,267)^2 + 5(-0,267)^2 + 5(1,533)^2 \\ &= 8,025 + 0,355 + 11,755 \\ &= 20,135 \end{aligned}$$

Para a SS_M , os graus de liberdade GL_M serão sempre menos um o número de parâmetros estimados menos um. Em resumo, o

valor será o número de grupos menos 1 (que representaremos por $k - 1$). Assim, no caso de três grupos, o número de graus de liberdade será 2 (em virtude dos cálculos da soma dos quadrados ter por base a média dos grupos, duas serão livres para variar na população se a terceira for mantida constante).

8.2.6 Soma dos Quadrados dos Resíduos (SS_R) ②

Sabemos que existem 43,74 unidades de variação para serem explicadas nos nossos dados e que nosso modelo pode explicar 20,14 dessas unidades (aproximadamente a metade). A soma final dos quadrados é a soma dos quadrados dos resíduos (SS_R), que informa quanto da variação total não pode ser explicado pelo modelo. Esse valor é a quantidade de variação causada por fatores estranhos, como diferenças individuais de peso, testosterona ou outros. Se já conhecemos SS_T e SS_M , então a maneira mais fácil de calcular o valor de SS_R é subtraindo SS_M de SS_T ($SS_R = SS_T - SS_M$); contudo, fazendo isso teremos pouca ideia do que estamos calculando e, é claro, se errarmos no cálculo da SS_M ou da SS_T (ou de ambos), a SS_R também estará errada. Vimos na seção 5.2.3 que a soma dos quadrados dos resíduos é a diferença entre o que o modelo prevê e o

que é observado de fato. Sabemos que para um dado participante, o modelo prevê a média do grupo que esse participante pertence. Portanto, SS_R é calculado pela diferença entre o escore obtido por essa pessoa e a média do grupo que ela pertence. Em termos gráficos, as linhas verticais na Figura 8.1 representam essa soma dos quadrados. Essas distâncias entre cada dado e a média do grupo são elevadas ao quadrado e então somadas para fornecer a soma dos quadrados dos resíduos, SS_R (veja a equação (8.6)):

$$SS_R = \sum (x_{ik} - \bar{x}_k)^2 \quad (8.6)$$

Agora, a soma dos quadrados para cada grupo representa a soma das diferenças ao quadrado entre o escore de cada participante no grupo e a média do grupo. Portanto, podemos expressar a SS_R como $SS_R = SS_{\text{Grupo1}} + SS_{\text{Grupo2}} + \dots + SS_{\text{Grupok}}$. Dado que sabemos a relação entre a variância e a soma dos quadrados, podemos utilizar a variância de cada grupo dos dados do Viagra para criar uma equação como a que foi feita para a soma total dos quadrados. Dessa forma, a SS_R pode ser expressa pela equação (8.7):

$$SS_R = \sum s_k^2(n_k - 1) \quad (8.7)$$

Isso significa que devemos pegar a variância de cada grupo (s_k^2) e multiplicá-la pelo número de pessoas naquele grupo menos um ($n_k - 1$). Depois de fazer isso para cada grupo, some todos os resultados. Para os dados do Viagra, teremos:

$$\begin{aligned} SS_R &= s_{\text{grupo 1}}^2(n_1 - 1) + s_{\text{grupo 2}}^2(n_2 - 1) \\ &\quad + s_{\text{grupo 3}}^2(n_3 - 1) \\ &= (1,70)(5 - 1) + (1,70)(5 - 1) + (2,50)(5 - 1) \\ &= (1,70 \times 4) + (1,70 \times 4) + (2,50 \times 4) \\ &= 6,8 + 6,8 + 10 \\ &= 23,60 \end{aligned}$$

O número de graus de liberdade para a quantidade SS_R (gl_R) pode ser obtido subtraindo-se o número de graus de liberdade do mo-

delo do número de graus de liberdade total, isto é, $gl_R = gl_T - gl_M = 14 - 2 = 12$. Colocando de outra forma, isso é $N - k$, ou seja, o total da amostra (N) menos o número de grupos (k).

8.2.7 Médias ao quadrado ②

O valor SS_M nos informa quanto da variação o modelo de regressão (por exemplo, a manipulação experimental) explica e o SS_R informa quanto da variação pode ser creditado a fatores estranhos. Contudo, por que esses valores são somas, eles serão influenciados pelo número de valores que foram adicionados (por exemplo, SS_M utiliza a soma de apenas três valores diferentes – as médias dos grupos – comparadas a SS_R e SS_T , que utilizam a soma de 14 valores diferentes). Para eliminar esse viés, é possível calcular a soma dos quadrados média (conhecida como *média dos quadrados*, MS), que são simplesmente as somas dos quadrados divididas pelos graus de liberdade. Dividimos pelo número de graus de liberdade em vez de pelo número de parâmetros utilizados para calcular a SS porque estamos tentando extrapolar para uma população e alguns parâmetros dentro da população serão constantes. Assim, para os dados do Viagra, vamos obter as seguintes médias dos quadrados:

$$\begin{aligned} MS_M &= \frac{SS_M}{gl_M} = \frac{20,135}{2} = 10,067 \\ MS_R &= \frac{SS_R}{gl_R} = \frac{23,60}{12} = 1,967 \end{aligned}$$

MS_M representa a quantidade média de variação explicada pelo modelo (isto é, a variação sistemática). Já MS_R é uma medida da quantidade média de variação explicada pelas variáveis estranhas (variação não-sistemática).

8.2.8 A razão F ②

A razão F é uma medida da variação explicada pelo modelo dividida pela variação explicada por fatores não-sistemáticos. Ela pode ser calculada pela divisão da média dos quadrados do modelo pela média dos quadrados dos resíduos:

$$F = \frac{MS_M}{MS_R} \quad (8.8)$$

Assim como o teste *t* independente, a razão *F* é uma medida da razão das variações sistemáticas para as não-sistemáticas. Como tal, ela é razão do efeito experimental para as diferenças individuais de desempenho. Um ponto interessante sobre a razão *F* é que, em virtude dela ser um quociente entre as variações sistemáticas e não-sistemáticas, se o seu valor for menor do que 1, então por definição ela representa um efeito não-significativo. Isso é verdade porque se a razão *F* é menor do que 1, significa que a MS_R é maior do que a MS_M , que em termos reais significa que existe mais variação não-sistemática do que sistemática. Você pode pensar nisso em termos do efeito de diferenças naturais na habilidade serem maiores do que as diferenças introduzidas pelo experimento. Nesse cenário, podemos estar seguros de que a nossa manipulação experimental não foi bem-sucedida (porque ela trouxe menos mudanças do que se tivéssemos deixado os participantes por si só!). Para os dados do Viagra a razão *F* é:

$$F = \frac{MS_M}{MS_R} = \frac{10,067}{1,967} = 5,12$$

Esse valor é maior do que 1, o que indica que a manipulação experimental teve um efeito além do efeito das diferenças de desempenho individuais. Contudo, ela não nos informa se a razão *F* é grande o suficiente para não ser um resultado casual. Para descobrir isso, devemos comparar o valor de *F* obtido contra o valor máximo que seria esperado apenas por acaso utilizando uma distribuição *F* com os mesmos graus de liberdade (esses valores podem ser encontrados no Apêndice A.3). Se o valor que obtivemos exceder esse valor crítico, podemos estar confiantes de que ele reflete um efeito da variável independente. Nesse caso, com 2 (gl_m) e 12 (gl_r) graus de liberdade, os valores críticos são 3,89 ($p = 0,05$) e 6,93 ($p = 0,01$), respectivamente. O valor observado, 5,12, é, dessa forma, significativo a 5%, mas não a 1%. Portanto, a significância exata fornecida pelo SPSS deve estar entre 0,05 e 0,01 (e de fato está).

8.2.9 Suposições da ANOVA ②

As suposições sob as quais a ANOVA é confiável são as mesmas dos testes paramétricos baseados na distribuição normal (veja a Seção 3.2). Isto é, os dados devem ser de uma população distribuída normalmente, as variâncias em cada uma das condições experimentais devem ser homogêneas, as observações devem ser independentes e a variável dependente deve ser mensurada pelo menos em uma escala de intervalo. Embora eu esteja sempre falando sobre a importância das hipóteses elas não são totalmente inflexíveis. Por exemplo, Lunney (1970) investigou o uso da ANOVA quando a variável dependente é dicotômica (podemos ter valores do tipo 0 e 1 apenas). Os resultados mostraram que quando os tamanhos dos grupos eram iguais, a ANOVA era precisa quando existiam pelo menos 20 graus de liberdade e a menor categoria continha no mínimo 20% de todas as respostas. Se a menor categoria de respostas continha menos do que 20% de todas as respostas, a ANOVA era precisa somente quando existiam 40 ou mais graus de liberdade. Esse estudo mostra que a ANOVA pode ser um procedimento bastante robusto.

Em termos de violação das hipóteses de homogeneidade da variância, a ANOVA é bastante robusta quando os tamanhos amostrais são iguais. Contudo, quando os tamanhos das amostras são diferentes, a ANOVA não é robusta a violações da homogeneidade da variância (por isso disse antes que vale a pena tentar coletar dados em grupos de mesmo tamanho!). Quando grupos com grandes tamanhos amostrais apresentam variâncias maiores do que grupos com tamanhos amostrais menores, a razão *F* resultante tende a ser conservadora. Quer dizer, é mais provável obter um resultado não-significativo mesmo quando existe uma diferença genuína na população. Já quando os grupos com grandes tamanhos amostrais apresentam variâncias menores do que os grupos com tamanhos amostrais menores, a razão *F* resultante tende a ser liberal. Ou seja, é mais provável obter um resultado significativo quando não existe uma diferença entre os grupos na população (ou seja, a taxa de Erro I não é controlada) – veja Glass, Peckham e Saunders (1972) para mais detalhes sobre o assunto.

Os problemas resultantes da violação da hipótese da homogeneidade das variâncias podem ser corrigidos (veja o Quadro 8.3). Assim, o problema mais sério é a violação da hipótese de independência. Scariano e Davenport (1987) mostraram que quando essa hipótese é violada (isto é, as observações entre os grupos são correlacionadas), a taxa de erro do Tipo I aumenta substancialmente. Por exemplo, utilizando o valor convencional de 5% para o erro do Tipo I da situação de observações independentes, se essas observações forem modificadas de modo que tenham uma correlação moderada (digamos, com um coeficiente de Pearson de 0,5), quando formos comparar três grupos de 10 observações por grupo, a taxa real de erro do Tipo I será 0,74 (um aumento substancial!). Portanto, se as observações forem correlacionadas, você pode pensar que está trabalhando com uma taxa de erro de aceitação de 5% (isto é, você irá encontrar um efeito significativo de modo incorreto apenas 5% das vezes) quando, de fato, a taxa de erro verdadeira está próxima de 75% (isto é, encontraremos um resultado significativo em 75% das vezes quando na realidade não existe tal efeito na população).

8.2.10 Contrastes planejados ②

A razão F nos informa apenas que o modelo ajustado aos dados é responsável por mais variação do que fatores estranhos, mas ela não informa quais grupos diferem. Assim, se o valor F é grande o suficiente para ser estatisticamente significativo, sabemos apenas que uma ou mais das diferenças entre as médias é estatisticamente significativa (por exemplo, tanto b_2 quanto b_1 são estatisticamente significativos). Portanto, após a realização da ANOVA precisamos executar análises adicionais para descobrir quais grupos diferem entre si. Na regressão múltipla, cada coeficiente b é testado individualmente por meio do teste t e devemos fazer o mesmo para a ANOVA. Contudo, precisaremos executar dois testes t , que deverá inflacionar o erro de conjunto (*familywise error rate*) (veja a Seção 8.2.1). Desse modo, precisamos encontrar uma maneira de comparar os diferentes grupos sem

inflacionar a taxa de erro do Tipo I. Existem duas formas de fazer isso. A primeira é dividir a variância de responsabilidade do modelo nas suas parcelas componentes e a segunda é comparar cada grupo (como se estivéssemos realizando vários testes t), mas utilizando um critério de aceitação mais restrito de forma que o erro de conjunto não ultrapasse o valor fixado (0,05). A primeira opção é executada utilizando-se as comparações planejadas (também conhecidas como **contrastes planejados**)³, e a segunda é realizada utilizando-se testes *post hoc* (veja a próxima seção). A diferença entre comparações planejadas e testes *post hoc* pode ser comparada à diferença entre testes uni e bilaterais, no sentido de que as comparações planejadas são feitas quando temos hipóteses específicas que queremos testar, ao passo que os testes *post hoc* são feitos quando não temos hipóteses específicas. Vamos examinar primeiro os contrastes planejados.

8.2.10.1 Escolhendo que contrastes realizar ②

No exemplo do Viagra, podemos ter hipóteses bastante específicas. Por um lado, esperamos que qualquer dose de Viagra altere a libido quando comparado ao grupo-placebo. Como uma segunda hipótese, podemos acreditar que uma alta dosagem deve aumentar mais a libido do que uma baixa dosagem. Para executar as comparações planejadas, essas hipóteses devem ser efetuadas *antes* dos dados serem coletados. É praticamente um padrão nas ciências sociais comparar as condições experimentais às condições de controle como o primeiro contraste e depois ver onde as diferenças estão entre os grupos-experimentais. A ANOVA está baseada na divisão da variação total nas suas duas partes componentes: a variação devido à manipulação experimental (SS_M) e a variação devido aos fatores não-sistemáticos (SS_R) (veja a Figura 8.2). As comparações planejadas levam essa lógica um passo além, dividindo a variação devido ao experimento em suas partes

³ Os termos “comparações” e “contrastes” são utilizados da mesma forma.

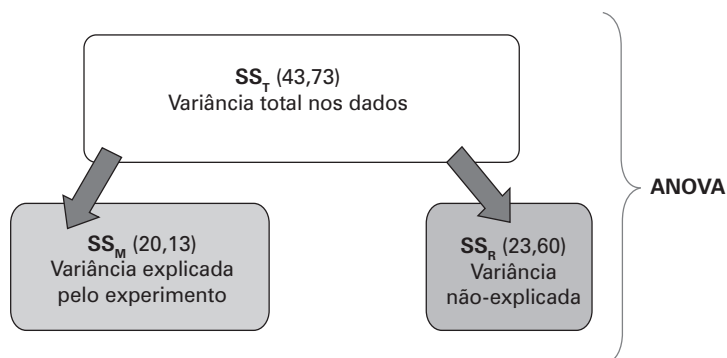


Figura 8.2 Partição da variância na ANOVA.

componentes (veja a Figura 8.3). As comparações exatas que são executadas dependem das hipóteses que se quer testar. A Figura 8.3 mostra a situação na qual a variância experimental é particionada de forma a mostrar quanto da variação é de responsabilidade das duas doses do medicamento comparadas com a condição placebo (*contraste1*). Então a variação explicada pela ingestão do Viagra é dividida para ver quanto dela é explicada pela alta e pela baixa dosagem (*contraste2*).

Em geral, os estudantes têm dificuldade de entender a noção de comparações planejadas, mas existem várias regras que podem ajudá-lo a saber o que fazer. É importante lembrar que estamos partindo uma parte da variação em porções independentes menores. Isso significa várias coisas. Primeiro, se um grupo é isolado em uma comparação, ele não deve reaparecer em outra comparação. Assim, na Figura 8.3, o contraste 1 envolve comparar o grupo-placebo com os grupos-experimentais; como o grupo-placebo foi

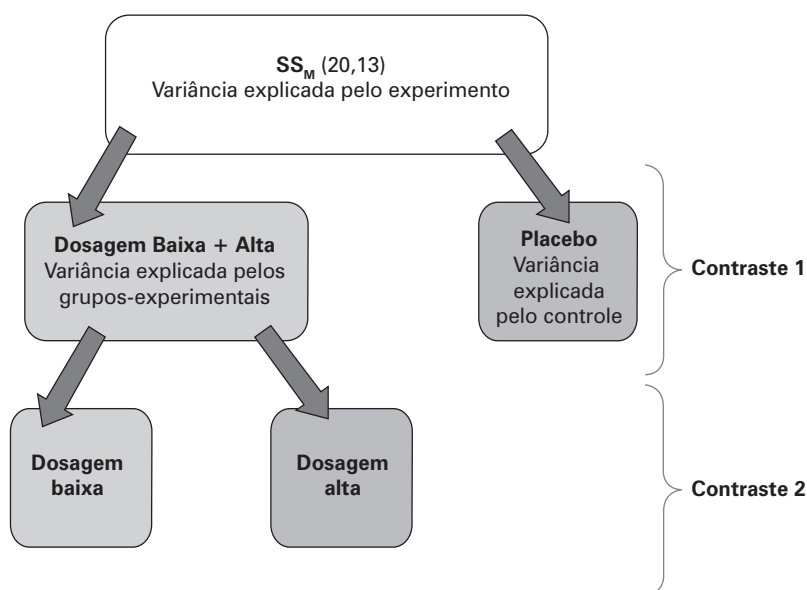


Figura 8.3 Partição da variância experimental em componentes.

isolado, ele não deve ser incorporado em qualquer outro contraste. Imagine a divisão da variância como fatiar um bolo. Você inicia com o bolo inteiro (a soma total dos quadrados) e depois corta o bolo em dois (SS_M e SS_R). Você então pega o pedaço que representa a SS_M e corta novamente em fatias menores. Uma vez que você tenha cortado uma fatia do bolo você não pode colocá-la de volta no bolo original, mas pode dividi-la em partes ainda menores. Da mesma forma, uma vez que a variância tenha sido dividida, ela não pode ser anexada a qualquer outra parte da variância; ela só pode ser novamente dividida em partes menores. Essa conversa sobre bolo está me deixando com fome, mas espero que tenha servido para ilustrar o que quero dizer.

Como uma segunda dica para selecionar contrastes, a regra da independência dos contrastes que acabei de explicar (a divisão do bolo!) significa que você deve sempre ter um contraste a menos que o número de grupos (assim, existirão $k - 1$ contrastes, onde k é o número de condições que estamos comparando). Finalmente, cada contraste deve comparar somente duas partes de variâncias. Essa regra final é a que permite que tiremos conclusões firmes sobre o que os contrastes nos dizem. A razão F nos informa que algumas das nossas médias diferem, mas não quais, e se executássemos um contraste em mais de duas partes de variâncias, enfrentaríamos o mesmo problema. Comparando apenas duas partes de variâncias, podemos estar seguros de que um resultado significativo representa uma diferença entre essas duas porções de variação experimental.

Em muitas pesquisas nas ciências sociais utilizamos pelo menos uma condição de controle e na grande maioria dos delineamentos experimentais previmos que as condições experimentais irão diferir da condição de controle (ou condições). Assim, **uma boa dica quando formos planejar comparações é comparar todos os grupos de condições experimentais com o grupo ou grupos-controle como sua primeira comparação.** Uma vez que tenhamos feito a primeira comparação, qualquer outra remanescente irá depender de quais grupos experimentais você prevê que irão diferir.

Para ilustrar esses princípios, a Figura 8.4 e a Figura 8.5 mostram os contrastes que podem ser feitos com um experimento de quatro grupos. Observe que nos dois cenários existem três possíveis comparações (uma a menos do que o número de grupos). Também, cada contraste compara somente duas porções da variância. E mais, nos dois cenários o primeiro contraste é o mesmo: os grupos-experimentais são comparados contra o grupo-controle (ou grupos). Na Figura 8.4 existe somente uma condição de controle e assim essa porção da variância é utilizada somente no primeiro contraste (porque ela não pode ser mais dividida). Na Figura 8.5 existem dois grupos-controle e assim a porção da variância devido às condições de controle (contraste 1) pode ser dividida novamente como forma de verificar se os escores nos dois grupos-controle diferem entre si (contraste 3).

Na Figura 8.4, o primeiro contraste contém uma parte da variância que é acarretada pelos três grupos-experimentais e essa porção da variância é dividida verificando primeiramente se os grupos E1 e E2 diferem do E3 (contraste 2). É igualmente válido utilizar o contraste 2 para comparar os grupos E1 e E3 com E2, ou comparar os grupos E2 e E3 com E1. As comparações exatas que você escolher dependerão das suas hipóteses. Para o contraste 2 na Figura 8.4 ser válido precisamos ter uma boa razão para esperar que o grupo E3 seja diferente dos outros dois grupos. A terceira comparação na Figura 8.4 depende da que foi escolhida para o contraste 2. O contraste 2 necessariamente deve envolver a comparação de dois grupos-experimentais contra um terceiro e os grupos-experimentais escolhidos para serem combinados devem ser separados na comparação final. Finalmente, observe nas Figuras 8.4 e 8.5 que quando um grupo foi isolado em uma comparação ele não é mais utilizado em qualquer contraste posterior.

Quando executamos um contraste planejado, comparamos “porções” de variância e essas porções às vezes consistem em vários grupos. Às vezes é confuso entender exatamente o que esses contrastes nos informam. Quando

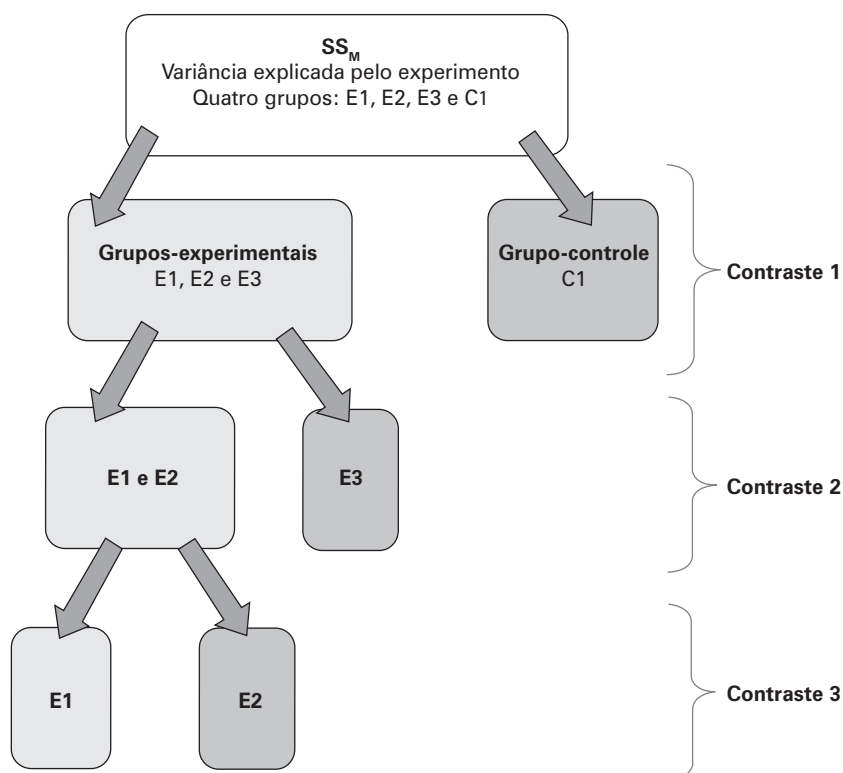


Figura 8.4 Partição da variância para comparações planejadas em um experimento de quatro grupos onde um é o grupo-controle.

O que um contraste planejado me informa?



you planeja um contraste que compara vários grupos a outro grupo, você está comparando as médias dos grupos em uma porção com a média do grupo em outra porção. Como exemplo, para os dados do Viagra eu sugeri que um primeiro contraste apropriado seria a comparação dos

grupos das duas dosagens (alta e baixa) com o grupo-placebo. As médias dos grupos são 2,20 (placebo), 3,20 (de baixa dosagem) e 5,00 (alta dosagem); assim, a primeira comparação – que compara os dois grupos-experimentais contra o placebo – é comparar 2,20 (média do grupo-placebo) com a média dos outros dois grupos ($(3,20 + 5,00)/2 = 4,10$). Se esse primeiro contraste for significativo, podemos

concluir que 4,10 é significativamente maior do que 2,20. Em termos do experimento, isso nos informa que a média desses grupos-experimentais é significativamente diferente da média do grupo-controle. Logicamente, isso significa que, se os erros padrão são os mesmos, o grupo-experimental com a média mais alta (o grupo de alta dosagem) será significativamente diferente da média do grupo-placebo. Contudo, o grupo-experimental com a média mais baixa (grupo de baixa dosagem) pode não diferir necessariamente do grupo-placebo; precisamos utilizar uma última comparação para esclarecer essas condições experimentais. Para os dados do Viagra, a comparação final é verificar se as duas condições experimentais diferem (isto é, as médias dos grupos de alta e baixa dosagem são significativamente diferentes). Se essa comparação for significativa,

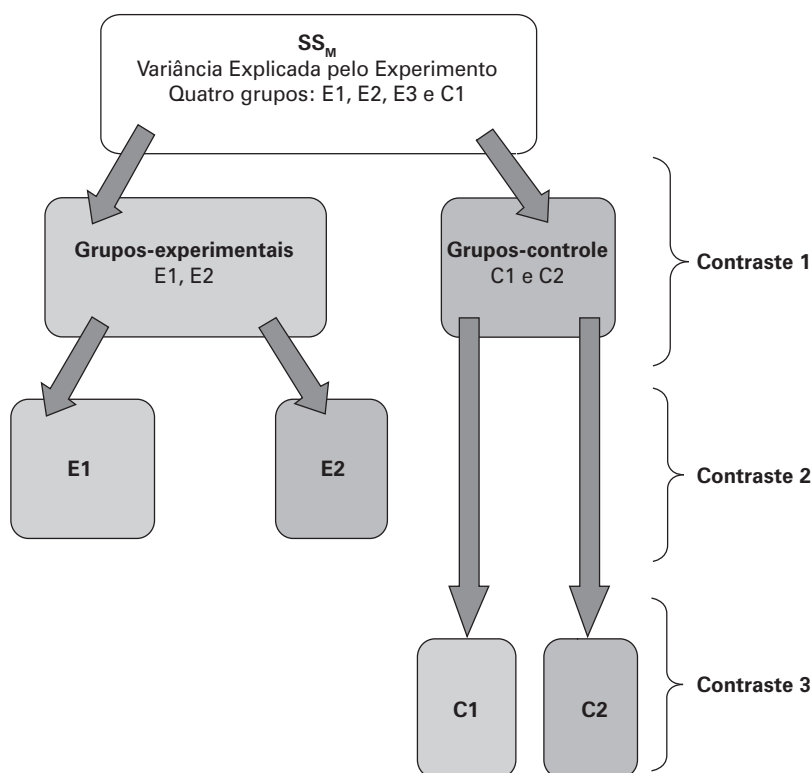


Figura 8.5 Partição da variância para comparações planejadas em um experimento de quatro grupos em que dois são grupos-controle.

podemos concluir que uma alta dose de Viagra afeta significativamente mais a libido do que uma baixa dosagem do medicamento. Se a comparação não for significativa, aceitamos que a dosagem do Viagra não afetou a libido. No primeiro caso, provavelmente as duas situações (alta e baixa dosagem) afetam a libido mais do que o placebo; já a segunda situação implica que uma dosagem baixa pode não ser diferente do que tomar o placebo. Contudo, a palavra *implica* é importante aqui: é possível que o grupo de baixa dosagem não difira do grupo-placebo. Para verificarmos isso, devemos executar um teste *post hoc*.

8.2.10.2 Definindo contrastes ponderados ③

Espero que a essa altura você já tenha alguma ideia de como planejar as suas com-
parações (se o seu cérebro ainda não explodiu!).

Adoraria dizer a você que o difícil já passou e que o SPSS irá magicamente executar as comparações que você selecionou, mas isso não seria verdade. Para que o SPSS execute as comparações planejadas, você precisa informar quais grupos deverão ser comparados, e fazer isso pode ser complicado. De fato, quando executamos contrastes atribuímos valores a certas variáveis no modelo de regressão (desculpe, mas vou ter que falar sobre regressão de novo) – da mesma forma que fizemos quando utilizamos a codificação *dummy* (auxiliar) para a ANOVA principal. Para executar os contrastes, simplesmente atribuímos certos valores a variáveis auxiliares no modelo de regressão. Enquanto anteriormente definimos os grupos-experimentais atribuindo às variáveis auxiliares os valores 0 e 1, quando executamos

os contrastes utilizamos valores diferentes para especificar quais grupos queremos comparar. Os coeficientes resultantes no modelo de regressão (b_2 e b_1) representam as comparações nas quais estamos interessados. Os valores atribuídos as variáveis *dummy* (auxiliares) são conhecidos como pesos (ponderações).

Esse procedimento é confuso, mas existem algumas regras básicas para atribuir valores às variáveis auxiliares a fim de obtermos as comparações desejadas. Explicarei essas regras antes de revelar como o processo funciona de fato. Lembre da seção anterior quando você leu essas regras e pense sobre o significado de “porção” de variação!

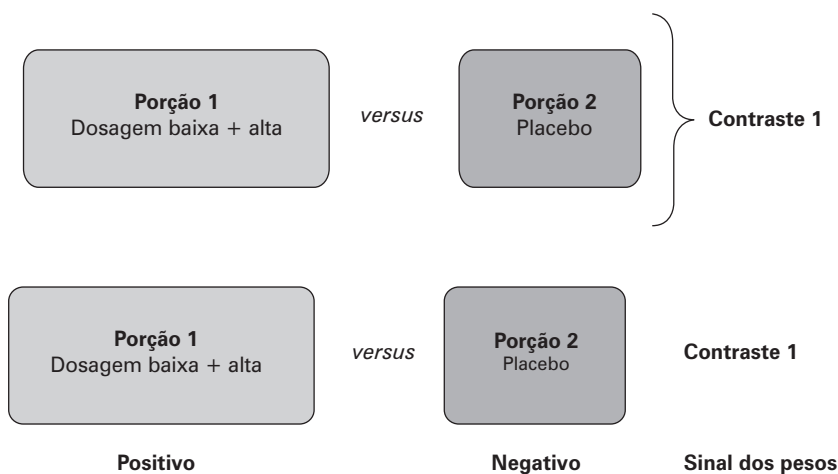
- **Regra 1:** Escolha comparações sensíveis. Lembre que você quer comparar somente duas porções de variação e que se um grupo é isolado em uma comparação, então esse grupo deve ser excluído de qualquer contraste subsequente.
- **Regra 2:** Grupos codificados com ponderações positivas serão comparados com grupos codificados com ponderações negativas. Assim, atribua a uma porção da variação ponderações positivas e à porção oposta, ponderações negativas.
- **Regra 3:** A soma das ponderações (pesos) para uma comparação deve ser zero: se você somar os pesos para um determinado contraste o resultado deve ser zero.

- **Regra 4:** Se um grupo não está envolvido em uma comparação, atribua automaticamente o peso 0. Se for dado o peso 0 a um grupo, elimine esse grupo de qualquer outro cálculo.
- **Regra 5:** Para um dado contraste, os pesos atribuídos aos grupos em uma porção da variação deve ser igual ao número de grupos na porção contrária da variação.

Vamos usar algumas dessas regras a fim de determinar as ponderações para os dados do Viagra. A primeira comparação que escolhemos foi comparar os dois grupos-experimentais contra o grupo-controle:

Portanto, a primeira porção da variação contém os dois grupos-experimentais e a segunda parte contém somente o grupo-placebo. A regra 2 diz que devemos atribuir a uma porção ponderações positivas e a outra, negativas. Não importa a ordem em que isso é feito, mas por conveniência vamos atribuir à porção 1 uma ponderação positiva e a 2, negativa.

Utilizando a regra 5 os pesos atribuídos aos grupos na porção 1 devem ser equivalentes aos valores atribuídos aos grupos na porção 2. Existe somente um grupo na porção 2, assim, atribuímos a cada grupo na porção 1 o peso de 1. Da mesma forma, atribuímos ao grupo na porção 2 um peso igual ao número de grupos na porção 1. Há dois grupos na porção 1, portanto, atribuímos ao grupo-placebo peso 2.



Depois, combinamos os sinais dos pesos com a magnitude e chegamos a -2 (placebo), 1 (dosagem baixa) e 2 (dosagem alta):

A regra 3 declara que para um dado contraste, a soma dos pesos deve ser zero; se as regras 2 e 5 forem seguidas, essa regra sempre será satisfeita (se você não seguiu adequadamente essas regras, isso irá aparecer quando os pesos forem somados). Vamos verificar somando os pesos: $1 + 1 - 2 = 0$.

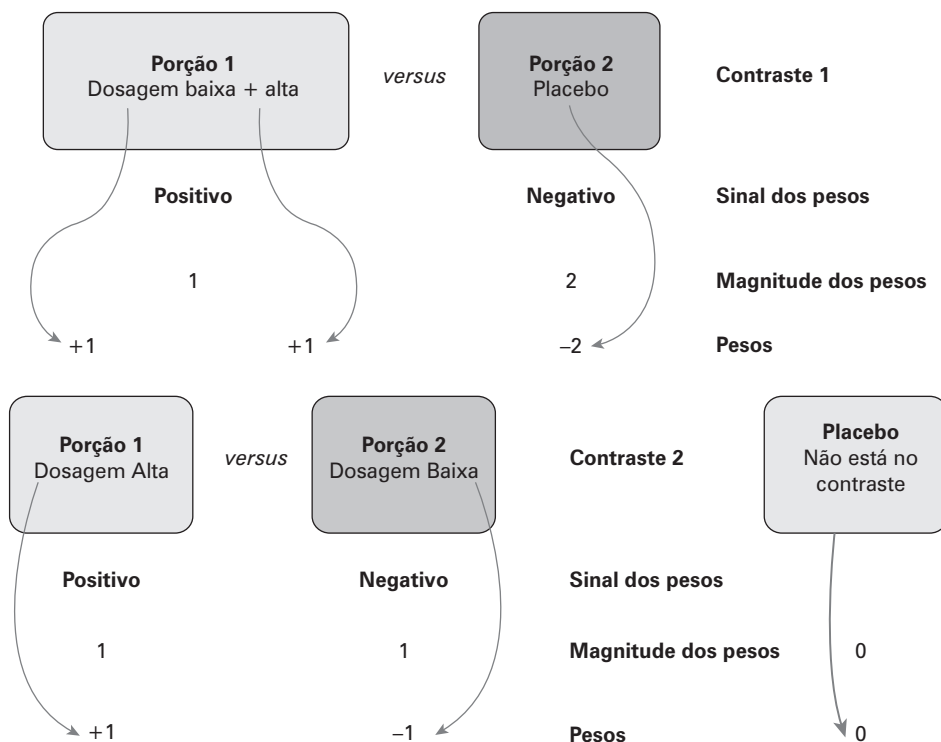
O segundo contraste será comparar os dois grupos-experimentais e assim iremos ignorar o grupo-placebo. A regra 4 nos diz que devemos atribuir a esse grupo uma ponderação igual a 0 (porque dessa forma iremos eliminá-lo de qualquer cálculo). Restam ainda duas porções de variação: a porção 1, que contém a variação do grupo de baixa dosagem, e a porção 2, que contém a variação do grupo de alta dosagem. Seguindo as regras 2 e 5 devemos atribuir a um grupo o peso $+1$ e ao outro grupo o peso -1 . O grupo-controle é ignorado (e assim ele terá


um peso de 0). Se adicionarmos os pesos do contraste 2, veremos que a soma deverá ser novamente zero, isto é, $+1 - 1 + 0 = 0$.

Os pesos para cada contraste são codificados para duas variáveis *dummy* (auxiliares) na equação 8.2. Assim, esses códigos podem ser utilizados num modelo de regressão múltipla no qual b_2 representa o contraste 1 (comparando os grupos-experimentais ao grupo-controle), b_1 representa o contraste 2 (comparando os grupos de alta e baixa dosagem) e b_0 é a média geral (equação (8.9)):

$$\text{Libido}_i = b_0 + b_1 \text{Contraste}_1 + b_2 \text{Contraste}_2 \quad (8.9)$$

Cada grupo é especificado agora não pelo esquema de códigos 0 e 1 que foi inicialmente utilizado, mas pelo esquema de códigos dos dois contrastes. Um código de -2 para o primeiro contraste e um código de 0 para o segundo contraste identificam os participantes no grupo-placebo. Da mesma forma, o grupo de alta dosagem





O que são contrastes ortogonais?

pode ser identificado pelo código 1 nas duas variáveis e o código de baixa dosagem tem o código 1 para um contraste e um código -1 para o outro (veja a Tabela 8.4).

É importante que os pesos para as comparações somem zero porque isso irá assegurar que você está comparando duas porções únicas de variação. Desse modo, o SPSS pode executar um teste *t*. Ainda mais importante é que quando você multiplica os pesos para um grupo em particular a soma desses produtos também deve ser zero (veja a última coluna da Tabela 8.4). Se a soma dos produtos é zero, podemos estar seguros de que os contrastes são **ortogonais** ou *independentes*. É importante para a interpretação que os contrastes sejam ortogonais. Quando utilizamos a codificação com variáveis auxiliares e executamos uma regressão nos dados do Viagra, comentei que não podíamos olhar para os testes *t* individuais executados nos coeficientes da regressão em virtude da inflação na taxa de erro de conjunto (veja a Seção 8.2.10 e a Safda 8.1 do SPSS). Contudo, se os contrastes são independentes, os testes *t* executados nos coeficientes *b* são, também, independentes e assim os valores *p* resultantes não se correlacionam. Pode parecer difícil assegurar que os pesos escolhidos para os contrastes sejam adequados para garantir a independência, mas contanto que as regras sejam seguidas você sempre terá em mãos um conjunto de comparações *ortogonais*. Portanto, é necessário sempre verificar se a soma dos produtos dos pesos dos contrastes é zero; se isso não ocorrer, volte para a lista de regras a fim de descobrir o que não deu certo (veja a última coluna da Tabela 8.4).



Mencionei anteriormente que quando utilizamos os códigos das variáveis auxiliares nos contrastes em um modelo de regressão, os valores *b* representam as diferenças entre as médias que os contrastes planejaram testar. Embora seja razoável acreditar em mim nessa questão, para os estudantes mais avançados quero me dar ao trabalho de mostrar como o modelo de regressão funciona. Assim, a próxima parte é para os corajosos; quem tem fobia a equações deve seguir direto para a próxima seção. Quando executamos contrastes planejados, o intercepto b_0 é igual à média geral (isto é, o valor previsto pelo modelo quando a pertinência a um grupo não é conhecida):

$$b_0 = \text{média geral} = \frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}}}{3}$$

Grupo-placebo: Se utilizarmos os códigos dos contrastes para o grupo-placebo (veja a Tabela 8.4), o valor previsto da libido será igual à média do grupo-placebo. A equação de regressão pode, portanto, ser expressa como:

$$\begin{aligned} \text{Libido}_i &= b_0 + b_1 \text{Contraste}_1 + b_2 \text{Contraste}_2 \\ \bar{X}_{\text{Placebo}} &= \left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}}}{3} \right) \\ &\quad + (-2b_1) + (b_2 \times 0) \end{aligned}$$

Rearranjando essa equação e multiplicando tudo por 3 (para eliminar a fração) obtemos:

$$\begin{aligned} 2b_1 &= \left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}}}{3} \right) - \bar{X}_{\text{Placebo}} \\ 6b_1 &= \bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}} - 3\bar{X}_{\text{Placebo}} \\ 6b_1 &= \bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} - 2\bar{X}_{\text{Placebo}} \end{aligned}$$

Tabela 8.4 Contrastes ortogonais para os dados do Viagra

Grupo	Variável auxiliar 1 (Contraste ₁)	Variável auxiliar 2 (Contraste ₂)	Produto Contraste ₁ × Contraste ₂
Placebo	-2	0	0
Baixa dosagem	1	-1	-1
Alta dosagem	1	1	1
Total	0	0	0

Podemos, então, dividir tudo por 2 para reduzir a equação à sua forma mais simples:

$$3b_1 = \left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}}}{2} \right) - \bar{X}_{\text{Placebo}}$$

$$b_1 = \frac{1}{3} \left[\left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}}}{2} \right) - \bar{X}_{\text{Placebo}} \right]$$

Essa equação mostra que b_1 representa a diferença entre as médias dos dois grupos-experimentais e a média do grupo-controle:

$$3b_1 = \left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}}}{2} \right) - \bar{X}_{\text{Placebo}}$$

$$= \frac{5 + 3,2}{2} - 2,2$$

$$= 1,9$$

Planejamos que o contraste 1 representasse a diferença entre a média dos grupos-experimentais e o grupo-controle, assim, deve estar claro agora que b_1 representa essa diferença. Os mais observadores devem ter notado que b_1 em vez de ser a verdadeira diferença entre as médias dos grupos-experimentais e a do grupo-controle é de fato um terço dessa diferença ($b_1 = 1,90/3 = 0,633$). A razão para essa divisão é que o erro de conjunto é controlado fazendo com que o coeficiente de regressão seja igual à diferença real dividida pelo número de grupos no contraste (no caso, 3).

Grupo de alta dosagem: Para a situação em que os códigos para o grupo de alta dosagem (veja a Tabela 8.4) são utilizados, o valor previsto da libido é a média para o grupo de alta dosagem; desse modo, a equação de regressão torna-se:

$$\text{Libido}_i = b_0 + b_1 \text{Contraste}_1 + b_2 \text{Contraste}_2$$

$$\bar{X}_{\text{Alta}} = b_0 + (b_1 \times 1) + (b_2 \times 1)$$

$$b_2 = \bar{X}_{\text{Alta}} - b_1 - b_0$$

Já sabemos o que b_1 e b_0 representam e assim colocamos esses valores na equação e multiplicamos por 3 para nos livrarmos de parte das frações:

$$b_2 = \bar{X}_{\text{Alta}} - b_1 - b_0$$

$$b_2 = \bar{X}_{\text{Alta}} - \left\{ \frac{1}{3} \left[\left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}}}{2} \right) - \bar{X}_{\text{Placebo}} \right] \right\}$$

$$- \left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}}}{3} \right)$$

$$3b_2 = 3\bar{X}_{\text{Alta}} - \left[\left(\frac{\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}}}{2} \right) - \bar{X}_{\text{Placebo}} \right]$$

$$- (\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}})$$

Se multiplicarmos tudo por 2 para nos livrarmos da outra fração, retirarmos os colchetes e parênteses e fizermos simplificações, teremos:

$$6b_2 = 6\bar{X}_{\text{Alta}} - (\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} - 2\bar{X}_{\text{Placebo}})$$

$$- 2(\bar{X}_{\text{Alta}} + \bar{X}_{\text{Baixa}} + \bar{X}_{\text{Placebo}})$$

$$6b_2 = 6\bar{X}_{\text{Alta}} - \bar{X}_{\text{Alta}} - \bar{X}_{\text{Baixa}} + 2\bar{X}_{\text{Placebo}}$$

$$- 2\bar{X}_{\text{Alta}} - 2\bar{X}_{\text{Baixa}} - 2\bar{X}_{\text{Placebo}}$$

$$6b_2 = 3\bar{X}_{\text{Alta}} - 3\bar{X}_{\text{Baixa}}$$

Finalmente, podemos dividir toda a equação por seis para descobrir o que b_2 significa:

$$b_2 = \frac{1}{2}(\bar{X}_{\text{Alta}} - \bar{X}_{\text{Baixa}})$$

Planejamos o contraste 2 para representar a diferença entre os grupos-experimentais:

$$\bar{X}_{\text{Alta}} - \bar{X}_{\text{Baixa}} = 5 - 3,2 = 1,8$$

Deve estar claro agora que b_2 representa essa diferença. Novamente, em vez de obtermos o valor real da diferença entre as médias dos grupos-experimentais, b_2 é, de fato, metade dessa diferença ($1,8/2 = 0,9$). O erro de conjunto é novamente controlado, fazendo o coeficiente de regressão igual à diferença real dividida pelo número de grupos envolvidos no contraste (nesse caso, 2).

Para ilustrar esses princípios, criei um arquivo denominado **Contrast.sav** em que os dados do Viagra estão codificados com base no esquema de códigos utilizado nessa seção. Execute uma análise de regressão múltipla nesses dados utilizando a variável **libido** como saída e usando as variáveis *dummies* (auxilia-

res) **dummy1** e **dummy2** como variáveis predictoras (deixe todas as opções por padrão). A ANOVA principal para o modelo é a mesma de quando o código das variáveis auxiliares (*dummies*) foram utilizados; contudo, os coeficientes de regressão agora são outros.



A Saída 8.2 do SPSS mostra os resultados dessa regressão. Observe que o intercepto é a média geral, 3,467 (viu, eu não estou mentindo!). Segundo,

o coeficiente da regressão para o contraste 1 é um terço da diferença entre a média das condições experimentais e a condição de controle (veja a seguir). Finalmente, o coeficiente da regressão para o segundo contraste é metade da diferença entre as médias dos grupos experimentais (veja a seguir). Assim, quando uma comparação planejada é realizada na ANOVA, um teste *t* é conduzido comparando a média de uma porção da variação com a média de uma porção diferente. A partir das significâncias dos testes *t* podemos ver que nossos grupos experimentais foram significativamente diferentes do grupo-controle ($p < 0,05$), mas que os grupos experimentais não foram significativamente diferentes ($p > 0,05$).

8.2.10.3 Comparações não-ortogonais ②

Gastei bastante tempo explicando como projetar comparações ortogonais apropriadas sem mencionar as possibilidades oferecidas pelos contrastes não-ortogonais. Contrastes

não-ortogonais são comparações que estão de alguma forma relacionadas e a melhor maneira de obtê-los é desobedecendo a regra número 1 da seção anterior. Utilizando a minha analogia do bolo, nas comparações não-ortogonais fatiamos o bolo e tentamos manter as fatias novamente juntas! Para o conjunto de dados do Viagra, por exemplo, um conjunto de contrastes não-ortogonais pode ter o mesmo contraste inicial (comparando os grupos-experimentais contra o grupo-controle), mas depois compara o grupo de baixa dosagem com o grupo-placebo. Isso viola a regra 1 porque o grupo-placebo está sozinho no primeiro contraste e foi utilizado novamente no segundo contraste. Os códigos para esse conjunto de contrastes são mostrados na Tabela 8.5; olhando para a última coluna fica claro que quando você multiplica e soma os códigos dos contrastes a soma não dá zero. Isso mostra que os contrastes não são ortogonais.

Nada existe de intrinsecamente errado em utilizarmos contrastes não-ortogonais. Contudo, se você escolher esse tipo de procedimento deve tomar muito cuidado em como irá interpretar os resultados. Com contrastes não-ortogonais as comparações feitas estão relacionadas, assim, a estatística teste resultante e o valor *p* estão correlacionados de alguma forma. Desse modo, você deve utilizar níveis de probabilidade mais conservadores

Os contrastes não-ortogonais são legítimos?



Saída 8.2 do SPSS

Coefficients^a (Coeficientes)

Model (Modelo)	Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	<i>t</i>	Sig.
	<i>B</i>	Std. Error (Erro padrão)	<i>Beta</i>		
1 (Constant) (Constante)	3.467	0.362		9.574	0.000
Dummy Variable 1 (Variável auxiliar 1)	0.633	0.256	0.525	2.474	0.029
Dummy Variable 2 (Variável auxiliary 2)	0.900	0.443	0.430	2.029	0.065

a. Dependent Variable: Libido (Variável dependente: Libido)

Tabela 8.5 Contrastes não-ortogonais para os dados do Viagra

Grupo	Variável auxiliar 1 (Contraste ₁)	Variável auxiliar 2 (Contraste ₂)	Produto Contraste ₁ × Contraste ₂
Placebo	-2	-1	2
Baixa dosagem	1	0	0
Alta dosagem	1	1	1
Total	0	0	3

para aceitar que um dado contraste é significativo (veja a Seção 8.2.11). Por essa razão, recomendo que você utilize comparações planejadas ortogonais sempre que possível.

8.2.10.4 Contrastes padrão ②

Embora em geral você projete seus próprios contrastes, existem contrastes especiais para certas situações. Alguns desses contrastes são ortogonais enquanto outros não. Muitos proce-

dimentos no SPSS permitem que você execute os contrastes mencionados nessa seção.

A Tabela 8.6 mostra os contrastes que estão disponíveis no SPSS para procedimentos tais como a regressão logística (veja a Seção 6.4.3), a ANOVA fatorial e a ANOVA de medidas repetidas (veja os Capítulos 9 e 10). Embora os códigos exatos não sejam fornecidos na Tabela 8.6, exemplos das comparações são feitos com as situações de três e de quatro grupos (onde os grupos são rotulados de 1, 2 e 3 e 1, 2,

Tabela 8.6 Contrastes padrão disponíveis no SPSS

Nome	Definição	Contraste	Três Grupos	Quatro Grupos
Desvios (Primeiro)	Compara o efeito de cada categoria (exceto a primeira) com o efeito experimental global	1	2 vs. (1, 2, 3)	2 vs. (1, 2, 3, 4)
		2	3 vs. (1, 2, 3)	3 vs. (1, 2, 3, 4)
		3		4 vs. (1, 2, 3, 4)
Desvios (último)	Compara o efeito de cada categoria (exceto a última) com o efeito experimental global	1	1 vs. (1, 2, 3)	1 vs. (1, 2, 3, 4)
		2	2 vs. (1, 2, 3)	2 vs. (1, 2, 3, 4)
		3		3 vs. (1, 2, 3, 4)
Simples (Primeiro)	Cada categoria é comparada com a primeira categoria	1	1 vs. 2	1 vs. 2
		2	1 vs. 3	1 vs. 3
		3		1 vs. 4
Simples (último)	Cada categoria é comparada com a última categoria	1	1 vs. 3	1 vs. 4
		2	2 vs. 3	2 vs. 4
		3		3 vs. 4
Repetido	Cada categoria (exceto a primeira) é comparada com a prévia	1	1 vs. 2	1 vs. 2
		2	2 vs. 3	2 vs. 3
		3		3 vs. 4
Helmert	Cada categoria (exceto a última) é comparada ao efeito médio de todas as categorias subsequentes	1	1 vs. (2, 3)	1 vs. (2, 3, 4)
		2	2 vs. 3	2 vs. (3, 4)
		3		3 vs. 4
Diferença (Helmert reversa)	Cada categoria (exceto a primeira) é comparada ao efeito médio de todas as anteriores	1	3 vs. (2, 1)	4 vs. (3, 2, 1)
		2	2 vs. 1	3 vs. (2, 1)
		3		2 vs. 1

3 e 4, respectivamente). Quando você codifica uma variável no editor de dados, o SPSS irá tratar o código de menor valor como o grupo 1, o próximo mais alto como grupo 2 e assim sucessivamente. Portanto, dependendo de quais comparações quer realizar, você deve codificar a variável de código de modo apropriado (e então utilizar a Tabela 8.6 como um guia sobre quais comparações o SPSS irá fazer). Os leitores espertos podem ter notado que alguns contrastes na Tabela 8.6 são ortogonais (os de Helmert e os contrastes diferenças) enquanto outros não são ortogonais (desvios, simples e repetidos). Você também pode ter notado que as comparações calculadas utilizando contrastes simples são as mesmas obtidas pela utilização das variáveis auxiliares na Tabela 8.2.

8.2.10.5 Contrastes polinomiais: análise da tendência ②

Um tipo de contraste deliberadamente omitido na Tabela 8.6 é o polinomial. Esse contraste testa tendências nos dados e na sua forma mais básica ele procura por tendências lineares (isto é, se a média dos grupos aumen-

ta de forma proporcional). Contudo, existem tendências mais complexas – como a quadrática, cúbica e quártica – que também podem ser examinadas. A Figura 8.6 mostra diagramas de tipos de tendência que podem existir em um conjunto de dados. A **tendência linear** representa uma mudança proporcional no valor da variável dependente ao longo de categorias ordenadas. Uma **tendência quadrática** é quando existe uma mudança na direção de uma linha (por exemplo, a linha é curva em um lugar). Um exemplo disso é quando um medicamento melhora o desempenho em uma tarefa no início, mas à medida que a dose aumenta o desempenho volta a cair. É óbvio que para achar uma tendência quadrática são necessários pelo menos três grupos (porque na situação de dois grupos não existem categorias suficientes da variável independente para que as médias da variável dependente mudem de uma forma e depois de outra). Uma **tendência cúbica** é quando existem duas mudanças na direção da tendência. Por exemplo, a média da variável dependente primeiro aumenta com as primeiras categorias da variável independente, depois ao longo de outras categorias as mé-

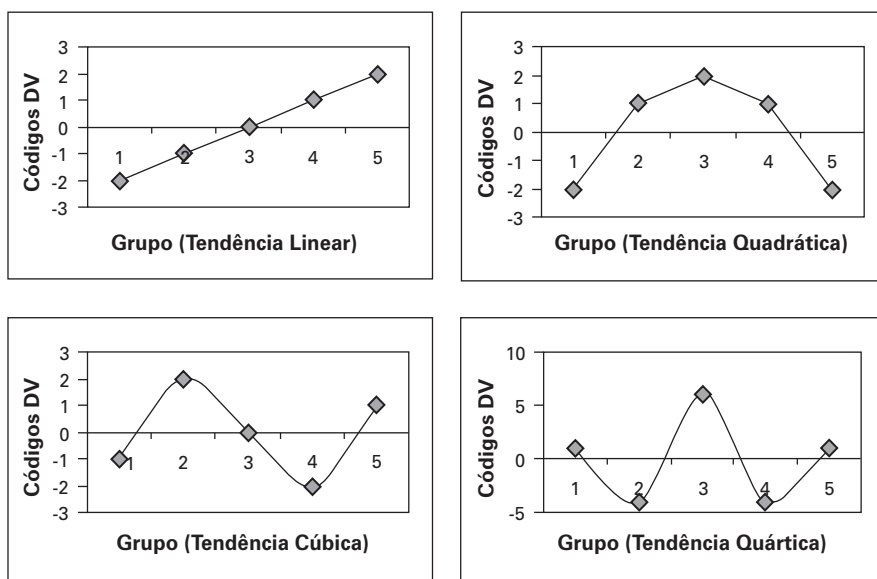


Figura 8.6 Tendências linear, quadrática, cúbica e quártica em cinco grupos.

dias diminuem, mas nas últimas categorias as médias se elevam novamente. Para ter duas mudanças na direção da média você deve ter pelo menos quatro categorias da variável independente. A última tendência com a qual você pode se deparar é a **quártica**, que apresenta três mudanças de direção (portanto, serão necessárias pelo menos cinco categorias da variável independente).

Tendências polinomiais devem ser examinadas em conjuntos de dados em que faz sentido ordenar as categorias da variável independente (por exemplo, se você administrou cinco doses de uma droga, faz sentido examinar as cinco doses na sua ordem de magnitude). Para os dados do Viagra existem somente três grupos, desse modo, podemos esperar encontrar somente uma tendência linear ou quadrática (e não teria sentido testar para tendências de ordens mais altas).

Cada uma dessas tendências apresenta um conjunto de códigos para as variáveis auxiliares no modelo de regressão, portanto, faremos o mesmo que foi feito para os contrastes planejados exceto que os códigos já foram escolhidos para representar o tipo de tendência de interesse. De fato, os diagramas na Figura 8.6 foram construídos traçando os valores dos códigos para os cinco grupos. Na verdade, se você adicionar códigos para uma dada tendência, a soma será igual a zero, e se os códigos forem multiplicados você vai encontrar que a soma dos produtos também será igual a zero. Dessa forma, esses contrastes são ortogonais. A vantagem desses contrastes é que não é necessário construir uma codificação própria, pois esses códigos já existem.

8.2.11 Procedimentos *post hoc* ②

Frequentemente, você não tem previsões específicas *a priori* sobre os dados coletados e, em vez disso, está interessado em explorar os dados para verificar a existência de quaisquer possíveis diferenças entre médias. Esse procedimento é, às vezes, denominado *mineração de dados* ou *análise de dados exploratória*. Eu acho que esses dois termos têm uma conotação de “manipulação” dos dados e prefiro pensar

nesses procedimentos como “encontrar as diferenças que eu teria previsto se fosse esperto o suficiente”.

Testes *post hoc* consistem em **comparações em pares** planejadas para comparar todas as diferentes combinações dos grupos sendo testados. Assim, é como tomar todos os pares de grupos possíveis e executar um teste *t* em cada um. Pode parecer idiota explicar isso (mas eu sou um pouco idiota) considerando o que eu já disse sobre o problema do crescimento do erro de conjunto. Contudo, as comparações aos pares controlam o erro de conjunto corrigindo o nível de significância de cada teste de forma que o erro do Tipo I total (α) ao longo de todas as comparações permaneça ao nível especificado (0,05). Existem várias formas de controlar o erro de conjunto. A mais popular (e mais fácil) é dividir o valor de α pelo número de comparações, assegurando, dessa forma, que o erro do Tipo I acumulado fique abaixo do valor especificado (0,05). Portanto, se forem executados 10 testes, utilizamos 0,005 como nosso critério de significância. Esse método é conhecido como **correção de Bonferroni** em homenagem a Carlo Bonferroni (Figura 8.7). Existe uma contrapartida no controle da taxa de erro de conjunto que é a perda de poder estatístico. Isso quer dizer que a probabilidade de rejeitar um efeito que existe aumenta (o erro do Tipo II). Se formos conservadores na taxa de erro do Tipo I para cada comparação, aumentamos a chance de deixar passar uma diferença genuína nos dados.

Contudo, quando consideramos procedimentos *post hoc* precisamos levar em conta três coisas: (1) o teste controla a taxa de erro do Tipo I; (2) o teste controla a taxa de erro do Tipo II (isto é, o teste apresenta um bom poder estatístico); e (3) o teste é confiável quando as suposições da ANOVA são violadas?

Eu adoraria discutir todos os detalhes tediosos dos tipos de testes *post hoc*, mas não há razão para tanto. Existem alguns textos excelentes disponíveis para quem quer saber mais (por exemplo, Toothaker, 1993; Klokars e Sax, 1986) e o SPSS também fornece 18 tipos de teste *post hoc*, portanto, eu teria que des-



Figura 8.7 Carlo Bonferroni, antes que a fama da sua correção o levasse à bebida, às drogas e ao fanatismo pela estatística.

matar muitos quilômetros de floresta para explicá-los. O importante é que você saiba quais testes *post hoc* executam melhor de acordo com os três critérios mencionados acima.

8.2.11.1 Procedimentos *post hoc* e erros do Tipo I (α) e Tipo II (β) ②



A taxa de erro do Tipo I e o poder estatístico de um teste estão conectados. Portanto, existe sempre uma contrapartida: se o teste é conservador (a probabilidade de erro do Tipo I é pequena), é provável

que ele tenha um baixo poder estatístico (a probabilidade de erro do Tipo II será alta). Assim, é importante que os procedimentos de comparações múltiplas controlem a taxa de erro do Tipo I, mas sem uma perda substancial no poder. Se um teste é muito conservador, provavelmente iremos rejeitar diferenças entre médias que são, na realidade, significativas.

A comparação de pares MDS (Menor Diferença Significativa) (*LDS – Least-Significant Difference*) não tenta controlar o erro do Tipo I e é equivalente à execução de múltiplos testes t. A única diferença é que a MDS requer que a ANOVA global seja significativa. O pro-

cedimento ENK (Estudentizado de Newman-Keuls) também é um teste bastante liberal e não controla o erro de conjunto. Os testes de Bonferroni e Tukey controlam a taxa de erro do Tipo I muito bem, mas são testes conservadores (eles têm poder estatístico deficiente). O de Bonferroni apresenta um poder maior quando o número de comparações é pequeno, já o de Tukey é mais poderoso quando testa uma grande quantidade de médias. O teste de Tukey geralmente tem um poder maior do que os testes de Dunn e Scheffé. O procedimento de Ryan, Einot, Gabriel e Welsch Q (REGWQ) apresenta um bom poder e um controle satisfatório do erro do Tipo I. De fato, quando queremos testar todos os pares de médias esse procedimento é provavelmente o melhor. Contudo, quando os tamanhos dos grupos são diferentes, esse procedimento não deverá ser utilizado.

8.2.11.2 Procedimentos *post hoc* e violações das hipóteses do teste ②

A maioria das pesquisas sobre testes *post hoc* tem se debruçado no problema do desempenho do teste quando os tamanhos dos grupos são diferentes (delineamento desbalanceado), quando as variâncias das populações são muito distintas e quando os dados não são normalmente distribuídos. A boa notícia é que muitos procedimentos de comparações múltiplas tem um bom desempenho sob pequenos desvios da normalidade. A má notícia é que elas apresentam um mau desempenho quando os tamanhos dos grupos são diferentes e quando as variâncias populacionais são desiguais.



O teste GT2 de Hochberg e de Gabriel foram projetados para lidar com situações em que os tamanhos amostrais são diferentes. O procedimento de

Gabriel é geralmente mais poderoso, mas pode se tornar muito liberal quando os tamanhos das amostras são bem diferentes. O GT2 de Hochberg não é confiável quando as variâncias populacionais são diferentes e deve ser utilizado somente quando você estiver seguro de que esse não é o caso. Existem várias comparações

múltiplas especialmente projetadas para situações em que as variâncias populacionais diferem. O SPSS fornece quatro opções para essa situação. O T2 de Tamhane, T3 de Dunnett, Games-Howell e C de Dunnett. O T2 de Tamhane é conservador e o C e T3 de Dunnett mantêm um controle muito rígido do erro do Tipo I. O procedimento de Games-Howell é o mais poderoso, mas pode ser liberal quando os tamanhos das amostras forem pequenos. Contudo, o de Games-Howell é mais preciso quando os tamanhos amostrais são diferentes.

8.2.11.3 Resumo dos Procedimentos *post hoc* ②

A escolha dos procedimentos de comparações múltiplas irá depender da situação que você tiver e se o mais importante é manter um controle restrito sobre o erro de conjunto ou ter um grande poder estatístico. Contudo, algumas orientações gerais podem ser colocadas (veja Toothaker, 1993). Quando os tamanhos das amostras são iguais e você tem certeza de que as variâncias populacionais são semelhantes, utilize o REGWQ ou Tukey, pois ambos têm um bom poder e um bom controle sobre o erro do Tipo I. O teste de Bonferroni é geralmente conservador, mas se você quer garantia sobre a taxa de erro do Tipo I, deve utilizá-lo. Se os tamanhos amostrais são levemente diferentes, use o procedimento de Gabriel porque ele tem um grande poder, mas se os tamanhos amostrais são muito diferentes, o recomendável é o GT2 de Hochberg. Se você tem dúvidas quanto à igualdade das variâncias populacionais, utilize o procedimento de Games-Howell porque ele oferece, em geral, o melhor desempenho. Recomendo que você execute o procedimento de Games-Howell junto com qualquer outro teste que você possa ter selecionado, em virtude da incerteza em relação à igualdade das variâncias populacionais.

Embora essas orientações gerais forneçam um guia a ser seguido, fique atento a outros procedimentos disponíveis e quando eles podem ser úteis (por exemplo, o teste de Dunnett é a única comparação múltipla que permite que você teste médias contra uma média controle).

8.3 REALIZANDO A ANOVA DE UM FATOR (ONE-WAY) NO SPSS ②

Espero que vocês gostem da teoria por trás da ANOVA, pois vamos colocá-la em prática executando uma ANOVA com os dados do Viagra. Como ocorreu com o teste *t* independente, precisamos entrar com os dados no editor utilizando uma variável de código para especificar a qual dos três grupos os dados pertencem. Assim, os dados devem ser dispostos em duas colunas (uma denominada **dose** que especifica quanto Viagra um dado participante tomou, e outra denominada **libido** que indica a libido da pessoa na semana seguinte). Os dados estão no arquivo **Viagra.sav**, mas eu recomendo que você os digite para praticar a entrada de dados. Codifiquei a variável de grupo de modo que 1 = placebo, 2 = baixa dosagem e 3 = alta dosagem (veja a Seção 2.4.4).

Para realizar uma ANOVA de um fator, precisamos primeiro acessar a caixa de diálogo principal utilizando o caminho de menu **Analyze⇒Compare Means⇒One-way Anova** (Analisar⇒Comparar médias⇒ANOVA de um fator) (Figura 8.8). Essa caixa de diálogo apresenta um espaço onde você pode listar uma ou mais variáveis dependentes e um segundo espaço para especificar a variável de grupo ou *fator*. Fator é outro termo para a variável independente e não deve ser confundido com os fatores que serão apresentados quando você aprender sobre análise de fatores. Uma coisa que eu não gosto no SPSS é que em vários procedimentos, tais como a ANOVA de um fator, o programa estimula o usuário a executar múltiplos testes, o que, como vimos, não é recomendável. Por exemplo, nesse procedimento você pode especificar diversas variáveis dependentes nas quais conduzir diferentes ANOVAs. Na verdade, se você tivesse medido diversas variáveis dependentes (digamos que você tivesse medido não apenas a libido, mas a excitação psicológica e a ansiedade também) seria preferível analisar esses dados utilizando a MANOVA em vez de tratar cada medida dependente de forma separada (veja o Capítulo 14). Para os dados do Viagra precisamos selecionar somente **libido** da lista de variáveis

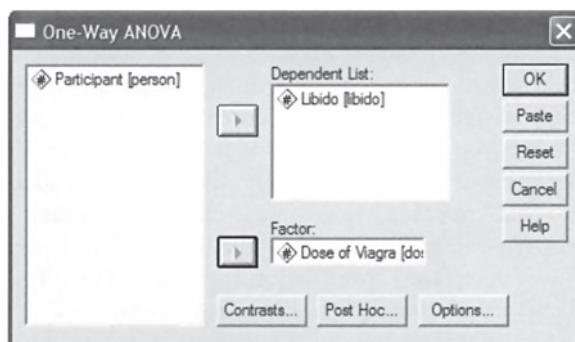


Figura 8.8 Caixa de diálogo principal para a ANOVA de um fator.

e transferi-la para o quadro denominado **Dependent List** (Lista Dependente) clicando em ►. Selecione então a variável de grupo **dose** e transfira-a para o quadro denominado **Factor** (Fator) clicando em ►.

8.3.1 Comparações planejadas utilizando o SPSS ②

Se você clicar em **Contrasts...** (Contrastes), vai acessar a caixa de diálogo que permite fazer as comparações planejadas descritas na Seção 8.2.10.

Essa caixa de diálogo é mostrada na Figura 8.9 e ela apresenta duas seções. A primeira seção serve para realizar uma análise de tendência. Se você quer testar alguma tendência nos dados, marque o espaço denominado **Polynomial** (Polinomial). Depois de

marcá-lo, selecione o grau do polinômio desejado. Os dados do Viagra apresentam somente três grupos e assim o maior grau de tendência que pode ser identificado é o quadrático (veja a Seção 8.2.10.3). Para realizar uma análise de tendência, é importante que você tenha codificado a variável de grupo de uma forma significativa. Assim, esperamos que a libido seja menor no grupo-placebo, aumente no grupo de baixa dosagem e aumente novamente no grupo de alta dosagem. Para detectar uma tendência que faça sentido, é necessário ter codificado esses grupos em uma ordem crescente. Fizemos isso codificando o grupo-placebo com o valor 1, o grupo de baixa dosagem com o valor 2 e o grupo de alta dosagem com o valor 3. Se tivéssemos codificado os grupos de uma forma diferente, isso influenciaria tanto a tendência a ser detectada quanto se ela fosse estatisticamente significativa.

Para os dados do Viagra existem

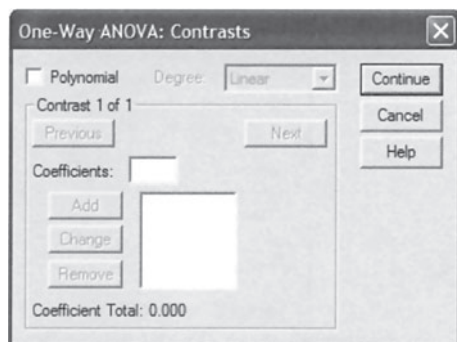
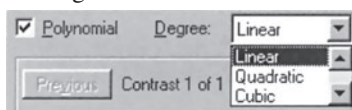


Figura 8.9 Caixa de diálogo para realizar comparações planejadas.

somente três grupos, assim, devemos selecionar a opção polinomial ☒ **Polynomial**, e depois selecionar o grau quadrático (dois) clicando em ▼ e então selecionando a opção **Quadratic** (Quadrática). Se a tendência quadrática for selecionada, o SPSS irá testar tanto essa opção quanto a linear.

A parte inferior da caixa de diálogo semelhante à da Figura 8.9 serve para especificar qualquer comparação planejada. Para fazer isso, precisamos informar ao SPSS quais pon-



derações ele deve atribuir para cada grupo. O primeiro passo é decidir quais comparações queremos fazer e então quais pesos devem ser atribuídos a cada grupo para cada um dos contrastes. Já vimos esse processo na Seção 8.2.10.2, assim, sabemos que os pesos para o primeiro contraste são -2 para o grupo-placebo, 1 para o grupo de baixa dosagem e novamente 1 para o grupo de alta dosagem. Iremos especificar esse contraste primeiro. É importante estar seguro de que o peso correto seja atribuído para cada grupo, desse modo, você deve lembrar que o primeiro peso atribuído deve ser a ponderação do primeiro grupo (isto é, o grupo codificado com o menor valor no editor de dados). Para os dados do Viagra, o grupo codificado com o menor valor é o placebo (que tem o código 1) e devemos entrar com a ponderação para esse grupo primeiro. Clique no espaço denominado **Coefficients** (Coeficientes) com o mouse e digite “ -2 ” e clique em **Add** (Adicionar). O passo seguinte é entrar com a ponderação para o segundo grupo, que para os dados do Viagra é o grupo de baixa dosagem (porque esse grupo está codificado no editor com o segundo valor mais alto). Clique novamente no espaço denominado **Coefficients** (Coeficientes) com o mouse e digite “ 1 ” clicando em seguida em **Add** (Adicionar). Finalmente, devemos entrar com a ponderação do último grupo que, para os dados do Viagra, é o grupo de alta dosagem (no editor esse grupo foi codificado com o valor mais alto). Clique em **Coefficients** (Coeficientes) com o mouse e digite “ 1 ” e, depois, clique em **Add** (Adicionar). O resultado deve ser semelhante ao da Figura 8.10.

Uma vez que os pesos foram atribuídos você poderá alterar ou remover qualquer um deles utilizando o mouse para selecionar a ponderação que quer mudar. O peso irá aparecer no espaço denominado **Coefficients** (Coeficientes), onde você poderá digitar o novo valor e então clicar em **Change** (Mudar). Alternativamente, você pode clicar em qualquer um dos pesos e removê-lo completamente clicando em **Remove** (Remover). Utilizando as ponderações, o SPSS calcula o coeficiente total

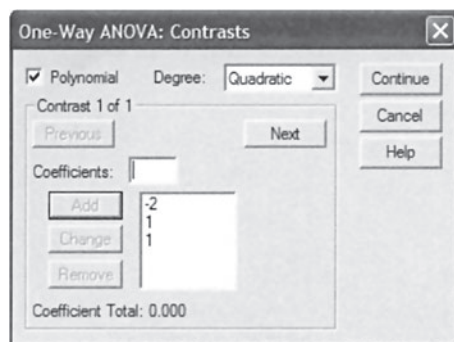


Figura 8.10 Caixa de diálogo dos contrastes completada para o primeiro contraste para os dados do Viagra.

que, como vimos na Seção 8.2.10.2, deve ser igual a zero. Se esse coeficiente for qualquer outro valor que não zero, você deve retornar e verificar se todas as regras para a atribuição de pesos foram seguidas.

Uma vez que você tenha especificado o primeiro contraste, clique em **Next** (próximo). Os pesos que você recém digitou desaparecerão e a caixa de diálogo estará pronta para a entrada do segundo contraste que, nesse caso, é o último contraste. Sabemos da Seção 8.2.10.2 que os pesos para o segundo contraste foram: 0 (para o grupo-placebo), $+1$ para o grupo de baixa dosagem e -1 para o grupo de alta dosagem. Podemos especificar esse contraste da mesma forma que antes. Lembrando que a primeira ponderação a ser entrada é a do grupo-placebo, devemos entrar o valor 0 como a primeira ponderação. Clique no espaço denominado **Coefficients** (Coeficientes) com o mouse e digite “ 0 ” e clique em **Add** (Adicionar). A seguir precisamos entrar com a ponderação para o grupo de baixa dosagem clicando novamente no espaço **Coefficients** (Coeficientes), digitando “ 1 ” e clicando em **Add** (Adicionar). Finalmente, será adicionada a ponderação para o grupo de alta dosagem da mesma forma só que agora com o cuidado de digitar o valor “ -1 ”. O resultado deve ser uma caixa de diálogo semelhante à da Figura 8.11. Note que a soma dos pesos é zero, como ocorreu com o contraste 1. Lembre-se de en-

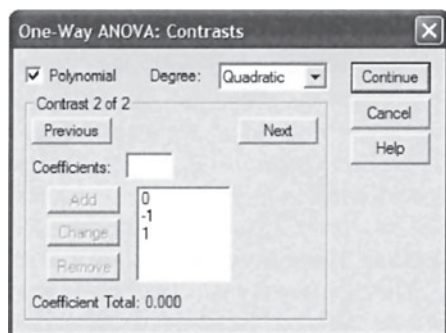


Figura 8.11 Caixa de diálogo dos contrastes preenchida para o segundo contraste para os dados do Viagra.

trar com o peso 0 para qualquer grupo que não estiver no contraste. Quando todos os contrastes planejados tiverem sido especificados, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

8.3.2 Testes *Post Hoc* no SPSS ②

Depois de informar ao SPSS quais as comparações planejadas a serem feitas, podemos escolher realizar testes *post hoc*. Teoricamente, se fizemos as comparações planejadas não precisamos realizar testes *post hoc* (porque as hipóteses de interesse já foram testadas). Da mesma forma, se tivéssemos escolhido realizar testes *post hoc* não seria necessário fazer os

contrastos planejados (porque não temos hipóteses para testar). Contudo, para poupar espaço, iremos realizar alguns testes *post hoc* para os dados do Viagra. Clique em **Post Hoc...** na caixa de diálogo principal para acessar a caixa de diálogo para os testes *post hoc* (Figura 8.12).

Na Seção 8.2.11.3, recomendei vários procedimentos *post hoc* para várias situações. Para os dados do Viagra, os tamanhos das amostras são iguais e não existe necessidade de realizar o teste de Gabriel. Devemos utilizar o teste de Tukey e o REGWQ e checar os resultados com o procedimento de Games-Howell. Temos uma hipótese específica que tanto o grupo de alta dosagem quanto o de baixa dosagem devem diferir do grupo-placebo e assim podemos utilizar o teste de Dunnett para examinar essas hipóteses. Uma vez que você tenha selecionado o teste de Dunnett, você pode mudar a categoria de controle (o padrão é utilizar a última categoria) para especificar que a primeira categoria será utilizada como categoria de controle (porque o grupo-placebo foi codificado com o valor mais baixo). Você pode, também, escolher se irá realizar um teste do tipo bilateral (☒ 2-sided) ou unilateral. Se você escolher o teste unilateral, deve prever se você acredita que a média do grupo-controle será menor do que a média de um grupo-expe-

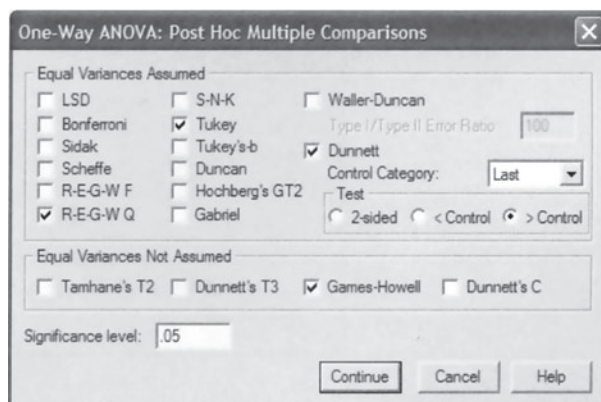


Figura 8.12 Caixa de diálogo para especificar testes *post hoc*.

rimental em particular (**<Control**) ou maior do que a média de um grupo-experimental em particular (**>Control**). Esses são todos os testes *post hoc* que você necessita especificar e quando a caixa de diálogo estiver completa ela deverá ser semelhante à da Figura 8.12. Clique em **Continue** (continuar) para voltar à caixa de diálogo principal.

8.3.3 Opções ②

As opções para a ANOVA de um fator são bem simples. Primeiro você pode solicitar estatísticas descritivas, o que irá produzir uma tabela de médias, desvios padrão, erros padrão, amplitudes e intervalos de confiança para as médias de cada grupo. Essa é uma opção útil, pois poderá auxiliar na interpretação dos resultados finais. Uma opção crucial a ser selecionada é o teste de homogeneidade das variâncias. Como ocorre com os testes t, há uma suposição de que as variâncias dos grupos são iguais e selecionando essa opção essa hipótese é testada. O SPSS utiliza o teste de Levene, que testa a hipótese de que as variâncias de cada grupo são iguais (veja a Seção 3.6). Se a hipótese de homogeneidade das variâncias for violada, o SPSS oferece duas versões alternativas da razão F: a de Brow-Forsythe (1974) e a de Welch (1951). Essas duas estatísticas são discutidas no Quadro 8.3 (e, se você está realmente entediado, olhe o arquivo **WelchF.pdf** no site www.artmed.com.br), mas é suficiente dizer que elas devem ser selecionadas apenas como garantia para o caso de a hipótese ser violada.

Existe, também, a opção de ter um diagrama das médias (selecionar a opção **Means plot**) (Figura 8.13). Se essa opção for selecionada, um diagrama de linha do grupo de médias será apresentado na saída. Finalmente, as opções nos deixam especificar se queremos excluir casos na base **listwise** ou em uma base **per analysis**. Essa opção é útil se você está realizando várias ANOVAs com diferentes variáveis dependentes. A primeira opção (**Exclude cases analysis by analysis**) (Excluir casos análise por análise) exclui qualquer caso que tiver um valor faltando (*missing*) tanto para a variável independente

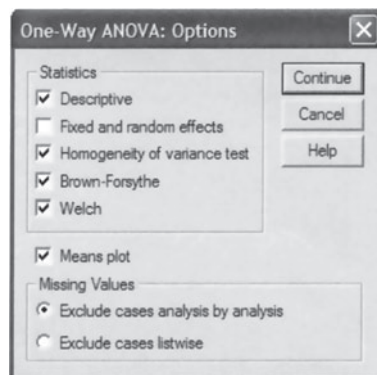


Figura 8.13 Opções para a ANOVA de um fator.

quanto para a variável dependente utilizadas nessa análise em particular. A opção **Exclude cases listwise** (Excluir casos por lista) excluirá de todas as análises qualquer caso que tenha um valor faltando para a variável independente ou qualquer variável dependente especificada. Assim, se você se acostumar a não realizar centenas de ANOVAs em diferentes variáveis (veja o Capítulo 14 sobre MANOVA) os valores padrão serão adequados. Quando você tiver selecionado as opções apropriadas, clique em **Continue** (continuar) para retornar à caixa de diálogo principal e clique em **OK** para rodar a análise.

8.4 SAÍDAS DA ANOVA DE UM FATOR (ONE WAY) ②

8.4.1 Saídas para a análise principal ②

Se você carregar os dados do Viagra (ou digitá-los) e selecionar todas as opções que eu sugeri, deverá verificar que a saída será semelhante a que segue. Se a sua saída for diferente não se assuste porque um de nós fez algo errado – espero que não eu. A Figura 8.14 mostra um diagrama de barras de erro dos dados do Viagra com uma linha superposta para apresentar a tendência geral das médias ao longo dos grupos. Esse gráfico não é automaticamente produzido pelo SPSS; contudo, um diagrama de linha será produzido se a opção **Means plot** (Representar médias) for selecionada (esse grá-

Quadro 8.3

O que eu faço na ANOVA quando a hipótese da homogeneidade das variâncias é violada? ③



Na Seção 8.2.9.1, mencionei que quando os tamanhos dos grupos são desiguais, violações na hipótese de homogeneidade das variâncias podem ter sérias consequências. O SPSS incorpora opções para duas razões F , que foram projetadas para serem mais robustas caso a homogeneidade das variâncias não seja válida. A primeira é a razão F de Brown e Forsythe (1974), que é bem fácil de explicar. Relatei anteriormente que quando os tamanhos dos grupos são diferentes e os grupos maiores apresentam as maiores variâncias, isso introduz um viés na razão F , que tende a ser mais conservadora. Se você voltar à equação (8.7), isso ficará claro, pois para calcular SS_R as variâncias são multiplicadas pelos seus tamanhos amostrais (menos 1), assim, nessa situação obteremos um grande tamanho amostral multiplicado por uma grande variância que irá inflar o valor do SS_R . Que efeito isso vai causar na razão F ? Bem, a razão F é proporcional a SS_M/SS_R , desse modo, se SS_R é grande, então a razão F será pequena (o motivo para ser mais conservadora: seu valor está sendo reduzido demais!). Brown e Forsythe contornaram esse problema ponderando as variâncias dos grupos, não pelos seus tamanhos amostrais, mas pelo inverso dos seus tamanhos amostrais (na verdade, eles utilizam n/N , o tamanho da amostra como uma proporção do tamanho total das amostras). Isso significa que o impacto de grandes tamanhos amostrais com grandes variâncias será reduzido:

$$F_{BF} = \frac{SS_M}{SS_{R_{BF}}} = \frac{SS_M}{\sum s_k^2 (1 - \frac{n_k}{N})}$$

Assim, para os dados do Viagra, a SS_M é a mesma de antes (20,135), dessa forma, a equação fica:

$$\begin{aligned} F_{BF} &= \frac{20,135}{s_{\text{grupo1}}^2 (1 - \frac{n_{\text{grupo1}}}{N}) + s_{\text{grupo2}}^2 (1 - \frac{n_{\text{grupo2}}}{N}) + s_{\text{grupo3}}^2 (1 - \frac{n_{\text{grupo3}}}{N})} \\ &= \frac{20,135}{1,70(1 - \frac{5}{15}) + 1,70(1 - \frac{5}{15}) + 2,50(1 - \frac{5}{15})} \\ &= \frac{20,135}{3,933} \\ &= 5,119 \end{aligned}$$

Essa estatística é avaliada utilizando os graus de liberdade do modelo e dos termos erro. Para o modelo, o gl_M é o mesmo de antes (isto é, $k - 1 = 2$), mas é feito um ajuste nos graus de liberdade dos resíduos, gl_R . Os cálculos desse valor bem como do F de Welch (1951) estão além dos objetivos deste livro, mas se você estiver interessado consulte o apêndice A.5 (ou o arquivo **WelchF.pdf** no site www.artmed.com.br), que detalha o F de Welch. Tenham em mente que essas duas técnicas fazem ajustes no F e nos graus de liberdade dos resíduos com o propósito de combater os problemas originados da violação da hipótese da homogeneidade das variâncias.

A pergunta óbvia é qual dos dois procedimentos é melhor. Tomarken e Serlin (1986) revisaram essas e outras técnicas e concluíram que ambas controlam bem a taxa de erro do Tipo I (isto é, quando não existem efeitos na população você, de fato, obtém um F não-significativo). Contudo, em termos de poder (isto é, qual teste é melhor para detectar um efeito quando ele existe), o teste de Welch parece se comportar melhor exceto quando existe uma média extrema com uma grande variância.



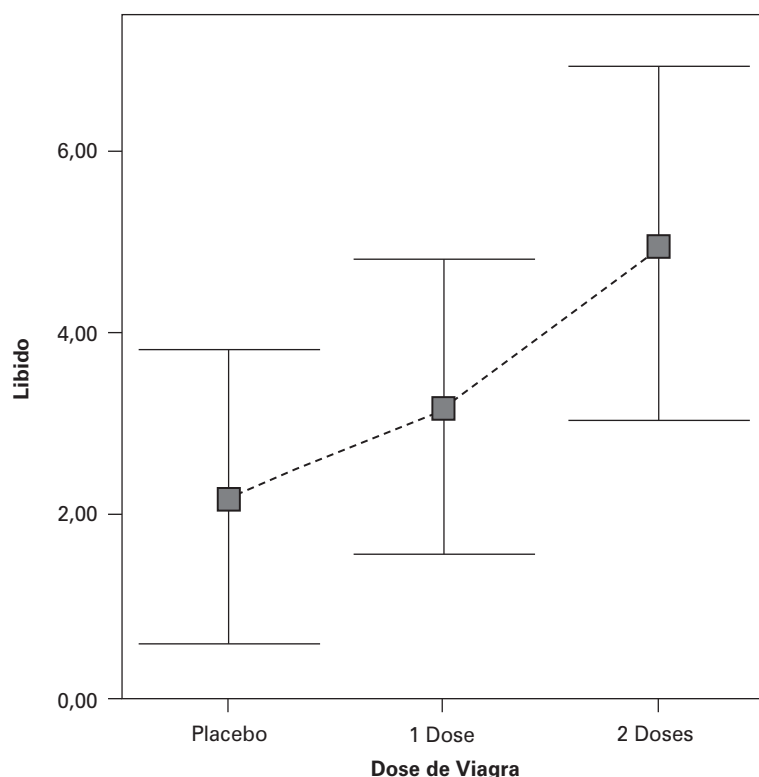


Figura 8.14 Diagrama de barras de erros para os dados do Viagra.

fico será parecido com o da Figura 8.14, mas sem as barras de erros). Escolhi apresentar um diagrama de barras de erro porque ele é mais interessante que apenas o diagrama de linha. Fica claro a partir desse gráfico que todas as barras de erro se sobrepõem, indicando que não existem diferenças entre os grupos (embora essa medida seja aproximada). A linha que une as médias indica uma tendência linear no sentido de que à medida que a dose de Viagra aumenta, cresce também proporcionalmente o nível médio da libido.

A Saída do SPSS 8.3 mostra a tabela das estatísticas descritivas para o procedimento de um fator dos dados do Viagra. Observe que as médias e os desvios padrão correspondem àqueles mostrados na Tabela 8.1. Além disso, o erro padrão é apresentado. Você deve lembrar que o erro padrão é o desvio padrão da distribuição amostral desses dados (assim, para o

grupo-placebo, se você retirar muitas amostras da população da qual esses dados vieram, as médias dessas amostras terão um desvio padrão de 0,5831). Intervalos de confiança para as médias também são apresentados. Você já deve estar familiarizado com o que um intervalo de confiança nos informa: se tomarmos 100 amostras da população da qual o grupo-placebo foi retirado e construirmos intervalos de confiança para as médias dessas amostras, 95 desses intervalos irão conter a média da população. Em outras palavras, é provável que a média da população esteja entre 0,5811 e 3,8189. Embora esses diagnósticos não sejam importantes agora, iremos nos referir a eles no decorrer da análise.

A próxima parte da saída é uma tabela que resume o teste de Levene (Saída do SPSS 8.4). Esse teste é projetado para verificar a hipótese nula de que as variâncias dos grupos são iguais. Ele é uma ANOVA conduzida nas diferenças

Saída do SPSS 8.3

Descriptives (Descritivas)

			N	Mean (Média)	Std. Deviation (Desvio Padrão)	Std. Error (Erro Padrão)	95% Confidence Interval for Mean (Intervalo de Confiança de 95% para a Média)		Minimum (Mínimo)	Maximum (Máximo)
							Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)		
Libido	Dose of Viagra (Dose de Viagra)	Placebo	5	2.2000	1.3038	0.5831	0.5811	3.8189	1.00	4.00
		Low Dose (Baixa Do- sagem)	5	3.2000	1.3038	0.5831	1.5811	4.8189	2.00	5.00
		High Dose (Alta Dosagem)	5	5.0000	1.5811	0.7071	3.0368	6.9632	3.00	7.00
		Total	15	3.4667	1.7674	0.4563	2.4879	4.4454	1.00	7.00

absolutas entre os dados observados e a média de onde os dados vieram.⁴ Nesse caso, o teste de Levene está, portanto, testando se as variâncias dos três grupos são significativamente diferentes. Se o teste de Levene é significativo (isto é, o valor de *Sig* é menor do que 0,05), podemos dizer que as variâncias são significativamente diferentes. Isso quer dizer que violamos uma das suposições da ANOVA e que devemos tomar algumas providências para corrigir isso. Como vimos no Capítulo 3, uma forma comum de retificar diferenças entre as variâncias dos grupos é transformar todos os dados: se as variâncias são diferentes, elas podem, às vezes, ser estabilizadas tomando-se a raiz quadrada de cada valor da variável dependente e depois analisando novamente esses valores transformados (veja o Capítulo 3 e Howell, 2002, pp. 342-349). Contudo, dado a aparente utilidade do F de Welch e do F de Brown-Forsythe e o fato de que as transformações muitas vezes não ajudam, você pode relatar o F de Welch (e eu sugiro que isso seja

feito em vez do F de Brown-Forsythe a menos que você tenha também uma média extrema que esteja causando problemas com a variância). Felizmente, para esses dados as variâncias são parecidas (daí o alto valor da probabilidade); de fato, se você observar a Saída 8.3 do SPSS verá que as variâncias do grupo-placebo e do grupo de baixa dosagem são idênticas.

Saída do SPSS 8.4
Test of Homogeneity of Variances (Teste
de homogeneidade das variâncias)

	Levene Statistic (Estatística de Levene)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Libido	0.092	2	12	0.913

A Saída do SPSS 8.5 mostra a tabela resumo da ANOVA principal. A tabela é dividida em efeitos entre os grupos (efeitos devidos ao modelo – o efeito experimental) e efeitos dentro dos grupos (essa é a variação assistemática dos dados). O efeito entre grupos é depois dividido nas componentes linear e quadrática e esses componentes são a análise de tendência descrita na Seção 8.2.10.5. O efeito entre grupos denominado *Combinado* é o efeito experimental geral. Nessa linha são informadas as somas dos quadrados para o modelo

⁴ O leitor interessado pode tentar fazer isso. Basta criar uma variável denominada **diff** (forma abreviada para diferença), que é cada escore subtraído da média do grupo a qual esse escore pertence. Depois, remova todos os sinais de menos (isto é, pegue os valores absolutos de **diff**) e realize uma ANOVA de um fator com **dose** como variável independente e **diff** como variável dependente. Você irá verificar que a razão F para essa análise é 0,092, que apresenta uma significância *p* = 0,913!

Saída do SPSS 8.5

ANOVA

				Sum of Squares (Soma dos Quadrados)	df (gl)	Mean Square (Quadrados Médios)	F	Sig. (Sig.)
Libido	Between Groups (Entre Grupos)	(Combined) (Combinado)		20.133	2	10.067	5.119	0.025
		Linear Trend (Tendência Linear)	Contrast (Contraste)	19.600	1	19.600	9.966	0.008
			Deviation (Desvios)	0.533	1	0.533	0.271	0.612
	Within Groups (Dentre Grupos)	Quadratic Trend (Tendência Quadrática)	Contrast (Contraste)	0.533	1	0.533	0.271	0.612
				23.600	12	1.967		
		Total		43.733	14			

($SS_M = 20,13$) e esse valor corresponde ao valor calculado na seção 8.2.5. Os graus de liberdade são iguais a 2 e os quadrados médios para o modelo correspondem aos valores calculados na seção 8.2.7 (10,067). As somas dos quadrados e as médias dos quadrados representam o efeito experimental. Esse efeito global é então dividido porque solicitamos ao SPSS que fizesse uma análise de tendência nos dados (iremos abordar novamente as tendências no momento devido). Como não especificamos isso na Seção 8.3.1, essas duas linhas na tabela resumo não serão produzidas. A linha rotulada de *Within groups* (dentro dos grupos) fornece detalhes da variação não-sistemática dentro dos dados (a variação devido a diferenças naturais individuais na libido e as diferentes reações ao Viagra). A tabela nos informa quanta variação assistemática existe (a soma dos quadrados residual, SS_R) e esse valor (23,60) corresponde ao valor calculado na Seção 8.2.6. A tabela fornece a quantidade média de variação assistemática, a média dos quadrados (MS_R), que corresponde ao valor (1,967) calculado na Seção 8.2.7. O teste sobre se as médias dos grupos são as mesmas é representado pela razão F para os efeitos entre grupos combinados. O valor dessa razão é 5,12, que é a mesma calculada na Seção 8.2.8. Finalmente, o SPSS informa a probabilidade de esse valor ter ocorrido por acaso. A coluna final rotulada *Sig.* mostra a probabilidade

da ocorrência de uma razão F do tamanho da obtida por acaso. Nesse caso, existe uma probabilidade de 0,025 de que uma razão F desse valor tenha ocorrido por acaso (isto é, uma chance de apenas 2,5%). Vimos em capítulos anteriores que os cientistas sociais utilizam um ponto de corte de 0,05 como critério de significância. Então, em virtude do valor da significância observada ser menor do que esse valor podemos dizer que existe um efeito significativo do Viagra.⁵ Contudo, nesse estágio não sabemos exatamente qual foi o efeito do Viagra (não sabemos quais grupos estão diferindo). É interessante notar aqui que obtemos um efeito experimental significativo embora nosso diagrama de barras de erros tenha indicado que nenhuma diferença seria encontrada. Essa contradição ilustra como o diagrama de barra de erros pode ser utilizado apenas como uma ilustração grosseira dos dados.

Sabendo que o efeito global do Viagra foi significativo, podemos agora observar a análise

⁵ Uma pergunta que meus alunos geralmente fazem é: “a significância da ANOVA é uni ou bilateral, e se ela é bilateral posso dividir o valor por 2 para obter o valor unilateral?”. A resposta é que para um teste unilateral você deve fazer uma hipótese direcional (isto é, a média para gatos é maior do que a dos cães). A ANOVA é um teste não-específico, assim, ela apenas informa genericamente se existe uma diferença ou não e, porque existem várias médias, você não pode formular uma hipótese direcional. Desse modo, não é válido dividir por dois o valor da significância.

de tendência. Essa análise divide o efeito experimental para ver se ele pode ser explicado por um relacionamento linear ou quadrático. Primeiro, vamos tentar uma componente linear. Essa comparação testa se as médias crescem com os grupos de uma forma linear. Novamente a soma dos quadrados e a média dos quadrados são fornecidas, mas o mais importante a ser notado é o valor da razão F e o valor correspondente da significância. Para a tendência linear a razão F é 9,97 e esse valor é significativo a um nível de 0,008. Portanto, podemos dizer que à medida que a dose de Viagra aumenta de nenhum Viagra para uma alta dose do medicamento, a libido cresce proporcionalmente. Na tendência quadrática, essa comparação é testar se o padrão das médias é curvilíneo (isto é, representado por uma curva que tem uma curvatura). O gráfico de barras de erros dos dados recomenda que as médias não podem ser representadas por uma curva e os resultados para a tendência quadrática confirmam isso. A razão F para a tendência quadrática não é significativa (de fato, o valor de F é menor do que 1, o que imediatamente indica que esse contraste não será significativo).

Finalmente, a Saída do SPSS 8.6 mostra as razões F de Welch e de Brow-Forsythe. Não precisamos delas, pois o teste de Levene não foi significativo, indicando que as variâncias podem ser consideradas iguais. Contudo, quando a homogeneidade da variância for violada você deve levar em conta esses valores *em vez* daqueles na tabela principal. Se você estiver interessado em como esses valores são calculados, verifique o Quadro 8.3 e o arquivo **WelchF.pdf** no *site* www.artmed.com.br. No entanto, sendo honesto, isso é bastante confuso

e eu recomendo que você fique com os valores calculados na Saída 8.6 do SPSS e acredite que eles fazem o que devem fazer (você também deve notar que os graus de liberdade do erro foram ajustados e precisa lembrar-se disso quando for relatar esses valores!).

8.4.2 Saídas para as comparações planejadas ②

Na Seção 8.31, solicitamos que o SPSS realizasse duas comparações planejadas: uma para testar se o grupo-controle era diferente dos dois grupos que receberam o Viagra e outro para ver se as duas doses de Viagra faziam diferença na libido. A Saída do SPSS 8.7 mostra os resultados das comparações que foram solicitadas para os dados do Viagra. A primeira tabela apresenta os contrastes dos coeficientes; esses valores são os que foram digitados na Seção 8.3.1 e vale a pena olhar essa tabela para verificar novamente que os contrastes estão comparando aquilo que eles deveriam comparar! Como uma regra prática rápida, lembre que quando realizamos comparações planejadas arranjamos os pesos de tal forma que comparamos qualquer grupo com um peso positivo contra qualquer grupo com pesos negativos. Portanto, a tabela dos pesos mostra que o contraste 1 compara o grupo-placebo contra os dois grupos-experimentais e o contraste 2 compara o grupo de baixa dosagem contra o de alta dosagem. É útil dar uma olhada nessa tabela para se certificar de que os pesos digitados no SPSS correspondem aos pesos que nós, de fato, pretendíamos informar ao SPSS!

A segunda tabela fornece a estatística para cada contraste. Note que as estatísticas são produzidas para situações nas quais as variâncias dos grupos são iguais e para quando elas são diferentes. Se o teste de Levene foi significativo, você deve ler a parte da tabela rotulada como: **Does not assume equal variances** (Não assume a igualdade de variâncias). Contudo, para esses dados o teste de Levene não foi significativo, portanto, podemos utilizar a parte da tabela denominada **Assume equal variances** (Assume a igualdade de variâncias). A tabela nos informa o valor do contraste, que é a soma ponderada da média dos grupos. Esse valor é obtido toman-

Saída do SPSS 8.6
Robust Tests of Equality of Means (Testes robustos de igualdade de médias)

Libido	Statistic ^a (Estatística)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Welch	4.320	2	7.943	0.054
Brown-Forsythe	5.119	2	11.574	0.026

a. Asymptotically F distributed (F assintoticamente distribuído)

Saída do SPSS 8.7

Contrast Coefficients (Coeficientes dos contrastes)

Contrast (Contraste)o	Placebo (Placebo)	Low Dose (Baixa Dosgaem)	High Dose (Alta Dosagem)
1	-2	1	1
2	0	-1	1

Contrast Tests (Testes dos contrastes)

		Contrast (Contraste)	Value of Contrast (Valor do contraste)	Std Error (Erro padrão)	t	df (gl)	Sig. (2-tailed) (Sig. bilateral)
Libido	Assume equal variances (Assume a igualdade de variâncias)	1	3.8000	1.5362	2.474	12	0.029
		2	1.8000	0.8869	2.029	12	0.065
	Does not assume equal va- riances (Não assume a igual- dade de variâncias)	1	3.8000	1.4832	2.562	8.740	0.031
		2	1.800	0.9165	1.964	7.720	0.086

do a média de cada grupo, multiplicando-a pelo peso do contraste de interesse e então somando todos os valores.⁶ A tabela também fornece o erro padrão de cada contraste e a estatística *t*. A estatística *t* é obtida dividindo o valor do contraste pelo erro padrão ($t = 3,8/1,5362 = 2,47$) e isso é comparado com o valor crítico da distribuição *t*. A significância do contraste é dada na coluna final e esse valor é bilateral. Utilizando o primeiro contraste como exemplo, se o tivéssemos utilizado para testar a hipótese geral de que os grupos-experimentais iriam diferir do placebo, deveríamos utilizar esse valor bilateral. Contudo, na realidade testamos a hipótese de que os grupos-experimentais irão aumentar a libido acima dos níveis observados no grupo-placebo: essa hipótese é unilateral. Obtidas as médias para esses grupos e considerando as hipóteses, podemos dividir os valores das significâncias por 2 a fim de obter as probabilidades unilaterais. Assim, para o contraste 1, podemos dizer que tomar Viagra aumenta significativamente a libido comparado ao grupo-controle ($p = 0,0145$). Para o contraste 2 temos novamente uma hipótese unilateral (de que uma alta dosa-

gem de Viagra irá aumentar a libido mais do que uma baixa dosagem) e as médias suportam essa hipótese. A significância para o contraste 2 nos informa que uma alta dose de Viagra aumenta significativamente a libido mais do que uma baixa dosagem ($p_{\text{unilateral}} = 0,065/2 = 0,0325$). Note que se não tivéssemos uma hipótese específica sobre qual grupo teria a média mais alta, deveríamos concluir que a dose de Viagra não teria um efeito significativo na libido. Por isso, é importante que geremos as hipóteses antes de coletar qualquer dado porque esse método de descoberta científica é mais poderoso.

Em resumo, existe um efeito global do Viagra na libido. As comparações planejadas revelaram que o Viagra aumentou significativamente a libido quando comparado ao grupo-controle, $t(12) = 2,47$, $p < 0,05$, e que uma alta dosagem aumentou significativamente a libido quando comparada com uma baixa dosagem, $t(12) = 2,03$, $p < 0,05$.

8.4.3 Saídas para os testes *post hoc* ②

Se não tivéssemos hipóteses específicas sobre o efeito que o Viagra exerce na libido, poderíamos executar testes *post hoc* para comparar todos os grupos de participantes entre eles. De fato, solicitamos ao SPSS que fizesse

⁶ Para o primeiro contraste, esse valor é:

$$\sum(\bar{X}W) = [(2,2 \times -2) + (3,2 \times 1) + (5,0 \times 1)] = 3,8$$

isso (veja a Seção 8.3.2), e os resultados dessa análise são mostrados na Saída 8.8 do SPSS. Essa tabela mostra os resultados do teste de Tukey (conhecido como Tukey HSD)⁷, do procedimento de Games-Howell e o teste de Dunnett, que foram especificados anteriormente. Se observarmos primeiro o teste de Tukey (porque não temos razão para duvidar que as variâncias populacionais sejam diferentes), fica claro a partir da tabela que cada grupo de participantes é comparado aos outros grupos restantes. Para

cada par de grupos, a diferença entre as médias dos grupos é apresentada bem como o erro padrão, o nível de significância e um intervalo de 95% de confiança para essa diferença. Primeiro, o grupo-placebo é comparado ao grupo de baixa dosagem e revela uma diferença não-significativa (*Sig.* é maior do que 5%), mas quando comparado ao grupo de alta dosagem existe uma diferença significativa (*Sig.* é menor do que 5%). Essa descoberta é interessante porque nossa comparação planejada mostrou que qualquer dose de Viagra produz um aumento significativo na libido, embora essas comparações indiquem que uma dosagem baixa não faça

⁷ A sigla HSD significa *Honestly Significant Difference* (Diferença Honestamente Significativa).

Saída do SPSS 8.8

Multiple Comparisons (Comparações múltiplas)

Dependent Variable: Libido (Variável dependente: libido)

	(I) Dose of Viagra (I Dose de Viagra)	(J) Dose of Viagra (J) Dose de Viagra)	Mean Difference (I-J) (Diferença Média I-J)	Std. Error (Erro Padrão)	Sig. (Sig.)	95% Confidence Interval (Intervalo de 95% Confiança)	
						Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
HSD de Tukey	Placebo	Low Dose (Baixa Dosagem)	-1.0000	0.88694	0.516	-3.3662	1.3662
		High Dose (Alta Do- sagem)	-2.8000*	0.88694	0.021	-5.1662	-0.4338
	Low Dose (Baixa Dosagem)	Placebo	1.0000	0.88694	0.516	-1.3662	3.3662
		High Dose (Alta Do- sagem)	-1.8000	0.88694	0.147	-4.1662	0.5662
	High Dose (Alta Dosagem)	Placebo	2.8000*	0.88694	0.021	0.4338	5.1662
		Low Dose (Baixa Dosagem)	1.8000	0.88694	0.147	-0.5662	4.1662
Games- Howell	Placebo	Low Dose (Baixa Dosagem)	-1.0000	0.82462	0.479	-3.3563	1.3563
		High Dose (Alta Do- sagem)	-2.8000*	0.91652	0.039	-5.4389	-0.1611
	Low Dose (Baixa Dosagem)	Placebo	1.0000	0.82462	0.0479	-1.3563	3.3563
		High Dose (Alta Do- sagem)	-1.8000	0.91652	0.185	-4.4389	0.8389
	High Dose (Alta Dosagem)	Placebo	2.8000*	0.91652	0.039	0.1611	5.4389
		Low Dose (Baixa Dosagem)	1.8000	0.91652	0.185	-0.8389	4.4389
Dunnett t (> control) ^a (> controle)	Low Dose (Baixa Dosagem)	Placebo	1.0000	0.88694	0.227	0.8697	
	High Dose (Alta Dosagem)	Placebo	2.8000*	0.88694	0.008	0.9303	

* A diferença média é significativa ao nível de 5%.

a O test t de Dunnett trata um grupo como controle e compara todos os outros contra esse grupo-controle.

feito. Por que existe essa contradição? (Pense um pouco nisso antes de prosseguir.)

Na Seção 8.2.10.2, expliquei que a primeira comparação planejada compara os grupos-experimentais com o grupo-placebo. Especificamente, ela compara a média dos dois grupos-experimentais ($(3,2 + 5,0)/2 = 4,1$) com a média do grupo-placebo (2,2). Assim, a diferença entre esses dois valores ($4,1 - 2,2 = 1,9$) é que foi considerada significativa. Nos testes *post hoc*, quando o grupo de baixa dosagem é comparado com o grupo-placebo, o contraste está testando se a diferença entre as médias desses dois grupos é significativa. A diferença, nesse caso, é de apenas 1, comparada com uma diferença de 1,9 para a comparação planejada. Essa explicação ilustra como é possível ter resultados aparentemente contraditórios entre as comparações planejadas e as comparações *post hoc*. Mais importante, ela ilustra o cuidado que devemos ter na interpretação dos contrastes planejados.

O grupo de baixa dosagem é então comparado com o grupo-placebo e com o grupo de alta dosagem. Note primeiro que o contraste envolvendo os grupos-placebo e de baixa dosagem é idêntico àquele recém-descrito. A única informação nova é a comparação entre as duas condições experimentais. As médias dos grupos diferem por 1,8, o que não é significativo. Esse resultado também contradiz as comparações planejadas (lembre que o contraste 2 comparou esses grupos e encontrou uma diferença significativa). Pense por que essa contradição está ocorrendo. Por um lado, testes *post hoc* por sua natureza são bilaterais (você os utiliza quando não tiver uma hipótese e você não prevê a direção da hipótese porque ela não existe!) e o contraste 2 foi significativo somente quando considerado uma hipótese unilateral. Contudo, mesmo com um nível bilateral a comparação planejada foi mais próxima na significância que o teste *post hoc* e esse fato ilustra que os procedimentos *post hoc* são mais conservadores (isto é, apresentam menos poder de detectar verdadeiras diferenças) que as comparações planejadas.

O restante da tabela descreve o teste de Games-Howell e uma rápida inspeção reve-

la o mesmo padrão de resultados: os únicos grupos que diferem significativamente foram o de alta dosagem e o placebo. Esses resultados confirmam nossas conclusões a partir do teste de Tukey porque mesmo se as variâncias populacionais não forem iguais (o que parece improvável uma vez que as variâncias amostrais são bastante semelhantes), o perfil dos resultados ainda será verdadeiro. Finalmente, o teste de Dunnett é descrito e você, espero, irá lembrar que solicitamos que o computador comparasse os dois grupos-experimentais contra o grupo-controle utilizando uma hipótese unilateral de que a média do grupo-controle seria menor do que as médias dos grupos-experimentais. Mesmo com uma hipótese unilateral, os níveis de libido com baixa dosagem não são diferentes do grupo-placebo. Contudo, no grupo de alta dosagem a libido é significativamente maior do que no grupo-placebo.

A tabela na Saída 8.9 do SPSS mostra os resultados do teste de Tukey e do REGWQ. Esses testes mostram subconjuntos de grupos que têm a mesma média. Portanto, o teste de Tukey cria dois subconjuntos de grupos com médias estatisticamente semelhantes. O primeiro desses subconjuntos contém o grupo-placebo e o de baixa dosagem (indicando que esses dois grupos apresentam médias semelhantes) e o segundo desses subconjuntos contém os grupos de alta e baixa dosagem. Esses resultados demonstram que o grupo-placebo tem uma média semelhante ao grupo de baixa dosagem, mas não ao de alta dosagem, e que o grupo de baixa dosagem apresenta média semelhante tanto ao de alta dosagem quanto ao grupo-placebo. Em outras palavras, os únicos grupos que apresentam médias significativamente diferentes são os grupos-placebo e de alta dosagem. O teste fornece um valor da significância para cada subconjunto e fica claro a partir desses valores que os grupos nos subconjuntos não têm médias significativas (como o indicado pelos valores *Sig.* que são maiores do que 0,05).

Esses cálculos utilizam a média harmônica dos tamanhos amostrais. A média harmônica é uma versão ponderada da média

Saída do SPSS 8.9

Libido

Dose of Viagra (Dose de Viagra)		N	Subset for alpha = .05 (Subconjunto para alfa = 0,05)	
			1	2
Tukey HSD ^a (HSD de Tukey)	Placebo	5	2.2000	
	Low Dose (Baixa Dosagem)	5	3.2000	3.2000
	High Dose (Alta Dosagem)	5		5.0000
	Sig.		0.516	0.147
Ryan-Einot-Gabriel- Welsch Range	Placebo	5	2.2000	
	Low Dose (Baixa Dosagem)	5	3.2000	3.2000
	High Dose (Alta Dosagem)	5		5.0000
	Sig.		0.282	0.065

Médias para subconjuntos homogêneos são apresentadas.
a Utiliza a média harmônica dos tamanhos amostrais = 5.

Dica da Samanta Ferrinho



- A ANOVA independente de um fator compara várias médias, quando essas médias vêm de grupos diferentes de pessoas; por exemplo, se você tem várias condições experimentais e utiliza diferentes participantes em cada condição.
- Quando você tiver gerado hipóteses específicas antes do experimento ser realizado utilize as *comparações planejadas*, mas se você não tiver hipóteses específicas, utilize os testes *post hoc*.
- Existem vários testes *post hoc* diferentes: quando você tiver amostras do mesmo tamanho e variâncias homogêneas utilize o REGWQ ou o HSD de Tukey. Se os tamanhos das amostras não diferem muito, utilize o procedimento de Gabriel, mas se os tamanhos amostrais são bem diferentes, use o GT2 de Hochberg. Se existir qualquer dúvida sobre a homogeneidade das variâncias, utilize o procedimento de Games-Howell.
- Teste a homogeneidade das variâncias utilizando o teste de Levene. Encontre a tabela com esse rótulo: se o valor na coluna denominada *Sig.* é menor do que 0,05, a hipótese foi violada. Nesse caso, vá para a tabela denominada *Robust Tests of Equality of Means* (Testes robustos de igualdade de médias). Se a homogeneidade das variâncias não foi violada (isto é, a significância do teste de Levene é maior do que 5%), vá para a tabela denominada ANOVA.
- Na tabela denominada ANOVA (ou Testes Robustos para a Igualdade de Médias – veja acima), olhe para a coluna *Sig.*: se o valor for menor do que 0,05, as médias dos grupos são significativamente diferentes.
- Para contrastes e testes *post hoc*, olhe novamente a coluna *Sig.* para descobrir se suas comparações são significativas (elas serão se o valor da significância for menor do que 0,05).

que leva em conta o relacionamento entre variâncias e tamanhos amostrais (veja Howell, 2002, p. 234). Embora você não precise saber os caminhos intrincados da média harmônica, é útil que o tamanho amostral harmônico seja utilizado, porque elimina o viés que po-

deria ser introduzido pelos tamanhos amostrais diferenciados. Assim, a ANOVA de um fator fornece resultados confiáveis mesmo com delineamentos desbalanceados (embora os delineamentos balanceados também apresentem vantagens!).

8.5 CALCULANDO O TAMANHO DE EFEITO ②

Uma coisa que você irá notar é que o SPSS não fornece rotineiramente um tamanho de efeito para a ANOVA independente de um fator. Contudo, vimos na equação (5.4) que:

$$R^2 = \frac{SS_M}{SS_T}$$

É claro que sabemos esses valores a partir da saída do SPSS. Assim, podemos simplesmente calcular r^2 utilizando o efeito entre grupos (SS_M) e a quantidade total de variância nos dados (SS_T) – embora por alguma razão bizarra ele seja geralmente chamado de **eta** ao quadrado, η^2 . Depois, basta tomar a raiz quadrada desse valor para obtermos o tamanho de efeito r :

$$\begin{aligned} r^2 &= \frac{SS_M}{SS_T} \\ &= \frac{20,13}{43,73} = 0,42 \\ r &= \sqrt{0,42} \\ &= 0,65 \end{aligned}$$

Utilizando o padrão para tamanhos de efeito, isso representa um grande efeito (acima do limite de 0,50 que caracteriza um grande efeito). Portanto, o efeito do Viagra na libido é um achado substancial.

Contudo, essa medida do tamanho de efeito é levemente tendenciosa porque é puramente baseada nas somas dos quadrados da amostra e nenhum ajuste é feito para o fato de que estamos tentando estimar o tamanho de efeito na população. Portanto, muitas vezes utilizamos uma medida mais complicada denominada **ômega quadrado** (ω^2). Essa estimativa ainda é baseada nas somas dos quadrados que vimos nesse capítulo, mas, como a razão F, ela utiliza a variância explicada pelo modelo e o erro da variância (nos dois casos, a variância média ou erro médio ao quadrado é usado);

$$\omega^2 = \frac{SS_M - (gl_M)MS_R}{SS_T + MS_R}$$

O gl_M na equação é o grau de liberdade para o efeito, que você pode obter da saída do SPSS (no caso do efeito principal ele é o número de condições experimentais menos 1). Assim, nesse exemplo temos:

$$\begin{aligned} \omega^2 &= \frac{20,13 - (2)1,97}{43,73 + 1,97} \\ &= \frac{16,19}{45,70} \\ &= 0,35 \\ \omega &= 0,60 \end{aligned}$$

Como você pode ver, isso nos levou a uma estimativa levemente superior àquela obtida pela utilização do r e, em geral, ω é uma medida mais precisa. Embora nas subseções na ANOVA eu utilizei o ômega como minha medida do tamanho de efeito, pense nele como se fosse r (porque ele é basicamente uma estimativa não tendenciosa de r de qualquer forma) e aplique os mesmos padrões para decidir quão substancial é um efeito.

Na maioria das vezes não é interessante termos um tamanho de efeito para toda a ANOVA porque ela está testando uma hipótese geral. Em vez disso, o que queremos são tamanhos de efeito para os contrastes (porque eles comparam apenas duas coisas e, assim, o tamanho de efeito é consideravelmente mais fácil de interpretar). Comparações planejadas são testadas com estatísticas t e, dessa forma, podemos utilizar a mesma equação da Seção 7.5.5:

$$r_{\text{contraste}} = \sqrt{\frac{t^2}{t^2 + gl}}$$

Sabemos o valor de t e da gl da Saída 8.7 do SPSS e, assim, podemos calcular r da seguinte forma:

$$\begin{aligned} r_{\text{contraste1}} &= \sqrt{\frac{2,474^2}{2,474^2 + 12}} \\ &= \sqrt{\frac{6,12}{18,12}} \\ &= 0,58 \end{aligned}$$

Se você pensar nos nossos padrões para o tamanho de efeito, isso representa um grande efeito (ele está acima de 0,50, o limite para um grande efeito). Portanto, além de ser estatisticamente significativo, esse efeito é grande e, assim, ele representa um achado substancial. Para o contraste 2, temos:

$$\begin{aligned} r_{\text{contraste2}} &= \sqrt{\frac{2,029^2}{2,029^2 + 12}} \\ &= \sqrt{\frac{4,12}{16,12}} \\ &= 0,51 \end{aligned}$$

Esse resultado é também um achado substancial e representa um grande tamanho de efeito.

8.6 RELATANDO RESULTADOS DA ANOVA INDEPENDENTE DE UM FATOR ②

Quando relatamos uma ANOVA precisamos fornecer detalhes da razão F e dos graus de liberdade a partir dos quais ela foi calculada. Para o efeito experimental nesses dados a razão F foi derivada pela divisão da média dos quadrados para o efeito pela média dos quadrados dos resíduos. Portanto, os graus de liberdade utilizados para determinar a razão F são os graus de liberdade para o efeito do modelo ($gl_M = 2$) e os graus de liberdade para os resíduos do modelo ($gl_R = 12$). Desse modo, a maneira correta de relatar o principal achado será:

- ✓ Existe um efeito significativo do Viagra nos níveis da libido, $F(2, 12) = 5,12$, $p < 0,05$, $\omega = 0,60$.

Note que o valor da razão F é precedido pelos valores dos graus de liberdade para o efeito. Ainda, raramente apresentamos o valor exato da significância da razão F ; em vez disso, apresentamos que o valor da significância p , foi menor do que o valor crítico para o teste de 0,05 e incluímos uma medida do tamanho de efeito. O contraste linear pode ser apresentado praticamente da mesma forma:

- ✓ Existe uma tendência linear significativa, $F(1, 12) = 9,97$, $p < 0,01$, $\omega = 0,62$, indicando que a medida que a dose de Viagra cresce, a libido aumenta proporcionalmente.

Note que os graus de liberdade mudaram para refletir como a razão F foi calculada. Eu incluí, também, uma medida do tamanho de efeito (tente calcular isso como foi feito para a razão F principal e veja se você obtém o mesmo valor). Ainda, relatamos que o valor F foi significativo a um valor menor do que o valor crítico de 0,01. Podemos, também, relatar os contrastes planejados:

- ✓ Os contrastes planejados revelaram que qualquer dose de Viagra aumentou significativamente a libido quando comparado com o placebo, $t(12) = 2,47$, $p < 0,05$ (unilateral), $r = 0,58$ e que uma alta dosagem aumentou significativamente a libido quando comparada com uma baixa dosagem, $t(12) = 2,03$, $p < 0,05$ (unilateral), $r = 0,51$.

Observe que nos dois casos eu declarei que foi utilizada uma probabilidade unilateral.

8.7 VIOLAÇÕES DAS HIPÓTESES DA ANOVA INDEPENDENTE DE UM FATOR ②

Mencionei várias vezes neste capítulo que a ANOVA é resistente a violações das suas hipóteses. Podemos também ver que há medidas que podem ser tomadas quando existe heterogeneidade das variâncias (Quadro 8.3). Contudo, existe uma alternativa, que é utilizar um grupo de testes (frequentemente denominados distribuição livre ou não-paramétricos, mas nenhum dos dois nomes é apropriado!). Bem, a ANOVA independente de um fator também apresenta uma contraparte não-paramétrica chamada de testes de Kruskal-Wallis. Se você não tiver dados normalmente distribuídos, ou teve outra hipótese violada, esse teste poderá ser uma alternativa útil para resolver o problema. Esse teste é descrito no Capítulo 13.

8.8 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Neste capítulo abordamos a análise de variância (ANOVA), que é o tópico dos próximos capítulos também. A ANOVA independente de um fator é utilizada em situações onde queremos comparar várias médias e os dados foram coletados utilizando diferentes participantes em cada condição. Comecei explicando que se fizermos muitos testes t nos mesmos dados, a nossa taxa de erro do Tipo I irá inflacionar. Portanto, utilizamos a ANOVA em vez dos testes t . Relatei como a ANOVA pode ser conceitualizada como um MLG (Modelo Linear Geral), assim ela é, de fato, o mesmo que uma regressão múltipla. Como em uma regressão múltipla, existem três médias importantes que utilizamos na ANOVA: a soma total dos quadrados, SS_T (uma medida da variabilidade nos dados), a soma dos quadrados do modelo, SS_M (uma medida de quanta variabilidade pode ser explicada pela nossa manipulação experimental) e SS_R (uma medida de quanta variabilidade não pode ser explicada pela manipulação experimental). Descobrimos que a razão F é simplesmente a razão entre a variância que podemos explicar pela variância que não podemos. Também descobrimos que uma razão F significativa nos informa apenas que nossos grupos diferem, mas não como eles diferem. Para descobrir onde estão as diferenças temos duas opções: determinar contrastes específicos para testar hipóteses (contrastos planejados) ou testar cada grupo contra cada um dos demais grupos (*testes post hoc*). A primeira é utilizada quando temos hipóteses formuladas antes da condição do experimento e a segunda é para explorar os dados quando nenhuma hipótese foi feita. Finalmente, descobrimos como programar esses procedimentos no SPSS.

8.9 TERMOS-CHAVE QUE DESCOBRIMOS

- Análise de Variância (ANOVA)
- Correção de Bonferroni
- F de Brown-Forsythe

- Tendência cúbica
- Graus de liberdade
- Desvios dos contrastes
- Contraste de Helmert
- ANOVA independente
- Ômega ao quadrado
- Ortogonal
- Comparações pareadas
- Contrastos planejados
- Contrastos polinomiais
- Contraste diferença (Contraste reverso de Helmert)
- Taxa de erro experimental
- Taxa de erro de conjunto
- Variância geral
- Média harmônica
- Tendência Quadrática
- Tendência Quártica
- Contraste repetido
- Contraste simples
- Ponderações (pesos)
- F de Welch
- Testes *post hoc*

8.10 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Imagine que eu esteja interessado em como diferentes métodos de ensino afetam o conhecimento dos estudantes. Notei que alguns professores são

arrogantes e distantes no seu estilo de ensinar e humilham qualquer um que se atrever a formular uma questão, enquanto outros estimulam e apoiam perguntas e comentários. Lecionei o mesmo conteúdo em três cursos de estatística. Para um grupo eu andava pela sala com uma grande vara e batia em qualquer um que fizesse uma pergunta idiota ou respondesse errado (*punitivo*). No segundo grupo utilizei meu método normal de ensino que é incentivar o aluno a discutir assuntos que eles achassem difíceis e fornecendo a qualquer um que se esforçasse uma recompensa (*incentivador*). No último grupo permaneci indiferente e não recompensei nem tampouco puni o esforço

(*indiferente*). Usei a nota nos alunos no exame (em percentual) como uma medida dependente. Com base nas teorias do condicionamento operante, espera-se que o método punitivo seja a forma malsucedida de reforço de aprendizagem, mas espera-se que o método incentivador seja bem-sucedido. Dessa forma, uma previsão é que o método incentivador irá produzir a melhor aprendizagem. Uma segunda hipótese é que o método punitivo irá retardar a aprendizagem de modo que ele é pior do que a abordagem indiferente. Os dados estão no arquivo **Teach.sav**. Execute uma ANOVA de um fator e utilize as comparações planejadas para testar as hipóteses que (1) o método incentivador resulta em melhores notas do que o método punitivo e o indiferente; e (2) o método indiferente irá produzir resultados significativamente melhores do que o punitivo. ②

- **Tarefa 2.** No Capítulo 13 (Seção 13.4) existem alguns dados sobre se comer refeições com soja reduz a contagem de espermatozoides. Dê uma olhada nessa seção, acesse os dados para esse exemplo, mas utilize a ANOVA para analisá-los. Qual é a diferença sobre o que você encontrou e o que está na Seção 13.4? Por que você acha que essa diferença surgiu? ②
- **Tarefa 3.** Estudantes (e professores) adotaram os seus celulares, o que é bastante preocupante dado as recentes controvérsias sobre a relação entre o uso de celulares e tumores cerebrais. A ideia básica é que telefones celulares emitem micro-ondas e, assim, manter um próximo à cabeça por um número elevado de horas é mais ou menos como colocar o cérebro num forno de micro-ondas e apertar o botão “cozinhar até que fique bem passado”. Se quiséssemos testar isso de forma experimental, poderíamos arranjar seis grupos de pessoas e fixar um telefone móvel na suas cabeças (de forma que eles não possam remover). Então, por controle remo-

to, ligaríamos os telefones por certo tempo cada dia. Após seis meses, mediríamos o tamanho de qualquer tumor (em mm³) próximo ao local da antena (logo atrás da orelha). Os seis grupos apresentariam 0, 1, 2, 3, 4 ou 5 horas diárias de micro-ondas do telefone por seis meses. Os dados desse exemplo estão no arquivo **Tumour.sav** (De Fiel e Hole, 2003, existe uma solução bastante detalhada no arquivo). ②

As respostas estão no arquivo **Answers (Chapter 8).pdf** no site www.artmed.com.br.

8.11 LEITURAS COMPLEMENTARES

- HOVELL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 2002. 5nd ed. Os Capítulos 11 e 12 fornecem uma abordagem detalhada da ANOVA e o Capítulo 16 cobre os MLG (Modelos Lineares Generalizados).
- IVERSEN, G. R., NORPOTH, H. *ANOVA*. Sage university paper series on quantitative applications in the social sciences, 07-001. Newbury Park (CA): Sage, 1987. 2nd ed. Texto de alto nível, mas uma boa leitura para aqueles que gostam de matemática.
- KLOCKARS, A. J., SAX, G. *Multiple comparisons*. Sage university paper series on quantitative applications in the social sciences, 07-061. Newbury Park (CA): Sage, 1986. De alto nível, mas com uma cobertura completa sobre comparações múltiplas – para mim esse livro é melhor do que o de Toothaker para comparações planejadas.
- ROSENTHAL, R. ROSNOW, R. L., RUBIN, D. B. *Contrasts and effect sizes in behavioral research: a correlational approach*. Cambridge: Cambridge University Press, 2000. Livro fantástico sobre comparações planejadas por três dos melhores escritores sobre estatística.
- ROSNOW, R. L., ROSENTHAL, R. *Beginning behavioral research: a conceptual primer*. Englewood Cliffs (NJ): Hall, 2005. 5ª ed. E eles escreveram outro ótimo livro!
- TOOTHAKER, L. E. *Multiple comparison procedures*. Sage university paper series on quantitative applications in the social sciences, 07-089. Newbury Park (CA): Sage, 1993.

Também de alto nível, mas fornece uma excelente abordagem sobre os procedimentos *post hoc*.

WRIGHT, D. B. *First steps in statistics*. London: Sage, 2002. Se esse capítulo for muito

complicado, o Capítulo 7 do Wright fornece uma introdução básica à ANOVA de fácil leitura.

ANÁLISE DE COVARIÂNCIA, ANCOVA (MLG 2)

9.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②



No capítulo anterior, vimos como a ANOVA de um fator pode ser caracterizada em termos de uma equação de regressão múltipla que utiliza variáveis *dummy* (auxiliares) para codificar a pertinência aos grupos. Além disso, no Capítulo 5 vimos como a regressão múltipla pode incorporar diversas variáveis previsoras contínuas. Dessa forma, não deve ser surpresa verificar que a equação de regressão para ANOVA pode ser estendida para incluir uma ou mais variáveis contínuas que podem prever a saída (ou variável dependente). Variáveis contínuas como essas – que não são parte da manipulação experimental principal, mas que têm influência na variável dependente – são conhecidas como *covariáveis* e podem ser incluídas em uma ANOVA. Quando mensuramos as covariáveis e as incluímos na ANOVA, denominamos o procedimento de **análise de covariância** (ou **ANCOVA**). Este capítulo está centrado nessa técnica. Começaremos apresentando a teoria da ANCOVA de uma forma breve (não se preocupe, não entrarei em

muitos detalhes) e uma hipótese adicional, a homogeneidade dos coeficientes de regressão. Também iremos mostrar um exemplo com a ajuda do SPSS e ver como interpretar e relatar os resultados dessa técnica.

9.2 O QUE É ANCOVA? ②

No último capítulo, usamos um exemplo que verificava os efeitos do Viagra na libido. Vamos pensar em outras coisas além do Viagra que podem influenciar a libido: bem, a mais óbvia é a libido do parceiro sexual do participante (são necessários dois para dançar um tango!), mas existem outras coisas, como medicamentos que podem inibir a libido (antidepressivos ou a pílula anticoncepcional) e a fadiga. Se essas variáveis (chamadas de **covariáveis**) são medidas, é possível controlar a influência que elas exercem na variável dependente pela sua inclusão no modelo de regressão. Do que aprendemos sobre regressão hierárquica (veja o Capítulo 5), deve ter ficado claro que se entrarmos com a covariável no modelo de regressão primeiro e depois entrarmos com as variáveis *dummies* (auxiliares) representando a manipulação experimental, será possível ver qual o efeito que uma

variável independente tem *após* o efeito da covariável. Assim, nós controlamos o efeito da covariável. Existem dois motivos para incluir covariáveis na ANOVA.

- **Para reduzir a variância do erro dentro do grupo:** Na discussão sobre a ANOVA e o teste *t* nos acostumamos com a ideia de que avaliamos o efeito de um experimento comparando o montante de variabilidade nos dados que o experimento pode explicar contra a variabilidade que ele não consegue explicar. Se conseguirmos explicar parte dessa variância “inexplicada” (SS_R) em termos de outras variáveis (covariáveis), reduziremos a variância do erro e determinaremos com maior precisão o efeito da variável independente (SS_M).
- **Eliminação dos confundidores:** Em qualquer experimento pode haver variáveis não medidas que confundem os resultados (isto é, variáveis que podem variar sistematicamente com a manipulação experimental). Se for sabido que uma variável pode influir na variável dependente sendo medida, a ANCOVA é ideal para remover a parcialidade dessa influência. Uma vez que uma possível variável de confusão foi identificada, ela pode ser medida e entrar na análise como uma covariável.

Existem outras razões para incluir covariáveis na ANOVA, mas como não pretendo descrever os cálculos da ANCOVA, recomendo que o leitor interessado consulte Wildt e Ahtola (1978) ou Stevens (1992, Capítulo 9).

Imagine que o pesquisador que conduziu o estudo sobre o Viagra no capítulo anterior subitamente se deu conta de que a libido do parceiro do participante poderia afetar a libido do próprio participante (especialmente porque a medida da libido foi comportamental). Ele repetiu o estudo num conjunto diferente de participantes, mas dessa vez medindo a libido do parceiro. A libido do parceiro foi medida levando-se em conta a frequência com que eles tentaram iniciar um contato sexual. No capítulo anterior, vimos que esse cenário experimental pode ser caracterizado em termos

de equação (8.2). Pense no que aprendemos sobre a regressão múltipla (Capítulo 5) e você verá que essa equação pode ser estendida para incluir essa covariável (veja a equação 9.1):

$$\begin{aligned}\text{Libido}_i &= b_0 + b_3 \text{Covariável}_i + b_2 \text{Alto}_i \\ &+ b_1 \text{Baixo}_i + \varepsilon_i \\ \text{Libido}_i &= b_0 + b_3 \text{Libido do Parceiro}_i \\ &+ b_2 \text{Alto}_i + b_1 \text{Baixo}_i + \varepsilon_i\end{aligned}\quad (9.1)$$

9.3 EXECUTANDO A ANCOVA NO SPSS ②

Os dados para esse exemplo estão na Tabela 9.1 e podem ser encontrados no arquivo **ViagraCovariate.sav**. A Tabela 9.1 mostra a libido dos participantes, dos parceiros e as médias (os desvios padrão estão entre parênteses) dos vários escores. Recomendo incluir esses dados no editor manualmente. Faça isso da mesma forma que foi feito para os dados do Viagra do capítulo anterior, exceto que uma variável a mais deve ser criada para incluir os valores da covariável.

9.3.1 Atribuindo dados ①


Os dados devem ser exibidos no editor de dados assim como eles estão na Tabela 9.1 (excluindo as linhas das médias e dos desvios padrão). Para tanto, crie uma variável código chamada de **dose** e use a opção **Labels** (Rótulos) para definir o valor dos rótulos (como no Capítulo 8, recomendo 1 = placebo, 2 = dose baixa, 3 = dose alta). Existem diferentes números de participantes em cada condição, assim, você precisa incluir nove valores de 1 nessa coluna (de modo que as primeiras nove linhas contenham o valor 1), seguido por oito linhas contendo o valor 2 e mais 14 linhas contendo o valor 3. Nesse ponto, você deve ter uma coluna com 30 linhas de dados. A seguir, crie uma segunda variável chamada **libido** e entre com os 30 escores que correspondem à libido de cada pessoa. Finalmente, crie uma terceira variável chamada de **partner (parceiro)** e use a opção **Labels** (Rótulos) para dar a essa variável o título de “*Partners libido*”

Tabela 9.1 Dados do arquivo ViagraCovariate.sav

Dose	Libido do Participante	Libido do(a) parceiro(a)
Placebo	3	5
	2	2
	5	6
	2	2
	2	3
	2	3
	7	7
	2	4
	4	5
\bar{X}	3,22	4,11
S	(1,79)	(1,76)
Dosagem baixa	7	10
	5	8
	3	6
	4	7
	4	7
	7	11
	5	9
	4	7
\bar{X}	4,88	8,13
S	(1,46)	(1,73)
Dosagem Alta	9	2
	2	3
	6	5
	3	4
	4	3
	4	3
	4	2
	6	0
	4	1
	6	3
	2	0
	8	1
	5	0
\bar{X}	4,85	2,08
S	(2,12)	(1,61)


(Libido do parceiro). Depois, entre com os 30 escores que correspondem à libido dos parceiros. Lembre que as médias (e desvios padrão) que eu incluí na tabela não são solicitadas pelo SPSS!

9.3.2 Análise principal ②

A maioria dos procedimentos da ANOVA fatorial no SPSS contém a possibilidade de incluir uma ou mais covariáveis. Entretanto, para delineamentos mais simples (a maioria dos que não incluem medidas repetidas) é melhor realizar a ANCOVA pelo procedimento fatorial geral (*general factorial*). Para acessar a caixa de diálogo principal, siga o seguinte caminho: **Analyze⇒General Linear Model⇒Univariate...** (Analisar⇒Modelo Linear Geral⇒Univariado) (veja a Figura 9.1).¹ A caixa de diálogo principal é semelhante àquela da ANOVA de um fator, exceto que existe um espaço para especificar as covariáveis. Selecione **libido** e coloque-a no espaço denominado *Dependent Variable* (Variável Dependente) clicando em . Selecione **dose** e transfira-a para o espaço denominado **Fixed Factor(s)** (Fator Fixo), depois selecione **partner** (parceiro) e coloque-a no espaço denominado *Covariate(s)* (Covariável).

9.3.3 Contrastes e outras opções ②

Existem várias caixas de diálogo que podem ser acessadas a partir da principal. Observe que se uma covariável for selecionada, os testes *post hoc* são desabilitados (você não pode acessar essa caixa de diálogo). Os testes *post hoc* não são projetados para situações nas quais uma covariável é especificada; entretanto, algumas comparações podem ser feitas usando contrastes.

Clique em  (contrastos) para acessar a caixa de diálogo *contrasts* (contrastos). Essa caixa de diálogo é diferente daquela que acessamos no Capítulo 8: naquela não é possível

¹ **Statistics⇒General Linear Model⇒GLM-General Factorial...** (Estatística⇒Modelo Linear Geral⇒MLG-Fatorial Geral) na versão 8.0 ou anteriores.

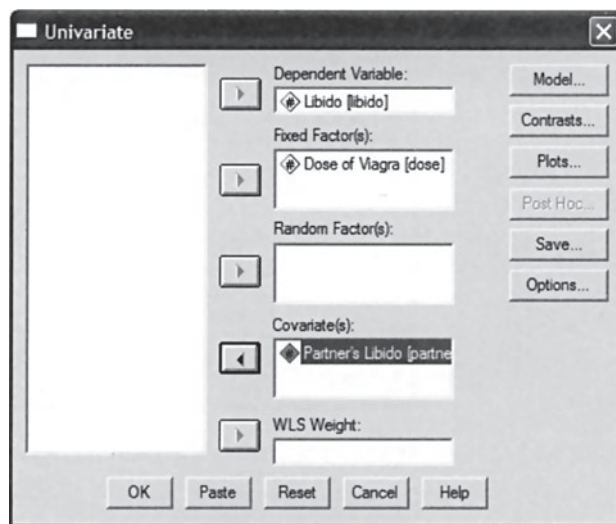


Figura 9.1 Caixa de diálogo principal para o MLG univariado.

entrar com códigos para especificar contrastes específicos. Em vez disso, é possível selecionar uma das várias possibilidades de contrastes padrão. Esses contrastes padrão foram listados na Tabela 8.6. Nesse exemplo havia uma condição de controle (o placebo codificado como o primeiro grupo), assim, um conjunto de contrastes sensíveis seria contrastes do tipo simples (**simple**) comparando cada grupo-experimental com o controle. O contraste padrão do SPSS é um contraste de desvio (**deviation**) e, para mudar isso, precisamos primeiro clicar em ☐ ao lado do quadro rotulado de *Contrast* (Contraste). Uma lista de contrastes irá aparecer automaticamente e você deve selecionar o tipo de contraste (nesse caso, **Simple** (Simple)) desta lista que irá desaparecer também automaticamente. Para contrastes simples você tem a opção de especificar a categoria de referência (que é a categoria contra a qual todos os outros grupos são comparados). Por padrão a categoria de referência é a última categoria: como nesse caso o grupo-controle foi a primeira categoria (assumindo que você codificou o placebo como 1), nós precisamos mudar essa opção selecionando ☒ **First** (Primeiro). Quando você selecionar um contraste novo, deve clicar em (Alterar) para registrar essa mudança.

A caixa de diálogo final deve ser semelhante à da Figura 9.2. Clique em (continuar) para retornar à caixa de diálogo principal.

Outra maneira de realizar os testes *post hoc* é clicando em (opções) para acessar opções da caixa de diálogo (veja a Figura 9.3). Para especificar os testes *post hoc*, selecione a variável independente (nesse caso, **dose**) do quadro denominado **Estimated Marginal Means: Factor(s)** (Médias marginais estimadas: Fator(es)) e **Factor Interactions** (In-

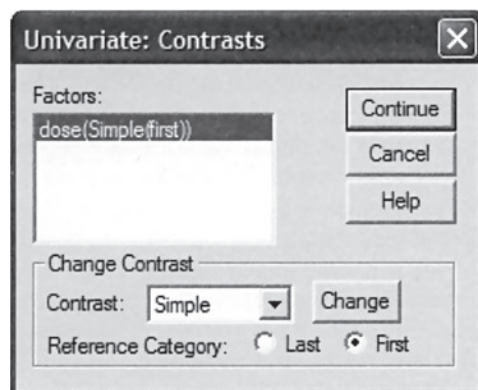


Figura 9.2 Opções para os contrastes padrão no MLG univariado.

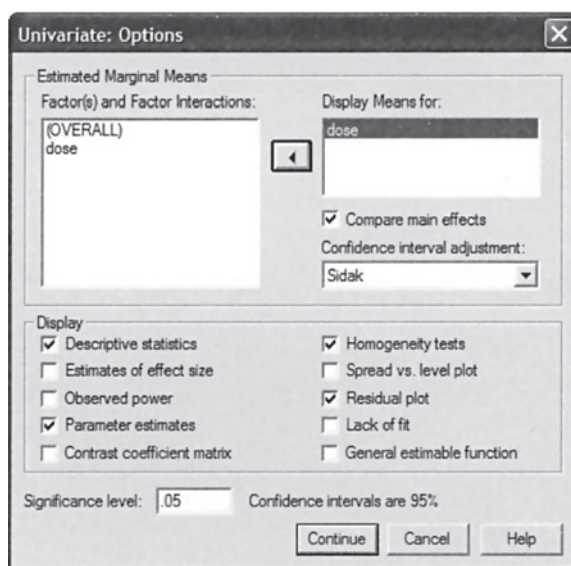


Figura 9.3 Caixa de diálogo **Options** (Opções) para o MLG univariado.

terações de fator) e transfira-a para o quadro denominado **Display Means** (Mostrar as Médias) clicando em . Uma vez que a variável foi transferida, a caixa rotulada de **Compare main effects** (Comparar efeitos principais) torna-se ativa e você deve selecionar essa opção (☒ **Compare main effects**). Se essa opção é selecionada, o quadro denominado **Confidence interval adjustment** (Ajuste do intervalo de confiança) torna-se ativo e você pode clicar em para escolher uma opção entre três possíveis níveis de ajuste. O padrão é não ter ajustamentos. A primeira opção é realizar o procedimento *post hoc* Tukey LSD (que não recomendo); a segunda é solicitar uma correção de Bonferroni (essa eu recomendo) e a opção final é uma correção de Sidak. A correção de Sidak é parecida com a correção de Bonferroni, mas menos conservadora; desse modo, deve ser selecionada se você está preocupado com a perda de poder associada aos valores corrigidos de Bonferroni. Para esse exemplo utilize a correção de Sidak (iremos usar a de Bonferroni mais adiante neste capítulo). Assim como a realização dos testes *post hoc* para a variável **dose** (dose), colocar a variável **dose** (dose) no quadro **Display Means**

for (Mostrar as Médias para) irá resultar numa tabela de médias marginais estimadas para essa variável. Essas médias fornecem uma estimativa das médias do grupo ajustadas (isto é, as médias após a contribuição da covariável). Quando você tiver selecionado as opções solicitadas (veja o Quadro 9.1), clique em **Continue** (continuar) para retornar à caixa de diálogo principal. Existem outras opções disponíveis na caixa de diálogo principal. Por exemplo, se você tem diversas variáveis independentes, pode representá-las graficamente umas contra as outras (o que é útil para interpretar efeitos de interação – veja a Seção 10.3.3). Para essa análise existe somente uma variável independente, assim, podemos clicar em **OK** a fim de executar a análise.

9.4 INTERPRETANDO AS SAÍDAS DA ANCOVA ②

9.4.1 Análise principal ②

A Saída 9.1 do SPSS mostra (como ilustração) a tabela da ANOVA para esses dados quando a covariável não está incluída. Fica claro a partir do valor da significância, maior do que 0,05, que não há diferenças na libido

Quadro 9.1

Opções para a ANCOVA ②

As opções restantes nessa caixa de diálogo são as seguintes:

- **Descriptives Statistics (Estatísticas descritivas):** Essa opção produz uma tabela com as médias e os desvios padrão de cada grupo.
- **Estimates of effect size (Estimativas do tamanho de efeito):** Essa opção produz o valor do eta quadrado (η^2), descrito na Seção 8.5, que é uma medida do tamanho de efeito experimental. Na verdade, o eta quadrado é o coeficiente de regressão (R^2) para uma linha de regressão não-linear (isto é, uma curva) passando supostamente pelas médias de todos grupos. Na população essa suposição é verdadeira, mas nas amostras não; portanto, o eta quadrado é geralmente viciado (veja a Seção 8.5 e Howell, 2002, Seção 11.11). Por esse motivo, essa opção é pouco recomendada e o tamanho de efeito pode ser melhor estimado utilizando-se o ômega ao quadrado (ω^2) – veja a Seção 9.7.
- **Observed power (O poder observado):** Essa opção fornece uma estimativa da probabilidade de que o teste estatístico detecte a diferença entre as médias dos grupos (veja a Seção 1.8.5). Essa medida não é muito utilizada porque se o teste F é significativo, a probabilidade de que o efeito tenha sido detectado será, é claro, alta. Da mesma forma, se as diferenças entre os grupos forem pequenas, o poder observado será baixo. O poder observado é de pouca utilidade e devo aconselhar que os cálculos do poder (exceto o do tamanho amostral) sejam feitos antes que o experimento seja conduzido (veja Cohen, 1988, 1992; Howell, 2002, para ver como fazer isso manualmente; Field, 1998b, para ver como fazer os cálculos usando o computador; ou utilize o *software* grátis G* Power disponível no *site* www.artmed.com.br ou em <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>).
- **Parameter estimates (Estimativas Paramétricas):** Essa opção produz uma tabela dos coeficientes de regressão e suas significâncias para as variáveis do modelo de regressão (veja a Seção 9.5).
- **Contrast coefficient matrix (Matriz de coeficiente de contraste):** Essa opção produz matrizes dos valores dos códigos utilizados por qualquer contraste na análise. É útil apenas para checar quais grupos estão sendo comparados em quais contrastes.
- **Homogeneity tests (Testes de homogeneidade):** Essa opção produz o teste de Levene de homogeneidade das variâncias (veja as Seções 3.6 e 8.4.1).
- **Spread vs. Level plot (Espalhamento vs. diagrama de nível):** Essa opção produz um diagrama que mostra a média de cada grupo de um fator (eixo-X) contra o desvio padrão daquele grupo (eixo-Y). É um recurso útil para verificar que não existe relacionamento entre a média e o desvio padrão. Se existe um relacionamento, talvez seja necessário estabilizar os dados utilizando uma transformação logarítmica (veja o Capítulo 3).
- **Residual plots (Diagrama dos resíduos):** Essa opção produz diagramas de valores residuais padronizados observados vs. previstos. Esses diagramas podem ser usados para avaliar a hipótese da igualdade de variância.

entre os três grupos; portanto, o Viagra parece não ter efeito significativo na libido. Você deve ter notado, também, que o total da variação a ser explicada (SS_T) foi 110,97 (*Total Corrected*), sendo que a manipulação experimental foi responsável por apenas 16,84 unidades (SS_M), enquanto 94,12 não foram explicadas (SS_R).

A Saída 9.2 do SPSS mostra os resultados do teste de Levene e a tabela da ANOVA quando a libido dos parceiros é incluída no modelo como uma covariável. O teste de Levene é significativo, indicando que as variâncias dos grupos não são iguais (portanto, a hipótese da homogeneidade da variância

Saída 9.1 do SPSS

Tests of Between-Subjects Effects (Testes de efeitos entre participantes)

Dependent Variable: Libido (Variável Dependente: Libido)

Source (Fonte)	Type III Sum of Squares (Soma dos quadrados do tipo III)	df (gl)	Mean Square (Média ao quadrado)	F	Sig.
Corrected Model (Modelo Corrigido)	16.844 ^a	2	8.422	2.416	0.108
Intercept (Intercepto)	535.184	1	535.184	153.533	0.000
DOSE (Dose)	16.844	2	8.422	2.416	0.108
Error (Erro)	94.123	27	3.486		
Total (Total)	683.000	30			
Corrected Total (Total corrigido)	110.967	29			

a. R Squared = .152 (Adjusted R Squared = .089 (R ao quadrado = 0,152 (R ao quadrado ajustado = 0,089))

foi violada). Entretanto, como foi mencionada na Seção 3.6, o teste de Levene não é necessariamente a melhor maneira de julgar se as variâncias são desiguais o suficiente para causar problemas. Uma boa maneira de se assegurar disso é verificar novamente os valores da maior e da menor variância. Para os nossos três grupos temos os desvios padrão de 1,79 (grupo-placebo), 1,46 (grupo de baixa dosagem) e 2,12 (grupo de alta dosagem) – veja a Tabela 9.1. Se elevarmos ao quadrado esses valores, teremos variâncias de: 3,20 (grupo-placebo), 2,13 (grupo de baixa dosagem) e 4,49 (grupo de alta dosagem). Podemos, então, dividir a maior variância pela menor: nesse caso, $4,49/2,13 = 2,11$. Se o valor resultante for menor do que 2, provavelmente não precisaremos nos preocupar muito; se for maior do que 2 (como é o caso aqui), provavelmente teremos problemas! Entretanto, não vamos nos preocupar muito com as diferenças entre as variâncias por enquanto.

Como interpreto a ancova?



O formato dessa tabela da ANOVA é, em grande parte, o mesmo que a da tabela sem a covariável, exceto que agora temos uma linha adicional com informações sobre a covariável

(**partner** (parceiro)). Observando os valores das significâncias, fica claro que a covariável prevê significativamente a variável dependente porque

esse valor é menor do que 0,05. Portanto, a libido de um participante é influenciada pela libido do parceiro. O mais interessante é que quando o efeito da libido do parceiro é removido, o efeito do Viagra se torna significativo (p é 0,016, o que é menor do que 0,05). A quantidade de variação de responsabilidade do modelo (SS_M) aumentou para 34,75 unidades (*corrected model* (modelo corrigido)), da qual o Viagra é responsável por 28,34 unidades. Ainda mais importante, a grande variação na libido que é de responsabilidade da covariável significa que a variância não-explicada (SS_R) foi reduzida para 76,22 unidades. Note que o SS_T não mudou; o que mudou foi como a variação total é explicada.

Esse exemplo ilustra como a ANCOVA pode auxiliar a exercer um controle experimental mais rígido, levando em consideração as variáveis de confusão, para nos dar uma medida “mais pura” do efeito da manipulação experimental. Sem levar em consideração a libido dos parceiros participantes, podemos concluir que o Viagra não teve efeito na libido, mas, na verdade, ele teve. Entretanto, o efeito da libido do parceiro parece mais forte do que o do Viagra. Retomando as médias do grupo da Tabela 9.1 para os dados da libido, fica claro que a significância da ANOVA reflete a diferença entre o grupo-placebo e os dois grupos-experimentais (porque os grupos de doses alta e baixa têm médias semelhantes – 4,88 e 4,85 – enquanto a média do grupo-placebo é bem mais baixa, 3,22). Contudo, será necessário realizar alguns contrastes para verificar isso.

Saída 9.2 do SPSS

Levene's Test of Equality of Error Variances^a
(Teste de Levene para a igualdade das variâncias do erro)

Dependent Variable: *Libido* (Variável dependente: libido)

<i>F</i>	<i>df1</i> (gl1)	<i>df2</i> (gl2)	<i>Sig.</i>
5.525	2	27	0.010

Testa a hipótese nula de que a variância do erro é a mesma em todos os grupos.
a. Delineamento: intercepto + Parceiro + Dose.

Tests of Between-Subjects Effects (Testes de efeitos entre participantes)

Dependent Variable: *Libido* (Variável dependente: libido)

<i>Source</i> (Fonte)	<i>Type III Sum of Squares</i> (Soma dos quadrados do tipo III)	<i>df</i> (gl)	<i>Mean Square</i> (Média ao quadrado)	<i>F</i>	<i>Sig.</i>
<i>Corrected Model</i> (Modelo Corrigido)	34.750 ^a	3	11.583	3.952	0.019
<i>Intercept</i> (Intercepto)	12.171	1	12.171	4.152	0.052
<i>PARTNER</i> (Parceiro)	17.906	1	17.906	6.109	0.020
<i>DOSE</i> (Dose)	28.337	2	14.169	4.833	0.016
<i>Error</i> (Erro)	76.216	26	2.931		
<i>Total</i> (Total)	683.000	30			
<i>Correlated Total</i> (Total correlacionado)	110.967	29			

a. R quadrado = 0,313 (R quadrado ajustado = 0,234)

A Saída do SPSS 9.3 mostra as estimativas dos parâmetros selecionadas na caixa de diálogo **options** (opções). Essas estimativas são calculadas usando a análise de regressão com a variável **dose** dividida entre duas variáveis *dummy* (auxiliares) codificadoras (veja a Seção 8.2.2 e a Seção 9.5). O SPSS codifica as duas variáveis auxiliares de forma que a última categoria (a categoria codificada com o valor mais alto no editor de dados – nesse exemplo, o grupo de alta dosagem) seja a de referência. Essa categoria de referência (rotulada de DOSE = 3 na saída) é codificada com 0 para as duas variáveis auxiliares (veja a Seção 8.2.2 para lembrar como a codificação com variáveis auxiliares funciona). DOSE = 2, portanto, representa a diferença entre o grupo codificado como 2 (baixa dosagem) e a categoria de referência (alta dosagem), e DOSE = 1 representa a diferença entre o grupo codificado como 1 (placebo) e a categoria de referência (alta dosagem). Os valores de *b* literalmente representam as diferenças entre as médias desses grupos e, assim, a significância dos testes *t* nos diz se as médias dos grupos diferem significativamente. Os graus de

liberdade para esses testes *t* podem ser calculados como na regressão normal (veja a Seção 5.2.4) utilizando $N - p - 1$, no qual N é o do tamanho da amostra total (nesse caso 30) e p é o número dos previsores (nesse caso, 3, as duas variáveis auxiliares e a covariável – veja a Equação (9.1)). Para esses dados, temos $gl = 30 - 3 - 1 = 26$.
Portanto, podemos concluir dessas estimativas que o grupo de alta dosagem difere significativamente do grupo-placebo (DOSE = 1 na tabela), mas que o grupo de alta dosagem difere significativamente, também, do grupo de baixa dosagem (DOSE = 2 na tabela). Essa última conclusão é peculiar porque contradiz o que inicialmente concluímos da ANOVA (lembre que as médias das dosagens altas e baixas eram praticamente idênticas) – você consegue imaginar por quê? (Isso será revelado no devido momento!). Por último, note o valor de *b* para a covariável (0,483). Esse valor nos diz que, mantendo tudo igual, se a libido do parceiro aumenta em uma unidade, a libido do participante aumentará apenas meia unidade (embora nada sugira uma ligação casual entre as duas). O sinal desse coe-

Saída 9.3 do SPSS

Parameter Estimates (Estimativas dos Parâmetros)

Dependent Variable: Libido (Variável dependente: libido)

Parameter (Parâmetro)	B	Std. Error (Erro Padrão)	t	Sig.	95% Confidence Interval (Intervalo de 95% de confiança)	
					Lower Bound (Limite inferior)	Upper Bound (Limite superior)
Intercept (Intercepto)	3.843	0.625	6.150	0.000	2.558	5.127
PARTNER (Parceiro)	0.483	0.196	2.472	0.020	0.081	0.085
(DOSE = 1)	-2.607	0.842	-3.095	0.005	-4.338	-0.876
(DOSE = 2)	-2.894	1.411	-2.051	0.050	-5.794	0.006
(DOSE = 3)	0 ^a

a. Esse parâmetro é fixado em zero porque ele é redundante.

ficiente nos mostra a direção do relacionamen-
to entre a covariável e a variável saída. Assim,
nesse exemplo, o coeficiente positivo significa
que a libido do parceiro tem um relacionamento
positivo com a libido do participante: quando
um aumenta, o outro também cresce. Um coefi-
ciente negativo significaria o contrário: à medi-
da que um aumenta, o outro diminui.

9.4.2 Contrastes ②

A Saída do SPSS 9.4 mostra o resultado da
análise dos contrastes especificados na Figura
9.2 e compara o nível 2 (baixa dosagem) com
o nível 1 (placebo) como uma primeira compa-
ração, e o nível 3 (alta dosagem) com o nível
1 (placebo) como uma segunda comparação.

Saída 9.4 do SPSS

Contrast Results (K Matrix) (Resultados dos contrastes (Matriz K))

Dose of Viagra (Dose do Viagra)			Dependent Variable (Variável Dependente)
Simple Contrast ^a (Contraste Simples)			Libido
Level 2 vs. Level 1 (Nível 2 vs. Nível 1)	Contrast Estimate (Estimativa do Contraste)		-0.287
	Hypothesized Value (Valor Hipotético)		0
	Difference (Estimate – Hypothesized) (Diferença – Hipotética – Estimada)		
	Std Erro (Erro Padrão)		-0.287
	Sig. (Sig.)		1.144
	95% Confidence Interval for Difference (Intervalo de Confiança de 95% para a diferença)		0.804
			-2.638
		Lower Bound (Limite Inferior)	
		Upper Bound (Limite Superior)	2.064
Level 3 vs. Level 1 (Nível 3 vs. Nível 1)	Contrast Estimate (Estimativa do Contraste)		2.607
	Hypothesized Value (Valor Hipotético)		0
	Difference (Estimate – Hypothesized) (Diferença – Hipotética – Estimada)		
	Std Erro (Erro Padrão)		2.607
	Sig. (Sig.)		0.842
	95% Confidence Interval for Difference (Intervalo de Confiança de 95% para a diferença)		0.005
			0.876
		Lower Bound (Limite Inferior)	
		Upper Bound (Limite Superior)	4.338

a Categoria de referência = 1.

Esses contrastes são consistentes com o que foi especificado: todos os grupos são comparados com o primeiro grupo. As diferenças dos grupos são apresentadas: o valor da diferença, o erro padrão, o valor da significância e o intervalo de 95% de confiança. Esses resultados mostram que o grupo de baixa dosagem não teve uma libido significativamente diferente do que a do grupo-placebo (contraste 1, $p = 0,804$), mas que o grupo da alta dosagem diferiu significativamente do grupo-placebo ($p = 0,005$). Esses resultados são consistentes com as estimativas de regressão paramétrica (na verdade, note que o contraste 2 é idêntico aos parâmetros da regressão para $\text{DOSE} = 1$ da seção anterior).

Novamente, tudo isso parece muito estranho: o efeito significativo da libido parece refletir uma diferença entre o grupo-placebo e os dois grupos do Viagra (que apresentam médias com valores semelhantes), mas os contrastes até agora contradizem essas conclusões. A razão para essa inconsistência é que a conclusão inicial foi baseada nas médias dos grupos que não tinham sido ajustadas para o efeito da covariável. Esses valores nada informam sobre as diferenças dos grupos refletidas na significância da ANCOVA. A Saída 9.5 do SPSS fornece os valores ajustados das médias dos grupos e são esses os valores que devem ser usados para a interpretação (essa é a razão principal para selecionar a opção **Display Means for** (Mostre as Médias para)). As médias ajustadas mostram um padrão bem diferente de respostas: parece que a significância da ANCOVA reflete a diferença entre o grupo de alta dosagem e os grupos de baixa dosagem e placebo. Os grupos de baixa dosagem e placebo parecem ter médias ajustadas bem similares, indicando que uma dosagem baixa de Viagra não aumenta a libido acima de níveis normais, sendo necessário para tanto uma alta dosagem! Essas conclusões sustentam o que sabemos dos contrastes e dos parâmetros da regressão, mas podem ser verificadas com os testes *post hoc* especificados no menu **options** (opções).

A Saída 9.6 do SPSS mostra os resultados das comparações *post hoc* corrigidas pelo método de Sidak que foram solicitadas como

Saída 9.5 do SPSS

Estimates (Estimativas)

Dependent Variable: Libido (Variável dependente: libido)

Dose of Viagra (Dose de Viagra)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
Placebo (Placebo)	3.313 ^a	0.572	2.138	4.489
Low Dose (Baixa Dosagem)	3.027 ^a	0.962	1.049	5.004
High Dose (Alta Dosagem)	5.920 ^a	0.644	4.597	7.244

a. As covariáveis que aparecem no modelo são avaliadas nos seguintes valores: Libido do Parceiro = 4,30.

parte da caixa de diálogo **options** (opções). A diferença significativa entre o grupo de alta dosagem e o placebo permanece, mas a diferença significativa entre os grupos de alta e baixa dosagem pelos parâmetros da regressão (Saída 9.3 do SPSS) desapareceu (p é somente 0,14). Essa contradição pode ser resultado de uma perda de poder nos testes *post hoc* (lembre que as comparações planejadas têm mais poder para detectar efeitos do que procedimentos *post hoc*). Entretanto, poderia haver outras razões para essas comparações serem não-significativas e devemos ter cautela nas nossas interpretações da significância da ANCOVA e das comparações subsequentes.

9.4.3 Interpretando as covariáveis ②

Já mencionei que as estimativas dos parâmetros (Saída 9.3 do SPSS) nos mostram como interpretar a covariável. Se o valor de b para a covariável for positivo, isso significa que a covariável e a variável de saída tem um relacionamento positivo (à medida que a covariável aumenta, o valor de saída também aumenta). Se o valor de b for negativo significa o oposto: a covariável e a variável

Saída 9.6 do SPSS

Pairwise Comparisons (Comparações aos pares)

Dependent Variable: Libido (Variável dependente: libido)

(I) Dose of Viagra (I) Dose de Viagra	(J) Dose of Viagra (J) Dose de Viagra	Mean Difference (Diferença Média)(I - J)	Std. Error (Erro Padrão)	Sig. ^a	95% Confidence Interval for Difference ^a (Intervalo de 95% de confiança para a diferença)	
					Lower Bound (Limite inferior)	Upper Bound (Limite Superior)
Placebo (Placebo)	Low Dose (Baixa Dosagem)	0.287	1.144	0.992	-2.631	3.205
	High Dose (Alta Dosagem)	-2.607*	0.842	0.014	-4.756	-0.458
Low Dose (Baixa Dosagem)	Placebo (Placebo)	-0.287	1.144	0.992	-3.205	2.631
	High Dose (Alta Dosagem)	-2.894	1.411	0.144	-6.493	0.706
High Dose (Alta Dosagem)	Placebo (Placebo)	2.607*	0.842	0.014	0.458	4.756
	Low Dose (Baixa Dosagem)	2.894	1.411	0.144	-0.706	6.493

Baseado nas médias marginais estimadas.
* A diferença média é significativa ao nível de 0,05.
a. Ajuste para as comparações múltiplas: Sidak.

de saída têm um relacionamento negativo (à medida que a covariável aumenta, a variável de saída diminui). Para esses dados o valor de *b* foi positivo, indicando que à medida que a libido do parceiro aumenta, também aumenta a do participante. Outra maneira de descobrir a mesma coisa é simplesmente construir um diagrama de dispersão da covariável contra a saída. Aprendemos sobre diagramas de dispersão na Seção 4.4, então dê uma olhada lá para descobrir como se faz um. A Figura 9.4 mostra o diagrama de dispersão resultante para esses dados e confirma o que já sabemos: o efeito da covariável é que à medida que a libido do parceiro aumenta, o mesmo acontece com a libido do participante (como mostra a inclinação da linha da regressão).

9.5 ANCOVA EXECUTADA COMO UMA REGRESSÃO MÚLTIPLA ②

Embora a ANCOVA esteja basicamente concluída, talvez seja interessante refazer a análise como uma regressão múltipla hierárquica. Como um exercício, entre esses dados e conduza uma análise você mesmo acrescentando duas variáveis auxiliares ao arquivo **ViagraCova-**

riate.sav que usamos neste capítulo (veja a Seção 8.2.2 para ajuda com a codificação da variável auxiliar). Se você não souber o que fazer, incluí um conjunto completo de dados chamado **ViagraCovariate Dummy.sav** no site www.artmed.com.br. Para conduzir a análise, usamos o procedimento de regressão (veja o Capítulo 5) com a **libido** como saída (variável dependente na terminologia do SPSS), depois entramos com a libido do parceiro como um preditor no primeiro bloco (**partner**) e, então, entramos com as duas variáveis *dummies* (auxiliares) no segundo bloco. Em todos os casos, o método de entrada deve ser o **Enter** (consulte a Seção 5.7 para ver como entrar com variáveis em uma regressão hierárquica, mas lembre-se de usar o método **Enter** para ambos os blocos). O resumo do modelo de regressão resultante (Saída 9.7 do SPSS) mostra a aderência do modelo primeiro quando somente a covariável é usada e em seguida quando tanto a covariável quanto as variáveis auxiliares são utilizadas.



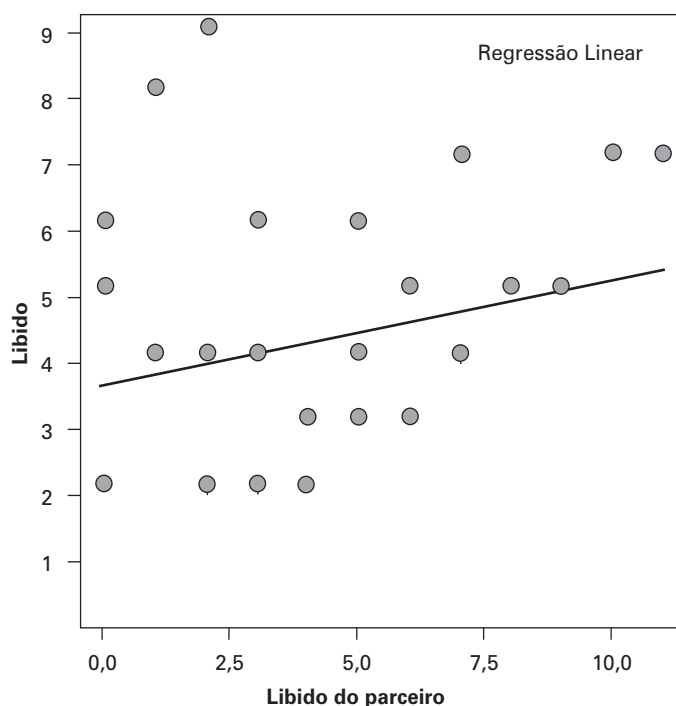


Figura 9.5 Diagrama de dispersão da libido do participante contra a libido do parceiro.

Portanto, a diferença entre os valores de R^2 ($0,313 - 0,058 = 0,255$) representa a contribuição individual da dose de Viagra. Assim, podemos dizer que a dose de Viagra é responsável por 25,5% da variação na libido enquanto a libido do parceiro é responsável por somente 5,8%. Essa informação adicional fornece uma percepção sobre a importância do Viagra. A próxima tabela é a da ANOVA, que é novamente dividida em duas seções. A metade superior representa apenas o efeito da covariável e a metade inferior representa todo o modelo (isto é, incluindo a covariável e as dosagens de Viagra). Note que na parte inferior da tabela da ANOVA (a parte do Modelo 2) todo o modelo (a libido do parceiro e as variáveis auxiliares) são responsáveis por 34,57 unidades da variância (SS_M), existem 110,97 unidades no total (SS_T) e a variância não-explicada (SS_R) é 76,22. A metade inferior, portanto, contém os mesmos valores que a tabela resumo da ANCOVA da saída 9.2 do SPSS. Esses valores são os mesmos da linha rotulada

linha *corrected model* (modelo corrigido) da tabela resumo da ANCOVA que encontramos quando executamos a análise como ANCOVA (veja a Saída 9.2 do SPSS).

A Saída 9.8 do SPSS mostra o restante da análise de regressão. Essa tabela de coeficientes de regressão é mais interessante. Novamente, essa tabela é dividida em duas e, assim, a parte inferior da tabela se parece com todo o modelo. Quando a dosagem do Viagra é considerada com a covariável, o valor de b para a covariável é 0,483, o que corresponde ao valor estimado do parâmetro na ANCOVA (Saída 9.3 do SPSS). Os valores de b para as variáveis auxiliares representam as diferenças entre as médias dos grupos de baixa dosagem e o grupo-placebo (**dummy1** – auxiliar 1) e o grupo de alta dosagem e o grupo-placebo (**dummy2** – auxiliar 2) – veja a Seção 8.2.2 para saber por quê. As médias dos grupos de alta e baixa dosagem foram de 4,88 e 4,85, respectivamente, e a média do grupo-placebo

Saída 9.7 do SPSS

Model Summary (Resumo do Modelo)

<i>Model</i> (Modelo)	<i>R</i>	<i>R Square</i> (R ao Quadrado)	<i>Adjusted R Square</i> (R ao Quadrado Ajustado)	<i>Std. Error of the Estimate</i> (Erro padrão da Estimativa)
1	0.240 ^a	0.058	0.024	1.932
2	0.560 ^b	0.313	0.234	1.712

a. Previsores (Constante), Libido do parceiro.

b. Previsores (Constante), Libido do parceiro, Variável Auxiliar 1 (Placebo versus Alta Dosagem), Variável Auxiliar 1

c. Placebo versus baixa dosagem.

ANOVA (Análise de Variância)^c*Dependent Variable: Libido* (Variável dependente: libido)

<i>Model</i> (Modelo)		<i>Sum of Squares</i> (Soma dos quadrados)	<i>df</i> (gl)	<i>Mean Square</i> (Média ao quadrado)	<i>F</i>	<i>Sig.</i>
1	<i>Regression</i> (Regressão)	6.413	1	6.413	1.717	0.201 ^a
	<i>Residual</i> (Resíduos)	104.554	28	3.734		
	<i>Total</i> (Total)	110.967	29			
2	<i>Regression</i> (Regressão)	34.750	3	11.583	3.952	0.019 ^b
	<i>Residual</i> (Resíduos)	76.216	26	2.931		
	<i>Total</i> (Total)	110.967	29			

a. Previsores (Constante), Libido do parceiro.

b. Previsores (Constante), Libido do parceiro, Variável Auxiliar 1 (Placebo versus Alta Dosagem), Variável Auxiliar 1 (Placebo versus Baixa Dosagem)

c. Variável dependente: Libido.

Saída 9.8 do SPSS

Coefficients^a (Coeficientes)

<i>Model</i> (Modelo)		<i>Unstandardized Coefficients</i> (Coeficientes não-padronizados)		<i>Standardized Coefficients</i> (Coeficientes padronizados)	<i>t</i>	<i>Sig.</i>
		<i>B</i>	<i>Std. Error</i> (Erro Padrão)	<i>Beta</i>		
1	<i>(Constant)</i> (Constante)	3.689	0.626		5.894	0.000
	<i>Partner's Libido</i> (Libido do Parceiro)	0.158	0.120	0.240	1.311	0.201
2	<i>(Constant)</i> (Constante)	1.236	0.986		1.253	0.221
	<i>Partner's Libido</i> (Libido do Parceiro)	0.483	0.196	0.737	2.472	0.020
	<i>Dummy Variable 1</i> (Placebo vs. Low) (Variável Auxiliar 1 (Placebo versus Baixa Dosagem))	-0.287	1.144	-0.066	-0.251	0.804
	<i>Dummy Variable 2</i> (Placebo vs. Low) (Variável Auxiliar 2 (Placebo versus Alta Dosagem))	2.607	0.842	0.672	3.095	0.005

a. Variável dependente: Libido.

foi de 3,22. Portanto, os valores de *b* para as duas variáveis auxiliares devem ser aproximadamente os mesmos ($4,88 - 3,22 = 1,66$ para **dummy1** (auxiliar 1) e $4,85 - 3,22 = 1,63$

para **dummy2** (auxiliar 2)). Os mais espertos podem notar da saída do SPSS que, de fato, os valores de *b* não são somente muito diferentes entre si (o que não deveria ser o caso

porque as médias do grupo de alta e baixa dosagem são praticamente as mesmas), mas são diferentes dos valores que acabei de calcular. Isso quer dizer que menti para vocês nas últimas 50 páginas sobre o que os valores de beta representam? Não, eu não sou tão horrível: essa aparente anormalidade ocorre porque os valores de b nessa regressão representam as diferenças entre *as médias ajustadas*, não as médias originais; isto é, a diferença entre a média de cada grupo e o placebo quando essas médias foram ajustadas para a libido do parceiro. Os valores ajustados são dados na saída 9.5 do SPSS e dessa tabela podemos ver que:

$$\begin{aligned} b_{\text{Auxiliar 1}} &= \bar{X}_{\text{Baixa(ajustada)}} - \bar{X}_{\text{Placebo(ajustado)}} \\ &= 3,027 - 3,313 = -0,286 \\ b_{\text{Auxiliar 2}} &= \bar{X}_{\text{Alta(ajustada)}} - \bar{X}_{\text{Placebo(ajustado)}} \\ &= 5,920 - 3,313 = -2,607 \end{aligned} \quad (9.12)$$

Esses são os valores que você pode ver na tabela do SPSS. Os testes t conduzidos nesses valores mostram que a significância da ANCOVA foi resultado de uma diferença signifi-

cativa entre os grupos de alta dosagem e placebo.² Não havia diferença significativa entre os grupos de baixa dosagem e do placebo. Você deve notar, também, que a significância dos valores de t é a mesma que vimos na tabela de contrastes na ANCOVA original (veja a saída 9.4 do SPSS). Finalmente, nós não sabemos se havia uma diferença entre os grupos de alta e baixa dosagem: para descobrir isso precisaremos usar diferentes codificações de variáveis auxiliares (talvez comparando a alta e a baixa com o placebo e, então, comparando a alta com a baixa como fazíamos para as comparações planejadas no Capítulo 8 – veja o Quadro 9.2).

² Como mencionei anteriormente neste capítulo, os graus de liberdade para esses testes t são $N - p - 1$, como em qualquer análise de regressão (veja a Seção 5.2.4). O N é o número do tamanho total da amostra (nesse caso, 30) e o p é o número de precursores (nesse caso, 3, as duas variáveis auxiliares e a covariável – veja a Equação (9.1). Para esses dados, temos $gl = 30 - 3 - 1 = 26$.

Quadro 9.2

Contrastes planejados para a ANCOVA ③



Você deve ter notado que embora possamos solicitar ao SPSS que faça certos contrastes padrão, não existe uma opção para especificar contrastes planejados como havia com a ANOVA independente de um fator (veja a Seção 8.3.1). Entretanto, esses contrastes podem ser feitos se executarmos a ANCOVA por meio do menu da regressão. Imagine que você escolha alguns contrastes planejados como no Capítulo 8, no qual o primeiro contraste comparou o grupo-placebo com todas as dosagens do Viagra e o segundo contraste comparou as dosagens alta e baixa (veja a Seção 8.2.10). Vimos nas Seções 8.2.10 e 8.3.1 que para fazer isso no SPSS precisamos entrar com certos números para codificar esses contrastes. Para o primeiro contraste, descobrimos que um conjunto apropriado de códigos seria -2 para o grupo-placebo e 1 para os grupos de dosagem alta e baixa. Para o segundo contraste, os códigos seriam 0 para o grupo-placebo, -1 para o grupo de dosagem baixa e 1 para o grupo de dosagem alta (veja a Tabela 8.4). Se você quiser fazer esses contrastes para a ANCOVA, deve inserir esses valores como duas variáveis auxiliares (*dummy*). Assim, tomando os dados desse exemplo, acrescentamos uma coluna chamada de **Dummy1** e naquela coluna colocamos o valor -2 para cada pessoa do grupo-placebo e o valor de 1 para os demais participantes. Precisamos, então, acrescentar uma segunda coluna chamada de **Dummy2** na qual colocamos um 0 para cada um dos participantes do grupo-placebo, -1 para todos no grupo de baixa dosagem e o valor de 1 para os do grupo de alta dosagem. Os dados completos seriam como os do arquivo **ViagraCovariateContrasts.sav** no site www.artmed.com.br.

(Continua)

Quadro 9.2 (Continuação)

Coefficients^a (Coeficientes)

Model (Modelo)	Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
	B	Std. Error (Erro Padrão)	Beta		
1 (Constant) (Constante)	3.689	0.626		5.894	0.000
Partner's Libido (Libido do Parceiro)	0.158	0.120	0.240	1.311	0.201
2 (Constant) (Constante)	2.009	0.986		2.038	0.052
Partner's Libido (Libido do Parceiro)	0.483	0.196	0.737	2.472	0.020
Dummy Variable 1 (Placebo vs. Low & High) (Variável Auxiliar 1 (Placebo versus Baixa e Alta))	0.387	0.238	0.276	1.623	0.117
Dummy Variable 2 (Low vs. High) (Variável Auxiliar 2 (Baixa versus Alta Dosagem))	1.447	0.705	0.617	2.051	0.050

a. Variável dependente: Libido.



Execute a análise como descrito na Seção 9.5. A saída resultante irá iniciar com um resumo do modelo e uma tabela da ANOVA que deve ser idêntica àquelas na Saída 9.7 do SPSS (porque fizemos as mesmas coisas que anteriormente, a única diferença é como a variância do modelo é subsequentemente separada com os contrastes). Os coeficientes de regressão para as variáveis auxiliares serão diferentes porque agora especificamos códigos distintos.

A primeira variável auxiliar compara o grupo-placebo com os grupos de alta e baixa dosagem. Assim, ela compara a média ajustada do grupo-placebo (3,313) com o valor médio das médias ajustadas para os grupos de baixa e alta dosagem $(3,027 + 5,920)/2 = 4,474$. O valor de b para a variável auxiliar deve, portanto, ser diferente entre esses valores: $4,474 - 3,313 = 1,16$. Entretanto, também descobrimos, numa seção complexa e chata (8.2.10.2) que esse valor é dividido pelo número de grupos dentro do contraste (isto é, 3), assim, ele será $1,16/3 = 0,387$ (como ele será na saída). A estatística t associada não é significativa, indicando que o grupo-placebo não foi significativamente diferente da média combinada dos grupos do Viagra.

A segunda variável auxiliar compara os grupos de alta e baixa dosagem, e assim o valor de b deve ser a diferença entre as médias ajustadas desses grupos: $5,920 - 3,027 = 2,89$. Também descobrimos na Seção 8.2.10.2 que esse valor é dividido pelo número dos grupos dentro do contraste (isto é, 2) e assim será $2,89/2 = 1,447$ (como na saída). A estatística t associada é significativa (sua significância é exatamente 0,05), indicando que o grupo de alta dosagem produziu uma libido significativamente mais alta do que o grupo de baixa dosagem após o controle do efeito da libido do parceiro.

Isso ilustra como você pode aplicar os princípios da Seção 8.2.10 para a ANCOVA. Embora o SPSS não forneça uma interface fácil para fazer os contrastes planejados, eles podem ser feitos se você usar os menus de regressão em vez dos da ANOVA!

9.6 SUPOSIÇÕES ADICIONAIS DA ANCOVA ③

9.6.1 Homogeneidade dos parâmetros da regressão ③

Quando uma ANCOVA é realizada, olhamos para o relacionamento total entre a saída

(variável dependente) e a covariável: ajustamos uma linha de regressão para todo o conjunto de dados, ignorando o grupo que a pessoa pertence. Ajustando esse modelo nós, portanto, assumimos que esse relacionamento é verdadeiro para todos os grupos de participantes. Por exemplo, se existe um relacionamento positivo

entre a covariável e a saída em um grupo, assumimos que existe um relacionamento positivo em todos os outros grupos também. Se, entretanto, o relacionamento entre a saída (variável dependente) e a covariável difere entre os grupos, todo o modelo de regressão não é preciso (ele não representa todos os grupos). Essa suposição é muito importante e é chamada de hipótese da *homogeneidade dos parâmetros da regressão*. A melhor maneira de entender essa suposição é imaginar um diagrama de dispersão para cada condição experimental com a covariável num eixo e a saída em outro. Se você calculou e traçou a linha de regressão para cada um desses diagramas de dispersão,

deve descobrir que as linhas de regressão são mais ou menos semelhantes (isto é, os valores de b em cada grupo devem ser iguais).

A Figura 9.5 mostra os diagramas de dispersão que exibem o relacionamento entre a libido do parceiro (a covariável) e a saída (a libido do participante) para cada uma das três condições experimentais. Em cada diagrama de dispersão um ponto representa os dados de um participante específico e as linhas são os parâmetros de regressão para o grupo particular (isto é, eles resumem o relacionamento entre a libido e a libido do parceiro mostrada pelos pontos). Deve ficar claro que existe um relacionamento positivo (as linhas dos parâ-

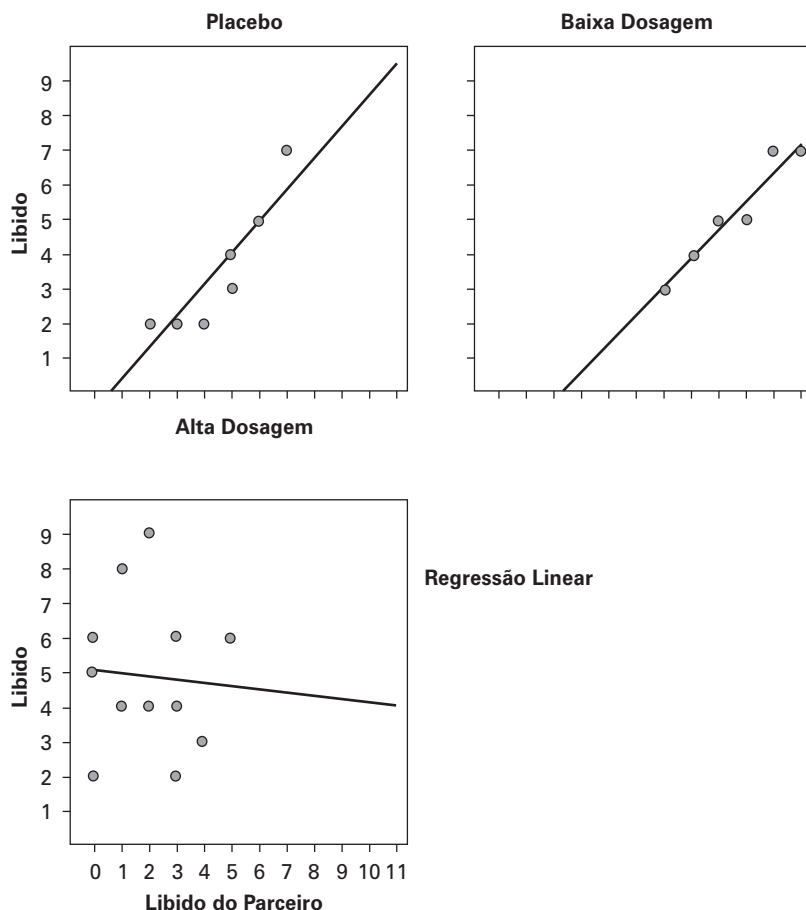


Figura 9.5 Diagrama de dispersão e linhas de regressão da libido do participante *versus* a libido do parceiro para cada uma das condições experimentais.

metros de regressão sobem da esquerda para a direita) entre a libido do parceiro e a libido do participante nas condições dos grupos-placebo e baixa dosagem. Entretanto, na condição de alta dosagem parece não haver relação entre a libido dos participantes e a dos parceiros (os pontos estão aleatoriamente espalhados e, na verdade, a linha do parâmetro de regressão desce da esquerda para a direita indicando uma relação levemente negativa. Essa observação nos fornece um indício para duvidar que exista uma homogeneidade nos parâmetros de dispersão (porque o relacionamento entre a libido do participante e do seu parceiro não é consistente entre os três grupos-experimentais).

9.6.2 Testando a homogeneidade dos parâmetros da regressão no SPSS ③

Para testar a hipótese da homogeneidade dos parâmetros da regressão precisamos executar novamente a ANCOVA, mas dessa vez utilizando um modelo customizado. Acesse a caixa de diálogo principal como anteriormente e coloque as variáveis nos mesmos quadros que antes (a janela final deve ficar como a da Figura 9.1). Para customizar o modelo precisamos acessar a caixa de diálogo **model** (Figura 9.6) clicando em **Model...** (modelo).

Para customizar o modelo, clique no círculo rotulado de **Custom** (Customizar) a fim de ativar a caixa de diálogo (Figura 9.6). As variáveis especificadas na caixa de diálogo principal estão listadas no lado esquerdo e são seguidas por uma letra indicando o seu tipo (F = fator fixo, C = covariável). Para testar a homogeneidade dos parâmetros da regressão, precisamos especificar um modelo que inclui a interação entre a covariável e a variável dependente. Normalmente, a ANCOVA inclui somente o efeito principal da dosagem e parceiro e não inclui o termo de interação. Para testar o termo de interação é importante incluir os efeitos principais de dosagem e parceiros a fim de que o termo de interação seja testado controlando os efeitos principais.

Portanto, comece selecionando **dose** e **partner** (parceiro)(você pode selecionar ambos ao mesmo tempo). Onde diz **Build Term(s)** (Construir Termo(s)) existe um menu tipo lista de escolha. Clique em **Interaction** para acessar esse menu e, depois, clique em **Main effects** (Efeitos principais). Clique em **Model** a fim de mover os efeitos principais de **dose** e **partner** para o quadro denominado **Model** (Modelo). Depois precisamos especificar o termo de interação. Para tanto, selecione **dose** e **partner** ao mesmo tempo e clique em **Interaction** para acessar o menu tipo lista de escolha novamente, mas dessa vez

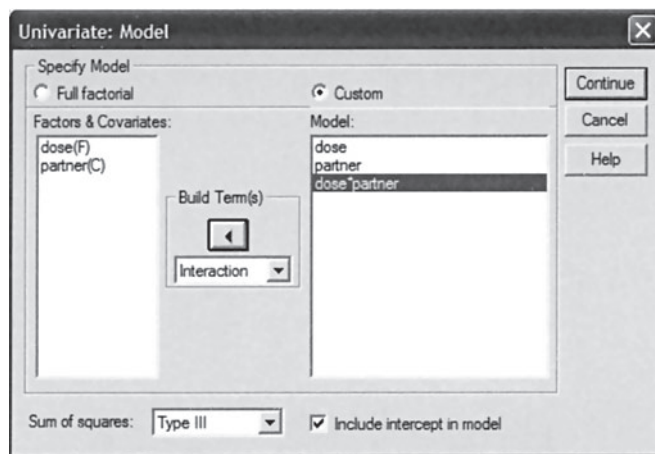



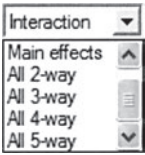


Figura 9.6 Caixa de diálogo *Model* (Modelo) do MLG univariado.

selecione **Interaction** (Interação). Clique em  para mover a interação de **dose** e **partner** para a caixa rotulada de **Model** (Modelo). A caixa de diálogo final deve ficar igual a da Figura 9.6. Você pode descobrir mais sobre como especificar efeitos na Seção 10.3.2. Tendo especificado os dois efeitos principais e o

termo de interação, clique em  (continuar) para retornar à caixa de diálogo principal e, então, em  para executar a análise.



A Saída 9.9 do SPSS mostra a tabela de resumo principal para a ANCOVA incluindo o

Saída 9.9 do SPSS

Tests of Between-Subjects Effects (Testes de efeitos entre participantes)

Dependent Variable: Libido (Variável Dependente: Libido)

Source (Fonte)	Type III Sum of Squares (Soma dos quadrados do tipo III)	df (gl)	Mean Square (Média ao quadrado)	F	Sig.
Corrected Model (Modelo Corrigido)	52.000 ^a	5	10.400	4.233	0.007
Intercept (Intercepto)	1.706	1	1.706	0.694	0.413
DOSE (Dose)	40.068	2	20.034	8.154	0.002
Partner (Parceiro)	22.430	1	22.430	9.129	0.006
Dose * Partner (Dose * Parceiro)	17.250	2	8.625	3.510	0.046
Error (Erro)	58.967	24	2.457		
Total (Total)	683.000	30			
Corrected Total (Total corrigido)	110.967	29			

a. R quadrado = 0,469 (R quadrado ajustado = 0,358).

Dica da Samanta Ferrinho



- A Análise de Covariância (ANCOVA) compara várias médias, mas com controle do efeito de outras variáveis (uma ou mais), denominadas covariáveis; por exemplo, se você tem várias condições experimentais e quer controlar a idade dos participantes.
- Na tabela denominada *Tests of between-subjects effects* (Testes entre os efeitos dos participantes), olhe para a coluna denominada *Sig.* tanto para a covariável quanto para a variável independente (a variável de agrupamento); se o valor for menor do que 0,05, isso significa que a covariável tem relação significativa com a variável de saída (a variável dependente); para a variável de agrupamento, isso significa que as médias dos grupos são significativamente diferentes após o controle para o efeito que a covariável tem na saída.
- Como ocorreu com a ANOVA, se você gerou hipóteses específicas antes do experimento, utilize as comparações planejadas (*planned comparisons*); se você não tem hipóteses específicas, utilize os testes *post hoc*. Embora o SPSS deixe você especificar certos contrastes padrão, outras comparações planejadas deverão ser feitas com o recurso do procedimento da regressão no SPSS.
- Para contrastes e testes *post hoc*, consulte a coluna denominada *Sig.* para descobrir se suas comparações são significativas (isso irá ocorrer se o valor da significância for menor do que 0,05).
- Teste as mesmas hipóteses que foram feitas para a ANOVA, mas, além disso, teste a hipótese da *homogeneidade dos parâmetros da regressão*. Isso deve ser feito customizando o modelo da ANCOVA no SPSS.

termo de interação. Os efeitos da dosagem do Viagra e a libido do parceiro continuam significativos, mas estamos interessados mesmo é no termo de interação. Assim, olhe para o valor de significância da covariável da saída da interação (**dose*partner**); se esse efeito é significativo, a hipótese da homogeneidade dos parâmetros da regressão foi violada. O efeito aqui é significativo ($p < 0,05$); portanto, a hipótese não é sustentável. Embora esse achado não seja surpreendente devido ao padrão dos relacionamentos mostrados na Figura 9.5, ele gera certa preocupação sobre a análise principal, especialmente à luz dos achados contraditórios das comparações múltiplas. Esse exemplo ilustra porque é importante testar as hipóteses e não simplesmente aceitar os resultados de uma análise.

9.7 CALCULANDO O TAMANHO DE EFEITO ②

Vimos no Quadro 9.1 que podemos fazer o SPSS produzir o eta quadrado (η^2), o que é apenas o r^2 calculado do efeito entre grupos, SS_M , dividido pela variância total dos dados, SS_T . Essa medida do tamanho de efeito é levemente tendenciosa, por isso recomendo não usá-la. Como ocorre com a ANOVA, é recomendável utilizar o ômega quadrado (ω^2), uma versão menos tendenciosa do eta quadrado (veja a Seção 8.5). Entretanto, como vimos na Seção 8.5, ele somente pode ser calculado quando tivermos um mesmo número de participantes em cada grupo (e não temos isto aqui!). Assim, estamos um pouco confusos!

Mas nem tudo está perdido porque, como eu já disse muitas vezes, o tamanho de efeito global não é tão interessante quanto o tamanho de efeito para comparações mais específicas. Essas são fáceis de calcular porque selecionamos parâmetros de regressão (veja a Saída do SPSS 9.3) e, assim, temos a estatística t para a covariável e comparações entre os grupos de baixa e alta dosagem e dos grupos-placebo e de alta dosagem. Essas estatísticas t têm $N - 2$ graus de liberdade (veja o Capítulo 5), onde N é o tamanho total da amostra (nesse caso,

30). Podemos usar a mesma equação da Seção 7.5.5.³

$$r_{\text{contraste}} = \sqrt{\frac{t^2}{t^2 + gl}}$$

Portanto, obtemos (ts da Saída 9.3 do SPSS):

$$\begin{aligned} r_{\text{contraste}} &= \sqrt{\frac{2,47^2}{2,47^2 + 28}} \\ &= \sqrt{\frac{6,10}{34,10}} \\ &= 0,42 \end{aligned}$$

$$\begin{aligned} r_{\text{Alta Dosagem vs. Placebo}} &= \sqrt{\frac{-3,095^2}{-3,095^2 + 28}} \\ &= \sqrt{\frac{9,58}{37,57}} \\ &= 0,50 \end{aligned}$$

$$\begin{aligned} r_{\text{Alta Dosagem vs. Baixa}} &= \sqrt{\frac{-2,051^2}{-2,051^2 + 28}} \\ &= \sqrt{\frac{4,21}{32,21}} \\ &= 0,36 \end{aligned}$$

Todos esses efeitos apresentam tamanhos de efeito de médio a grande (eles estão entre 0,3 e 0,5). Portanto, além de serem estatisticamente significativos, esses efeitos são descobertas importantes.

9.8 RELATANDO RESULTADOS ②

Relatar a ANCOVA é o mesmo que relatar uma ANOVA, exceto que precisamos relatar o efeito da covariável também. Para a covariável e o efeito experimental, fornecemos detalhes sobre a razão F e os graus de liberdade dos quais

³ A rigor, devemos usar um procedimento um pouco mais elaborado quando os grupos não são iguais. Mas isto está fora do âmbito do livro. Rosnow, Rosenthal e Rubin (2000) fornecem um relato muito claro dessa situação.

ele foi calculado. Em ambos os casos, a razão F foi derivada pela divisão da média dos quadrados do efeito pela média dos quadrados do resíduo. Portanto, os graus de liberdade usados para avaliar a razão F são os graus de liberdade do efeito do modelo ($gl_M = 1$ para a covariável e 2 para o efeito experimental) e os graus de liberdade para os resíduos do modelo ($gl_R = 26$ para a covariável e para o efeito experimental) – veja a Saída 9.2 do SPSS. Portanto, a maneira correta de relatar a descoberta principal seria:

- A covariável, libido do parceiro, está significativamente relacionada com a libido do participante, $F(1, 26) = 6,11, p < 0,05, r = 0,42$. Também houve um significativo efeito do Viagra nos níveis da libido após o controle para o efeito da libido do parceiro, $F(2, 26) = 4,83, p < 0,05$.

Podemos, também, relatar alguns contrastes:

- Contrastes planejados revelaram que tomar uma dosagem alta de Viagra aumentou significativamente a libido em comparação tanto ao grupo-placebo, $t(26) = -3,10, p < 0,05, r = 0,50$ quanto ao grupo de baixa dosagem, $t(26) = -2,05, p < 0,05, r = 0,36$.

9.9 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

Este capítulo mostrou como o modelo linear generalizado (MLG), descrito no Capítulo 8, pode ser estendido para incluir variáveis adicionais. As vantagens de fazer isso são que podemos manter o controle sobre os fatores em vez de manipular experimentalmente o que poderia influenciar nossa medida de saída. Isso nos dá um controle experimental mais restrito e pode, também, ajudar a explicar alguma parte da nossa variância do erro, fornecendo uma medida mais pura da manipulação experimental. Não vimos muita teoria sobre a ANCOVA, apenas aprendemos conceitualmente como o modelo de regressão pode ser expandido para incluir essas variáveis adicionais (*covariáveis*). Em vez disso, fomos direto ao exemplo, que foi

observar o efeito do Viagra na libido (como no Capítulo 8), mas incluindo a libido do parceiro como uma covariável. Expliquei como fazer a análise no SPSS e interpretar os resultados, mas também mostrei como a mesma saída poderia ser obtida executando a análise como uma regressão. Isso foi feito para mostrar que tanto a ANOVA quanto a ANCOVA são apenas formas de regressão! De qualquer modo, acabamos observando uma hipótese adicional que deve ser considerada quando realizando a ANCOVA: a hipótese da homogeneidade dos parâmetros da regressão. Isso simplesmente significa que o relacionamento entre a covariável e a variável da saída deve ser o mesmo em todos os grupos-experimentais. Também vimos como testar essa hipótese no SPSS. Nos próximos capítulos iremos observar situações em que teremos mais do que uma manipulação experimental.

9.10 TERMOS-CHAVE QUE DESCOBRIMOS

- Média ajustadas
- Análise da covariância (ANCOVA)
- Homogeneidade dos parâmetros da regressão
- Controlar
- Correção de Sidak
- Covariáveis

9.11 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Ser perseguido é uma experiência muito perturbadora e incômoda (para a pessoa perseguida) na qual alguém (o perseguidor) constantemente asse-

dia ou é obcecado por outra pessoa. Ela pode assumir muitas formas, desde o envio de cartas ameaçando matar seu gato se você não retribuir o amor que o perseguidor tem por você a literalmente segui-lo por toda parte numa tentativa desesperada de ver qual CD você compra num sábado (como se fosse qualquer outro além do Fugazi!). Um psicólogo, que cansou de ser persegui-

do por pessoas, decidiu tentar duas terapias diferentes em diferentes grupos de perseguidores (25 perseguidores em cada grupo – essa variável é chamada de **Group** (Grupo)). Ao primeiro grupo de perseguidores ele deu o que chamou de terapia cruel-para-ser-gentil. Essa terapia era baseada em punição para comportamentos de perseguição; resumindo, cada vez que o perseguidor o seguia ou mandava uma carta o psicólogo o atacava com um choque elétrico até ele parar com o comportamento de perseguição. Com isso, era esperado que os perseguidores tivessem uma reação contrária a qualquer coisa que lembrasse perseguição. A segunda terapia era a psicodinâmica, uma vertente recente da terapia psicodinâmica de Freud que reconhece a farsa que é esse tipo de tratamento (ou seja, você poderia dizer que ela é baseada na teoria de Freud!). Os perseguidores eram hipnotizados e regressavam a sua infância, e o terapeuta também discutia o pênis (a não ser que fosse mulher; nesse caso, eles discutiam a falta dele), o pênis do seu pai, o pênis do seu cachorro, o pênis do gato da rua e o pênis de quem viesse à sua cabeça. No final da terapia, o psicólogo media o número de horas na semana que o perseguidor gastava perseguindo a sua presa (essa variável é chamada de **stalk2** (perseguir2)). O psicólogo acreditava que o sucesso da terapia, para começar, dependia de quão grave era o problema, assim, antes da terapia ele media o número de horas que o paciente gastava perseguindo como um indicador do quanto a pessoa perseguia (essa variável é chamada de **stalk1** (perseguir1)). Os dados estão no arquivo **Stalker.sav**. Analise o efeito da terapia no comportamento de perseguição após a terapia, controlando o total do comportamento de perseguição antes da terapia. ②

- **Tarefa 2:** Um gerente de *marketing* de uma conhecida companhia de bebidas estava

interessado nos benefícios terapêuticos de certos refrigerantes na cura de ressacas. Ele levou 15 pessoas para sair de noite e as embbedou. Quando acordaram na manhã seguinte, desidratados, ele deu para cinco delas água, para outros cinco Lucozade (uma gostosa bebida inglesa à base de glicose) e aos cinco restantes uma marca líder de cola (essa variável é chamada de **drink** (bebida)). Ele mediu a sensação de bem-estar (numa escala de 0 = eu me sinto como se estivesse morto a 10 = eu me sinto saudável) duas horas mais tarde (essa variável é chamada de **well** (bem)). O gerente queria saber qual bebida produziu o maior nível bem-estar. Entretanto, ele se deu conta de que era importante controlar o quão bêbado a pessoa tinha ficado na noite anterior, e mediu isso numa escala de 0 (= tão sóbrio quanto uma freira) a 10 (= se debaten-do como um peixe fora da água, no chão, numa poça do seu próprio vômito). Os dados estão no arquivo **HangoverCure.sav**. Realize uma ANCOVA para ver se as pessoas se sentiram melhores depois de tomar as diferentes bebidas quando controlado o quão bêbados estavam na noite anterior. ②

As respostas estão no arquivo **Answers (Chapter 9).pdf** e a tarefa 1 tem uma interpretação completa em Field e Hole (2003).

9.12 LEITURAS COMPLEMENTARES

- HOWELL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 2002. 5ª Ed. Capítulo 16.
- RUTHERFORD, A. *Introducing ANOVA and ANCOVA: A GLM Approach*. London: Sage, 2000.
- WILDT, A. R., AHTOLA, O. *Analysis of covariance*. Sage University paper series on quantitative applications in the social sciences. 07-12. Newbury Park (CA): Sage, 1978. Esse texto é de alto nível e bastante completo se você quiser saber sobre a matemática por trás da ANCOVA.

10

ANOVA FATORIAL (MLG 3)

10.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②

Nos dois capítulos anteriores, vimos situações nas quais tentamos testar diferenças entre grupos quando havia uma única variável independente (isto é, uma variável que havia sido manipulada). Entretanto, no início do Capítulo 8, afirmei que uma das vantagens da ANOVA era que poderíamos olhar para os efeitos de mais de uma variável independente (e como essas variáveis interagem). Este capítulo estende o que já sabemos sobre ANOVA à observação de situações onde existem duas variáveis independentes. Já vimos no capítulo anterior que é muito fácil incorporar uma segunda variável à estrutura da ANOVA quando aquela variável é uma variável contínua (isto é, não dividida em grupos), mas agora iremos ver situações onde existe uma segunda variável independente sistematicamente manipulada designando pessoas a diferentes situações.

10.2 TEORIA DA ANOVA FATORIAL (ENTRE GRUPOS) ②

10.2.1 Delineamentos fatoriais ②

A **ANOVA fatorial** é usada quando você tem duas ou mais variáveis independentes

(essas variáveis são, às vezes, denominadas **fatores**, por isso o nome ANOVA *fatorial*). Existem vários tipos de delineamentos fatoriais:

- ***Delineamento fatorial independente:***

Quando existem diversas variáveis independentes ou previsoras e cada uma foi mensurada usando participantes diferentes (entre grupos).

- ***Delineamento fatorial relacionado:***

Quando diversas variáveis independentes ou previsoras foram mensuradas, mas os mesmos participantes foram usados em todas as condições (medidas repetidas).

- ***Delineamento misto:*** Quando diversas variáveis independentes (ou previsoras) foram mensuradas; algumas foram mensuradas com diferentes participantes e outras usaram os mesmos participantes.

Esta seção estende a ANOVA de um fator para o caso fatorial e os capítulos subsequentes aos delineamentos de medidas repetidas, delineamentos fatoriais de medidas repetidas e finalmente, delineamentos misturados.



Quadro 10.1**Nomeando ANOVAs ②**

As ANOVAs podem ser bem confusas porque existem muitas delas. Você irá, frequentemente, se deparar com afirmações como “uma ANOVA de dois fatores independente foi conduzida” ou “uma ANOVA de medidas repetidas de dois fatores foi conduzida” em periódicos. Essas afirmações podem parecer confusas, mas na verdade elas são bem fáceis. Todas as ANOVAs têm duas coisas em comum: elas envolvem uma quantidade de variáveis independentes e essas variáveis podem ser mensuradas usando os mesmos ou diferentes participantes. Se os mesmos participantes foram usados, geralmente usamos a frase *medidas repetidas*, e se diferentes participantes foram utilizados, usamos a palavra *independente*. Quando existem duas ou mais variáveis independentes, é possível que algumas variáveis usem os mesmos participantes enquanto outras usem diferentes participantes. Nesse caso, usamos a palavra *mista*. Quando nomeamos uma ANOVA estamos, simplesmente, dizendo ao leitor quantas variáveis independentes usamos e como elas foram mensuradas. Em termos gerais, podemos escrever o nome de uma ANOVA como:

- Uma (*número de variáveis independentes*) maneira como essas variáveis foram medidas pela ANOVA

Lembrando isso você pode entender o nome de qualquer ANOVA que aparecer. Veja esses exemplos e tente descobrir quantas variáveis foram usadas e como elas foram mensuradas:

- ANOVA de um fator independente.
- ANOVA de dois fatores de medidas repetidas.
- ANOVA de dois fatores mista.
- ANOVA de três fatores independentes.

As repostas deveriam ser:

- Uma variável independente mensurada usando participantes diferentes.
- Duas variáveis independentes, ambas mensuradas usando os mesmos participantes.
- Duas variáveis independentes: uma mensurada usando participantes diferentes e a outra mensurada usando os mesmos participantes.
- Três variáveis independentes, todas mensuradas usando participantes diferentes.

10.2.2 Um exemplo com duas variáveis independentes ②

Ao longo deste capítulo, usaremos um exemplo que tem duas variáveis independentes. Isso é conhecido como a ANOVA de dois fatores. Você verá as ANOVAs referidas como de um fator, de dois fatores, de três fatores e assim por diante e o número antes da palavra “fatores” nos diz quantas variáveis independentes havia: um fator tem uma variável independente, três fatores têm três (veja o Quadro 10.1). De qualquer modo, ficaremos com duas. Uma antropologista estava interessada nos efeitos do álcool na seleção de acompanhantes em uma boate. Seu fundamento era que após o consu-

mo do álcool, percepções subjetivas de atrativos físicos se tornariam menos precisas (o famoso efeito do álcool – “Não há gente feia você é que bebeu pouco”). Ela também estava interessada em saber se esse efeito era diferente para homens e mulheres. Ela escolheu 48 estudantes: 24 homens e 24 mulheres. Ela, então, levou grupos de oito participantes a uma boate e não deu álcool para eles (os participantes receberam drinques placebo de cerveja sem álcool), 2 canecas de uma cerveja forte ou 4 canecas* de

* N. de T.: Medida de capacidade utilizada em alguns países da Europa (especialmente Grã-Bretanha) e Estados Unidos. Na Grã-Bretanha, ela equivale a 568 ml e nos Estados Unidos, 473 ml. Esses valores podem variar em outros países.

Tabela 10.1 Dados para o efeito álcool (você é que bebeu pouco)

Álcool Gênero	Nada		2 Canecas ¹		4 Canecas	
	Feminino	Masculino	Feminino	Masculino	Feminino	Masculino
	65	50	70	45	55	30
	70	55	65	60	65	30
	60	80	60	85	70	30
	60	65	70	65	55	55
	60	70	65	70	55	35
	55	75	60	70	60	20
	60	75	60	80	50	45
	55	65	50	60	50	40
Total	485	535	500	535	460	285
Média	60,625	66,875	62,50	66,875	57,50	35,625
Variância	24,55	106,70	42,86	156,70	50,00	117,41

1. Uma caneca = 1 Pint

uma cerveja forte. No final da noite, ela tirou uma foto da pessoa com quem o participante estava conversando. Depois, reuniu um grupo de juízes independentes para avaliar a atratividade da pessoa em cada fotografia (numa escala de 0 de 100). Os dados estão na Tabela 10.1 e no arquivo **goggles.sav**.

10.2.3 A Soma total dos Quadrados (SS_T) ②



A ANOVA de dois fatores é conceitualmente muito similar à ANOVA de um fator. Basicamente, nós ainda encontramos o total da soma dos erros

ao quadrado (SS_T) e decompomos essa variância em uma variância que pode ser explicada pelo experimento (SS_M) e uma variância que não pode ser explicada (SS_R). Entretanto, na ANOVA de dois fatores, a variância explicada pelo modelo é feita não somente por uma manipulação experimental, mas duas. Portanto, decompomos a soma dos quadrados do modelo em uma variância explicada pela primeira variável independente (SS_A), uma variância explicada pela segunda variável independente (SS_B) e uma variância explicada pela interação dessas duas variáveis ($SS_{A \times B}$).

Basicamente, começamos da mesma forma como fizemos com a ANOVA de um fator.

Isto é, calculamos quanta variabilidade existe entre os escores quando ignoramos a condição experimental a que eles pertencem. Lembre a ANOVA de um fator (Equação 8.4) que a SS_T é calculada utilizando a seguinte equação:

$$SS_T = s^2_{\text{total}} (N - 1)$$

A variância total é simplesmente a variância de todos os escores quando ignoramos o grupo ao qual eles pertencem. Assim, se tratarmos os dados como um grande grupo, eles ficarão da seguinte forma:

65	50	70	45	55	30
70	55	65	60	65	30
60	80	60	85	70	30
60	65	70	65	55	55
60	70	65	70	55	35
55	75	60	70	60	20
60	75	60	80	50	45
55	65	50	60	50	40

Média Geral = 58,33

Se calcularmos a variância de todos esses escores, teremos 190,78 (tente isso na sua calculadora se você não confia em mim). Usamos 48 escores para gerar esse valor e, assim, N é 48. Dessa forma, a equação anterior fica:

$$\begin{aligned} SS_T &= s^2_{\text{total}} (N - 1) \\ &= 190,78(48 - 1) \\ &= 8966,66 \end{aligned}$$

Os graus de liberdade para esse SS será $N - 1$, ou 47.

10.2.4 A Soma dos Quadrados do Modelo (SS_M) ②

O próximo passo é calcular a soma dos quadrados do modelo. Como sugeri antes, essa soma dos quadrados é, então, dividida em três componentes: a variância explicada pela primeira variável independente (SS_A), a variância explicada pela segunda variável independente (SS_B) e a variância explicada pela interação dessas duas variáveis ($SS_{A \times B}$).

Antes de dividirmos a soma dos quadrados do modelo em suas partes componentes, primeiro precisamos determiná-la. Temos agora 8966,66 unidades de variância total para serem explicadas e nosso primeiro passo é determinar o quanto dessa variância é explicada pelo total das nossas manipulações experimentais (ignorando qual das duas variáveis independentes é responsável). Quando fizemos a ANOVA de um fator calculamos a soma dos quadrados do modelo verificando a diferença entre a média de cada grupo e a média total (veja a Seção 8.2.5). Podemos fazer o mesmo aqui. Efetivamente temos seis grupos-experimentais se combinarmos todos os níveis das duas variáveis independentes (três doses para os participantes masculinos e três doses para os femininos). Assim, dado que temos seis grupos de pessoas diferentes podemos aplicar a equação para a soma dos quadrados do modelo que usamos para a ANOVA de um fator (Equação 8.5):

$$SS_M = \sum n_k (\bar{x}_k - \bar{x}_{\text{geral}})^2$$

A média geral é a média de todos os escores (acima, calculamos isso como sendo 58,33) e n_k é o número de escores em cada grupo (isto é, o número de participantes em cada um dos seis grupos-experimentais; oito nesse caso). Portanto, a equação fica:

$$\begin{aligned} SS_M = & 8(60,625 - 58,33)^2 + 8(66,875 - 58,33)^2 \\ & + 8(62,500 - 58,33)^2 + 8(66,875 - 58,33)^2 \\ & + 8(57,500 - 58,33)^2 \\ & + 8(35,625 - 58,33)^2 \end{aligned}$$

$$\begin{aligned} = & 8(2,295)^2 + 8(8,545)^2 + 8(4,17)^2 + \\ & 8(8,545)^2 + 8(-0,83)^2 + 8(-22,705)^2 \\ = & 42,1362 + 584,1362 + 139,1112 \\ & + 584,1362 + 5,5112 + 4124,1362 \\ = & 5479,167 \end{aligned}$$

Os graus de liberdade para essa SS será o número de grupos usados, k , menos 1. Nós usamos seis grupos, assim, $gl = 5$.

Nesse estágio, sabemos que o modelo (nossas manipulações experimentais) podem explicar 5479,167 unidades da variância do total de 8966,66 unidades. O próximo estágio é dividir essa soma dos quadrados do modelo para ver quanta variância é explicada pelas nossas variáveis independentes separadamente.

10.2.4.1 O efeito principal do Gênero (SS_A) ②

Para calcular a variância de responsabilidade da primeira variável independente (nesse caso, gênero), precisamos agrupar os escores no conjunto de dados de acordo com o gênero a que pertencem. Assim, basicamente, ignoramos a quantidade de bebida que foi ingerida e colocamos todos os escores masculinos em um grupo e todos os escores femininos em outro grupo. Assim, os dados ficarão dessa maneira (note que o primeiro quadro contém as três colunas femininas da nossa tabela original e o segundo quadro contém as três colunas masculinas):

A ₁ Feminino			A ₂ Masculino		
65	70	55	50	45	30
70	65	65	55	60	30
60	60	70	80	85	30
60	70	55	65	65	55
60	65	55	70	70	35
55	60	60	75	70	20
60	60	50	75	80	45
55	50	50	65	60	40

Média Feminina
= 60,21

Média Masculina
= 56,46

Podemos, então, aplicar a equação para a soma dos quadrados do modelo que usamos para calcular a soma total dos quadrados:

$$SS_A = \sum n_k (\bar{x}_k - \bar{x}_{\text{geral}})^2$$

A média geral é a média de todos os escores (acima) e n_k é o número dos escores em cada grupo (isto é, o número de homens e mulheres: 24 nesse caso). Portanto, a equação fica:

$$\begin{aligned} SS_{\text{Gênero}} &= 24(60,21 - 58,33)^2 + 24(56,46 - 58,33)^2 \\ &= 24(1,88)^2 + 24(-1,87)^2 \\ &= 84,8256 + 83,9256 \\ &= 168,75 \end{aligned}$$

Os graus de liberdade para esse SS será o número de grupos usados, k , menos 1. Usamos dois grupos (masculinos e femininos), assim, $gl = 1$.

10.2.4.2 O efeito principal do Álcool (SS_B) ②

Para calcular a variância de responsabilidade da segunda variável independente (nesse caso, álcool) precisamos agrupar os escores do conjunto de dados original de acordo com a quantidade de álcool ingerida. Assim, basicamente ignoramos o gênero do participante e apenas colocamos todos os escores sem bebida em um grupo, os escores dos que tomaram 2 canecas em outro grupo e os escores dos que tomaram 4 canecas num terceiro grupo. Assim, os dados ficarão com a seguinte aparência:

B₁: Sem bebida	B₂: 2 canecas	B₃: 4 canecas
65	70	55
50	45	30
70	65	65
55	60	30
60	85	30
60	70	55
65	65	55
60	70	35
55	60	60
75	70	20
60	80	50
75	45	45
55	60	50
65	60	40
Média sem bebida = 63,75	Média após 2 canecas = 64,6875	Média após 4 canecas = 46,5625

Podemos, então, aplicar a mesma equação para a soma dos quadrados do modelo que usamos para o total da soma dos quadrados do modelo e para o efeito principal de gênero:

$$SS_B = \sum n_k (\bar{x}_k - \bar{x}_{\text{geral}})^2$$

A média geral é a média de todos os escores (58,33 como antes) e n_k é o número dos escores em cada grupo (isto é, o número dos escores em cada um dos quadros acima, nesse caso, 16). Portanto, a equação fica:

$$\begin{aligned} SS_{\text{Álcool}} &= 16(63,75 - 58,33)^2 + 16(64,6875 - 58,33)^2 \\ &\quad + 16(46,5625 - 58,33)^2 \\ &= 16(5,42)^2 + 16(6,3575)^2 \\ &\quad + 16(-11,7675)^2 \\ &= 470,0224 + 646,6849 + 2215,5849 \\ &= 3332,292 \end{aligned}$$

Os graus de liberdade para essa SS será o número de grupos usados menos 1 (veja a Seção 8.2.5 da ANOVA de um fator). Usamos três grupos e, assim, $gl = 2$.

10.2.4.3 O efeito interação ($SS_{A \times B}$) ②

O estágio final é calcular quanto da variância é explicada pela interação das duas variáveis. A maneira mais fácil de fazer isso é lembrar que a SS_M é constituída de três componentes (SS_A , SS_B e $SS_{A \times B}$); portanto, dado que conhecemos SS_A e SS_B , podemos calcular o termo de interação usando a subtração:

$$SS_{A \times B} = SS_M - SS_A - SS_B$$

Portanto, para esses dados, o valor é:

$$\begin{aligned} SS_{A \times B} &= SS_M - SS_A - SS_B \\ &= 5479,167 - 168,75 - 3332,292 \\ &= 1978,125 \end{aligned}$$

Os graus de liberdade podem ser calculados da mesma maneira, mas são também produto dos graus de liberdade para os efeitos principais (qualquer método funciona):

$$\begin{aligned} gl_{A \times B} &= gl_M - gl_A - gl_B & gl_{A \times B} &= gl_A \times gl_B \\ &= 5 - 1 - 2 & &= 1 \times 2 \\ &= 2 & &= 2 \end{aligned}$$

10.2.5 A Soma dos Quadrados dos Resíduos (SS_R) ②

A soma dos quadrados dos resíduos é calculada da mesma maneira que para a ANOVA

de um fator (veja a Seção 8.2.6) e, novamente, representa diferenças individuais no desempenho ou a variância que não pode ser explicada pelos fatores que foram sistematicamente manipulados. Vimos na ANOVA de um fator que o valor é obtido calculando o erro ao quadrado entre cada ponto de dados e a média correspondente do grupo. Uma forma alternativa para expressar isso foi como (veja a Seção 8.7):

$$SS_R = s_{\text{grupo1}}^2(n_1 - 1) + s_{\text{grupo2}}^2(n_2 - 1) + s_{\text{grupo3}}^2(n_3 - 1) + \dots + s_{\text{grupom}}^2(n_m - 1)$$

Assim, usamos as variâncias individuais dentro de cada grupo e as multiplicamos pelo número de pessoas dentro do grupo (n_k) menos um. Temos as variâncias individuais de cada grupo na nossa tabela de dados original (Tabela 10.1) e há oito pessoas em cada grupo (portanto, $n_k = 8$); assim, a equação fica:

$$\begin{aligned} SS_R &= s_{\text{grupo1}}^2(n_1 - 1) + s_{\text{grupo2}}^2(n_2 - 1) \\ &\quad + s_{\text{grupo3}}^2(n_3 - 1) + s_{\text{grupo4}}^2(n_4 - 1) \\ &\quad + s_{\text{grupo5}}^2(n_5 - 1) + s_{\text{grupo6}}^2(n_6 - 1) \\ &= (24,55)(8 - 1) + (106,7)(8 - 1) \\ &\quad + (42,86)(8 - 1) + (156,7)(8 - 1) \\ &\quad + (50)(8 - 1) + (117,41)(8 - 1) \\ &= (24,55 \times 7) + (106,7 \times 7) \\ &\quad + (42,86 \times 7) + (156,7 \times 7) \\ &\quad + (50 \times 7) + (117,41 \times 7) \\ &= 171,85 + 746,9 + 300 + 1096,9 \\ &\quad + 350 + 821,87 \\ &= 3487,52 \end{aligned}$$

Os graus de liberdade para cada grupo serão um a menos do que o número de escores por grupo (isto é, 7). Portanto, se acrescentarmos os graus de liberdade para cada grupo, temos um total de $6 \times 7 = 42$.

10.2.6 As razões F ②

Cada efeito na ANOVA de dois fatores (os dois efeitos principais e a interação) tem sua própria razão F. Para calcular essas estatísticas, primeiro temos que calcular os quadrados médios para cada efeito tomando a soma dos quadrados e dividindo-a pelos graus de liber-

dade respectivos (volte à Seção 8.2.7). Também precisamos dos quadrados médios para o termo residual. Assim, para esse exemplo teremos quatro quadrados médios calculados da seguinte forma::

$$\begin{aligned} MS_A &= \frac{SS_A}{gl_A} = \frac{168,75}{1} = 168,75 \\ MS_B &= \frac{SS_B}{gl_B} = \frac{3332,292}{2} = 1666,146 \\ MS_{A \times B} &= \frac{SS_{A \times B}}{gl_{A \times B}} = \frac{1978,125}{2} = 989,062 \\ MS_R &= \frac{SS_R}{gl_R} = \frac{3487,52}{42} = 83,036 \end{aligned}$$

A razão F para as duas variáveis independentes e sua interação são calculadas dividindo seus quadrados médios pelos quadrados residuais médios. Novamente, se você pensar na ANOVA de um fator, esse processo é exatamente o mesmo!

$$\begin{aligned} F_A &= \frac{MS_A}{MS_R} = \frac{168,75}{83,036} = 2,032 \\ F_B &= \frac{MS_B}{MS_R} = \frac{1666,146}{83,036} = 20,065 \\ F_{A \times B} &= \frac{MS_{A \times B}}{MS_R} = \frac{989,062}{83,036} = 11,911 \end{aligned}$$



Cada uma dessas razões F pode ser comparada aos valores críticos (com base nos seus graus de liberdade que podem ser diferentes para cada efeito)

para nos dizer se há uma probabilidade desses efeitos refletirem dados que surgiram por acaso ou refletirem um efeito das nossas manipulações experimentais (esses valores críticos podem ser encontrados no Apêndice A.3). Se uma razão F observada exceder o valor crítico correspondente, ela é significativa. O SPSS irá calcular essas razões F e a significância exata para cada uma, mas espero ter mostrado para você nesta seção que a ANOVA de dois fatores é basicamente a mesma que a ANOVA de um fator exceto que a soma dos quadrados do

modelo é dividida em três partes: o efeito de cada uma das variáveis independentes e o efeito de como essas variáveis interagem.

10.3 ANOVA FATORIAL UTILIZANDO O SPSS ②



10.3.1 Entrando com dados e utilizando a janela principal de diálogos ②

Para entrar com esses dados no editor de dados do SPSS, lembre a regra básica: **os níveis de uma variável entre grupos são colocados em uma única coluna**. Para aplicar essa regra a esses dados, precisamos criar duas variáveis código no editor de dados. Essas colunas irão representar o gênero e o consumo de álcool. Assim, crie uma variável denominada **gender** (gênero) no editor de dados e ative a caixa de diálogo **labels** (rótulos). Você deve definir o valor dos rótulos para representar os dois gêneros. Nós temos muita experiência com valores codificados, assim, você deve estar muito feliz designando códigos numéricos a diferentes grupos. Recomendo usar os códigos masculino = 0 e feminino = 1. Uma vez feito isso, você pode entrar com o valor de 0 ou 1 nessa coluna indicando o grupo a que a pessoa pertence. Crie uma segunda variável denominada **alcohol** (álcool) e designe códigos de grupo usando a caixa de diálogo **labels** (rótulos). Sugiro que você codifique essa variável com os seguintes três valores: placebo (sem álcool) = 1, 2 canecas = 2 e 4 canecas = 3. Você agora pode entrar com os valores 1, 2 ou 3 nessa coluna para representar a quantidade de álcool consumida pelo participante. Lembre que se você ativar opção valores dos rótulos, verá texto no editor de dados em vez dos códigos numéricos. Agora, a maneira como essa codificação funciona é a seguinte:

Uma vez que você criou as duas variáveis codificadas, você pode criar uma terceira variável para colocar os valores da variável dependente. Chame essa variável **attract** (atração) e use a opção **labels** (rótulos) para dar a ela o nome completo de “*attractiveness of date*” (atratividade do companheiro). Nesse exemplo existem duas variáveis independen-

Gênero	Álcool	O participante é
0	1	Homem que não consumiu álcool
0	2	Homem que consumiu 2 canecas
0	3	Homem que consumiu 4 canecas
1	1	Mulher que não consumiu álcool
1	2	Mulher que consumiu 2 canecas
1	3	Mulher que consumiu 4 canecas

tes e diferentes participantes foram usados em cada condição: portanto, podemos usar o procedimento geral da ANOVA fatorial do SPSS. Esse procedimento foi projetado para analisar delineamentos fatoriais entre grupos.

Para acessar a caixa de diálogo principal para a ANOVA fatorial geral, use o caminho do arquivo **Analyze⇒General Linear Model⇒Univariate...** (Analisar⇒Modelo Linear Geral⇒Univariado...). A caixa de diálogo resultante é mostrada na Figura 10.1. Primeiro selecione a variável dependente **attract** (atração) da lista de variáveis no lado esquerdo da caixa de diálogo e a transfira para o espaço denominado **Dependent Variable** (Variável dependente) clicando em . No espaço denominado **Fixed Factor(s)** (Fator(es) fixos), precisamos colocar qualquer variável relevante para a análise. Selecione **alcohol** (álcool) e **gender** (gênero) da lista de variáveis (essas variáveis podem ser selecionadas simultaneamente clicando numa, pressionando o mouse e arrastando o ponteiro sobre a outra variável) e transfira-as para a caixa **Fixed Factor(s)** (Fatores fixos) clicando em . Existem outros espaços que estão disponíveis para a realização de análises mais complexas como a ANOVA de fatores aleatórios e a ANCOVA fatorial. A ANOVA de fatores aleatórios está além do escopo deste livro (leitores interessados devem consultar Jackson e Brashers, 1994) e a ANCOVA fatorial simplesmente estende os princípios descritos no Capítulo 9.

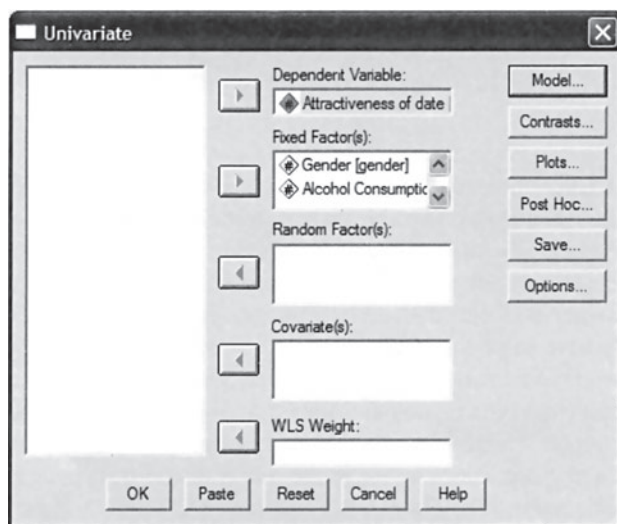


Figura 10.1 Caixa de diálogo principal para a ANOVA univariada.

10.3.2 Modelos customizados ②

Por padrão, o SPSS conduz uma análise fatorial completa (isto é, ela inclui todos os efeitos principais e interações de todas as variáveis independentes especificadas na caixa de diálogo principal). Entretanto, você pode querer customizar o modelo utilizado a fim de testar algumas coisas. Para acessar a caixa de diálogo **modelo** (modelo), clique em **Model...** na caixa de diálogo principal (veja a Figura 10.2). Você no-

tará que, por definição, o modelo fatorial completo está selecionado. Mesmo com ele selecionado, existe a opção na parte de baixo para mudar o tipo da soma dos quadrados utilizado na análise. Embora tenhamos aprendido sobre as somas dos quadrados e o que elas representam, não discuti sobre as diferentes maneiras de calculá-las. Não é necessário entender os cálculos das diferentes formas das somas

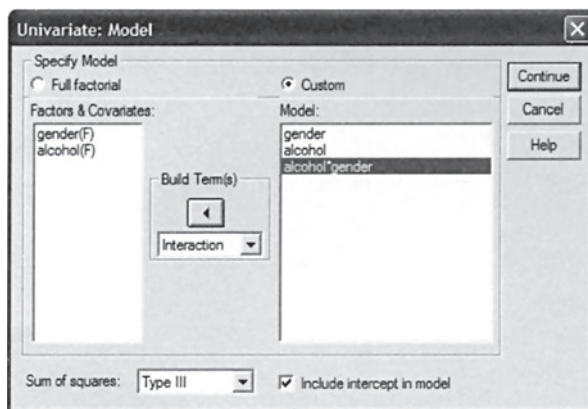
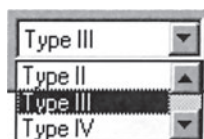
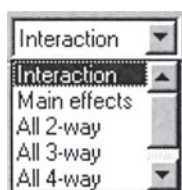


Figura 10.2 Modelos customizados na ANOVA.

dos quadrados, mas é importante que você conheça os usos de alguns tipos diferentes. Por padrão, o SPSS usa as somas dos quadrados do Tipo III que têm a vantagem de permanecerem constantes independentemente das frequências das células. Assim, elas podem ser usadas tanto com delineamentos balanceados ou não balanceados (isto é, diferentes números de participantes nos grupos), por isso elas são a opção padrão. As somas dos quadrados do Tipo IV são como as do Tipo III, com a exceção de poderem ser usadas com dados nos quais valores estejam faltando (*missing data*). Desse modo, se você tiver dados faltando no seu delineamento, deve utilizar a somas dos quadrados do Tipo IV.



Para customizar um modelo, clique no círculo denominado **Custom** (Customizar) para ativar a caixa de diálogo. As variáveis especificadas na caixa de diálogo principal serão

listadas no lado esquerdo e seguidas por uma letra em colchetes indicando o tipo da variável (F = fator fixo, R = fator aleatório, C = covariável). Você pode selecionar uma ou diversas variáveis dessa lista e transferi-las para o quadro denominado **Model** (Modelo) tanto como efeitos principais, como interações. Por definição, o SPSS transfere variáveis como termos de interação, mas existem muitas opções que nos permitem entrar com os efeitos principais ou as interações de um, dois ou três fatores. Essas opções evitam alguns problemas de ter que selecionar diversas combinações de variáveis (porque, por exemplo, você pode selecionar três variáveis, transferi-las como interações de dois fatores e elas irão criar combinações de três variáveis para você). Embora a seleção do modelo tenha usos importantes (veja a Seção 9.6), é provável que você queira executar a análise fatorial completa na maioria das vezes.

10.3.3 Ilustrando interações ②

Uma vez que as variáveis relevantes tenham sido selecionadas, você deve clicar em **Plots...** (traçar) para acessar uma caixa de diálogo semelhante à da Figura 10.3. A caixa de diálogo

plots permite que você selecione gráficos de linhas para os seus dados e eles serão úteis para a interpretação dos efeitos de interação. Temos somente duas variáveis independentes, assim, existe somente um gráfico que vale a pena observar (o que mostra os níveis de independência de uma variável contra a outra.). Selecione **alcohol** (álcool) da lista de variáveis do lado esquerdo da caixa de diálogo e transfira-a para o espaço denominado **Horizontal Axis** (Eixo Horizontal) clicando em **Add**. No espaço denominado **Separate Lines** (Linhas Separadas), precisamos colocar a variável independente restante: **gender** (gênero). Na verdade, independentemente da maneira como as variáveis são apresentadas, você deve usar o seu discernimento para produzir o gráfico mais adequado ao interesse. Quando você mudar as duas variáveis independentes para a caixa apropriada, clique em **Add** (adicionar) e esse gráfico será acrescentado à lista na parte de baixo da caixa de diálogo. Deveria estar claro que você pode traçar uma grande variedade de gráficos e, se você tiver uma terceira variável independente, terá uma opção de traçar diferentes gráficos para cada nível dessa terceira variável. O diagrama selecionado nos auxiliará na interpretação de qualquer interação entre gênero e consumo de álcool. Quando você tiver terminado de especificar os gráficos, clique em **Continue** (continuar) para retornar à caixa de diálogo principal.

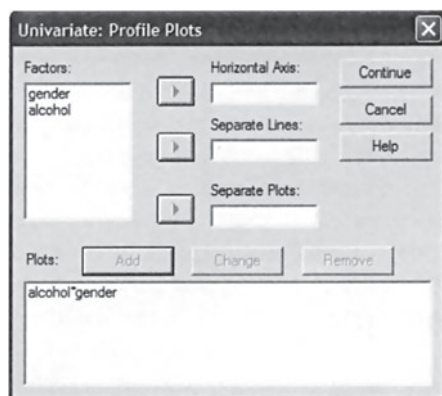


Figura 10.3 Caixa de diálogo para traçar gráficos.

10.3.4 Contrastes ②

Vimos no Capítulo 8 que é útil complementar a ANOVA com contrastes que dividam os efeitos principais e nos digam onde estão as diferenças entre grupos. Para a ANOVA de um fator independente, o SPSS tem um procedimento para a entrada de códigos que definam os contrastes que queremos fazer. Entretanto, para a ANOVA de dois fatores não existe essa facilidade e, em vez disso, estamos restritos a fazer um de muitos contrastes padrão. Esses contrastes padrão estão descritos na Tabela 8.6. Para ser honesto, esses contrastes irão fornecer o que você quer em muitas situações diferentes, mas se eles não fornecerem, e você quiser definir seus próprios contrastes, isso deverá ser feito utilizando a sintaxe. Determinar contrastes com a sintaxe não será, provavelmente, muito interessante para a maioria das pessoas, portanto não os incluí aqui; entretanto, porque sempre me perguntam muito sobre isso, e eu sou um completo masoquista, preparei um guia bem detalhado de como fazer isso no arquivo **ContrastsUsingSyntax.pdf** que está no *site* www.artmed.com.br. Assim, se você quiser saber mais, dê uma olhada nesse material.

Voltando ao assunto, podemos utilizar os contrastes padrão nesse exemplo. O efeito do gênero tem somente dois níveis e, assim, não precisamos de contrastes para o efeito principal. O efeito do álcool tem três níveis: sem álcool, 2 canecas e 4 canecas. Poderíamos, novamente, selecionar um contraste simples para essa variável e usar a primeira categoria como uma referência. Isso compararia o grupo de 2 canecas ao grupo sem álcool e, então, compararia a categoria de 4 canecas ao grupo sem álcool. Desse modo, os grupos com álcool seriam comparados ao grupo sem álcool. Podemos, também, selecionar um contraste *repetido*. Isso compararia o grupo de 2 canecas ao grupo sem álcool e, então, o grupo de 4 canecas ao grupo de 2 canecas (assim, ele se move entre os grupos comparando cada grupo ao anterior). Isso pode ser útil. Poderíamos, também, fazer um contraste Helmert, o qual compara cada categoria contra todas as categorias subsequentes, nesse caso, compara-

ria o grupo sem álcool às categorias restantes (isto é, todos os grupos que beberam álcool) e, então, moveria a categoria de 2 canecas e a compararia à categoria de 4 canecas. Qualquer uma dessas comparações seria boa, mas elas somente nos dão contrastes para os efeitos principais. Na realidade, na maior parte do tempo, realmente queremos contrastes para o nosso termo de interação e eles podem ser obtidos somente por meio da sintaxe (bem, parece que você terá que consultar o arquivo sugerido no final das contas!)

Para conseguir contrastes para o efeito principal do álcool, clique em **Contrasts...** (contrastes) na caixa de diálogo principal. Usamos a caixa de diálogo **contrasts** (contrastes) antes, na Seção 9.3.3, assim, volte àquela seção para auxiliá-lo a selecionar um contraste Helmert para a variável álcool. Uma vez que os contrastes tenham sido selecionados (Figura 10.4), clique em **Continue** (continuar) para retornar à caixa de diálogo principal.

10.3.5 Testes *post hoc* ②

Os testes *post hoc* são obtidos clicando em **Post Hoc...** na caixa de diálogo principal para acessar a caixa de diálogo testes *post hoc* (Figura 10.5). A variável **gender** (gênero) tem somente dois níveis, assim, não precisamos selecionar os testes *post hoc* para aquela variável (porque qualquer efeito significativo pode somente refletir diferenças entre homens e mulheres). Entretanto, havia três níveis da variável álcool (sem álcool, 2 canecas e 4 ca-

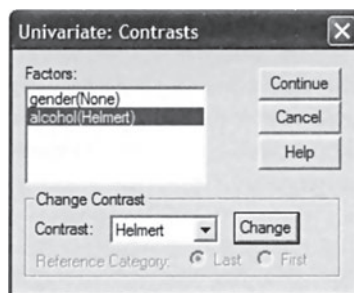


Figura 10.4 Caixa de diálogo para a obter contrastes padronizados.

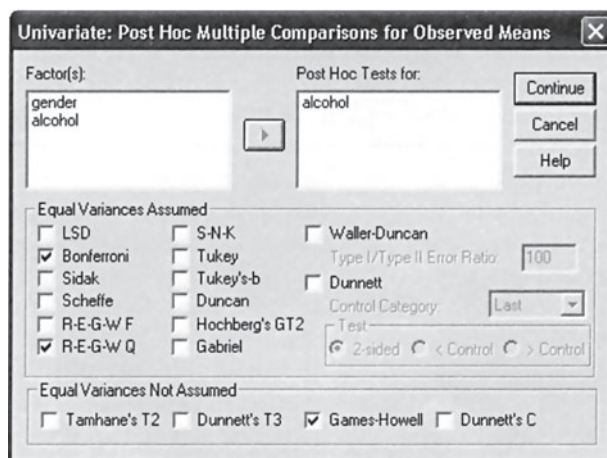


Figura 10.5 Caixa de diálogo para os testes *post hoc*.

necas): portanto, é necessário conduzir os testes *post hoc*. Primeiro você deve selecionar a variável **alcohol** (álcool) da caixa denominada **Factors** (Fatores) transferi-la para a caixa denominada **Post hoc Tests** (Testes post hoc). Minhas recomendações sobre quais testes *post hoc* usar estão na Seção 8.2.11 (não quero me repetir). É suficiente dizer que você deve se-

leccionar aqueles que foram selecionados na Figura 10.5! Clique em **Continue** para retornar à caixa de diálogo principal.

10.3.6 Opções ②

Clique em **Options...** (Opções) para ativar a caixa de diálogo semelhante à da Figura 10.6.

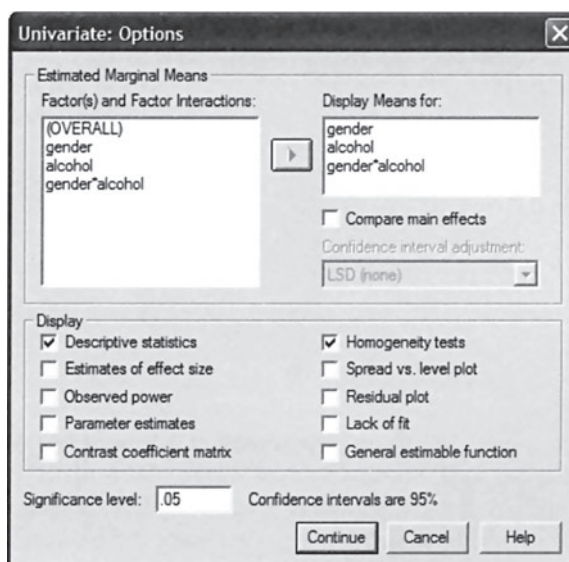


Figura 10.6 Caixa de diálogo *Options* (Opções).

As opções para a ANOVA fatorial são bem simples. Inicialmente, você pode solicitar algumas estatísticas descritivas que serão apresentadas em uma tabela contendo as médias, desvios padrão, erros padrão, amplitudes e intervalos de confiança para as médias de cada grupo. Essa é uma opção útil porque ela ajuda a interpretar os resultados finais. Uma opção vital a ser selecionada é o teste de homogeneidade das variâncias. Assim como no teste *t*, existe a hipótese de que as variâncias dos grupos são iguais e, selecionando essa opção, testamos se a hipótese é satisfeita. O SPSS usa o teste de Levene para testar a hipótese que as variâncias dos grupos são iguais. Uma vez selecionadas essas opções, clique em **Continue** para retornar à caixa de diálogo principal e clique em **OK** para executar a análise.

10.4 SAÍDAS DA ANOVA FATORIAL ②

10.4.1 Saídas para as análises preliminares ②

A Saída do SPSS 10.1 mostra a saída inicial da ANOVA fatorial. Essa tabela de estatísticas descritivas foi produzida porque marcamos a opção **Descriptive statistics** (Estatísticas descritivas) na caixa de diálogo **options** (opções) (veja a Figura 10.6), e ela mostra as médias, desvios padrão e o número de participantes em todas as condições do experi-

mento. Assim, por exemplo, podemos ver que na condição placebo os homens cantaram uma mulher que foi avaliada em 67% na escala de atração enquanto as mulheres selecionaram homens avaliados em 61% nessa mesma escala. Essas médias serão úteis na interpretação de qualquer efeito que surja na análise.

10.4.2 O teste de Levene ②

A Saída do SPSS 10.2 mostra os resultados do teste de Levene. Abordamos o teste de Levene na Seção 3.6 e inúmeras vezes depois! Em resumo, o teste de Levene é usado para avaliar a viabilidade da hipótese da igualdade das variâncias (homogeneidade da variância). O teste de Levene verifica se há alguma diferença significativa entre as variâncias do

Saída do SPSS 10.2

Levene's Test of Equality of Error Variiances^a (Teste de Levene para a Igualdade das Variâncias dos Erros)

	<i>F</i>	<i>df1</i> (gl1)	<i>df2</i> (gl2)	<i>Sig.</i>
<i>Attractiveness of Date</i> (Atratividade do Companheiro)	1.527	5	42	0.202

Testa a hipótese nula de que a variância dos erros da variável dependente é igual em todos os grupos.

a Delineamento: Intercepto + álcool + Gênero + Álcool.

* Gênero

Saída do SPSS 10.1

Descriptive Statistics (Estatísticas Descritivas)

<i>Alcohol Consumption</i> (Consumo de Álcool)			<i>Mean</i> (Média)	<i>Std. Deviation</i> (Desvio padrão)	<i>N</i>
<i>Gender</i> (Gênero)					
<i>Attractiveness of Date</i> (Atratividade do Companheiro)	Placebo	<i>Male</i> (Masculino)	66.8750	10.3294	8
		<i>Female</i> (Feminino))	60.6250	4.9552	8
		<i>Total</i> (Total)	63.7500	8.4656	16
	2 Pints	<i>Male</i> (Masculino)	66.8750	12.5178	8
		<i>Female</i> (Feminino))	62.5000	6.5465	8
		<i>Total</i> (Total)	64.6875	9.9111	16
	4 Pints	<i>Male</i> (Masculino)	35.6250	10.8356	8
		<i>Female</i> (Feminino))	57.5000	7.0711	8
		<i>Total</i> (Total)	46.5625	14.3433	16
	Total	<i>Male</i> (Masculino)	56.4583	18.5026	24
		<i>Female</i> (Feminino))	60.2083	6.3381	24
		<i>Total</i> (Total)	58.3333	13.8123	48

grupo e, assim, um resultado não-significativo (como encontrado aqui) é um indicativo de que a hipótese foi satisfeita. Se o teste de Levene é significativo, algumas providências devem ser tomadas para equalizar as variâncias pela transformação dos dados (tirar a raiz quadrada de todos os valores da variável dependente pode algumas vezes atingir esse objetivo – veja o Capítulo 3 e Howell, 2002, p.347)

10.4.3 A tabela principal da ANOVA ②

A Saída do SPSS 10.3 é a parte mais importante da saída porque ela nos diz se qual-quer uma das variáveis independentes teve um efeito na variável dependente. É importante observar os valores da significância das variáveis independentes na tabela. A primeira coisa a ser notada é que existe um efeito significativo do álcool (porque o valor de significância é menor do que 0,05). A razão F é altamente significativa, indicando que a quantidade de álcool consumida afetou significativamente quem o participante tentou cantar. Isso significa que, no geral, quando ignoramos se o participante era homem ou mulher, a quantidade de álcool influenciou na seleção do seu companheiro. O efeito da variável tomada isoladamente é conhecido como *efeito principal*. A melhor maneira para se ver o que isso significa é observar um diagrama de barras da atração média em cada nível de álcool (ignore completamente o gênero). Esse gráfico

pode ser traçado usando as médias na saída do SPSS 10.1 (viu, eu disse que esses valores seriam de alguma utilidade!) e a opção **Bar...** (Barra) do menu **Graphs** (Gráficos).

A Figura 10.7 mostra claramente que, quando você ignora o gênero, a atratividade geral do parceiro selecionado é semelhante às situações em que não foi bebido álcool e quando 2 canecas foram bebidas (as médias desses grupos são aproximadamente iguais). Portanto, esse efeito principal significativo é *provável* que reflita na queda na atratividade dos parceiros selecionados quando 4 canecas foram bebidas. Essa descoberta parece indicar que a pessoa está disposta a aceitar um companheiro menos atrativo depois de 4 canecas.

A próxima parte da saída 10.3 do SPSS nos informa sobre o efeito principal do gênero. Dessa vez a razão F não é significativa ($p = 0,161$, o que é maior do que 0,05). O que esse efeito significa é que, no geral, quando ignoramos quanto álcool foi bebido, o gênero do participante não influenciou a atratividade do parceiro que o participante selecionou. Em outras palavras, outras coisas sendo iguais, homens e mulheres igualmente selecionaram acompanhantes atraentes. Construir um diagrama de barras da atratividade média dos acompanhantes para homens e mulheres (ignorando quanto álcool foi consumido) revela o significado desse efeito principal. A Figura 10.8 mostra que a atratividade média

Saída 10.3 do SPSS

Tests of Between-Subjects Effects (Testes de efeitos entre participantes)

Dependent Variable: Attractiveness of Date (Variável dependente: Atratividade do companheiro)

Source (Fonte)	Type III Sum of Squares (Soma dos quadrados do tipo III)	df (gl)	Mean Square (Média ao quadrado)	F	Sig.
Corrected Model (Modelo Corrigido)	5479.167 ^a	5	1095.833	13.197	0.000
Intercept (Intercepto)	163333.333	1	163333.333	1967.025	0.000
ALCOHOL (Álcool)	3332.292	2	1666.146	20.065	0.000
GENDER (Gênero)	168.750	1	168.750	2.032	0.161
GENDER*ALCOHOL (Gênero*Álcool)	1978.125	2	989.062	11.911	0.000
Error (Erro)	3487.500	42	83.036		
Total (Total)	172300.000	48			
Corrected Total (Total corrigido)	8966.667	47			

a. R ao quadrado = 0,611 (R ao quadrado ajustado = 0,565).

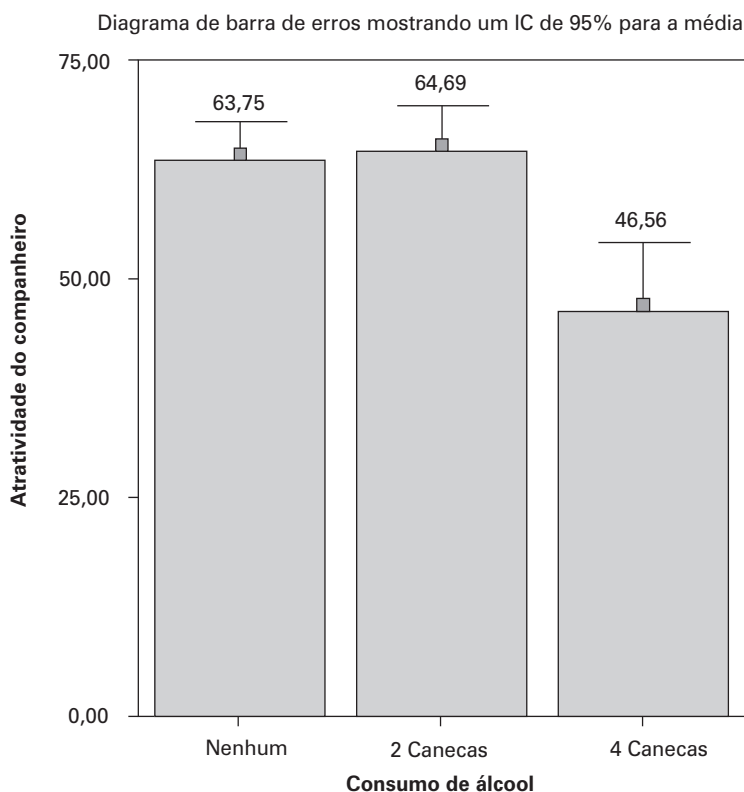


Figura 10.7 Gráfico mostrando o efeito principal do álcool.

dos companheiros dos participantes homens e mulheres foi muito parecida (as médias são diferentes por somente 4%). Portanto, esse efeito não-significativo reflete o fato de que a atratividade média foi similar. Podemos concluir disso que, *ceretis paribus**, homens e mulheres escolhem igualmente companheiros atrativos.

Finalmente, a Saída 10.3 do SPSS nos informa sobre a interação entre o efeito do gênero e o efeito do álcool. O valor de F é altamente significativo (porque o valor de p é menor do que 0,05). O que isso realmente significa é que o efeito do álcool na seleção de parceiros foi diferente para participantes homens do que para mulheres. A saída do SPSS inclui um gráfico que solicitamos (veja a Figura 10.3), que nos diz algo sobre a natureza desse efeito de interação (veja a Figura 10.9).

A Figura 10.9 mostra claramente que para as mulheres o álcool tem pouco efeito: a atratividade do parceiro selecionado é estável nas três condições (como está mostrado pela linha quase horizontal).

Entretanto, para os homens, a atratividade de suas parceiras é estável somente quando uma pequena quantidade foi ingerida, mas rapidamente declina quando bebem mais. As linhas não-paralelas normalmente indicam um efeito de interação significativa. Nesse gráfico em particular as linhas realmente se cruzam, o que indica uma grande interação entre as variáveis independentes. A interação nos diz que o álcool tem pouco efeito na seleção do acompanhante



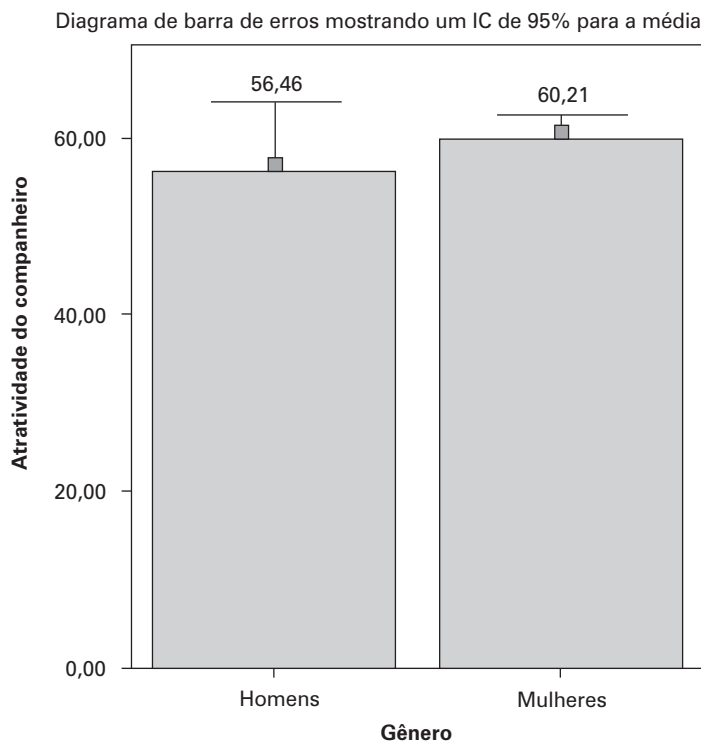


Figura 10.8 Gráfico para mostrar o efeito do gênero na seleção do companheiro.

até que 4 canecas tenham sido bebidos e que o efeito do álcool é predominante somente nos participantes masculinos. Resumindo, os resultados mostram que as mulheres mantêm seu alto padrão na seleção de seus companheiros apesar do álcool enquanto os homens, quando bebem algumas cervejas, tentam pegar qualquer coisa que ande! Um ponto interessante demonstrado por esses dados é que concluímos antes que o álcool significativamente afetava a atratividade do acompanhante na seleção (o efeito principal **álcool**); entretanto, o efeito de interação nos diz que isso somente é verdade com homens (as mulheres parecem não serem afetadas). Isso demonstra como os efeitos principais podem ser enganosos: geralmente as interações entre as variáveis são mais interessantes num delineamento fatorial.

10.4.4 Contrastes ②

A Saída 10.4 do SPSS mostra os resultados do nosso contraste Helmert no efeito do álcool. Isso nos ajuda a analisar o efeito do álcool. O topo da tabela mostra o contraste para *Level 1 vs. Later* (Nível 1 vs. Outros) o que nesse caso significa o grupo sem álcool comparado aos dois grupos com álcool. Isso testa se a média do grupo sem álcool (63,75) é diferente dos grupos dois 2 canecas e 4 canecas combinados $(64,69 + 46,56)/2 = 55,625$. Isso é uma diferença de 8,125 $(63,75 - 55,63)$ mostrada por ambas as *Estimativa de Contraste* e a *Diferença* na tabela. O importante a ser verificado é o valor de Sig. que nos diz se a diferença é significativa. Ela é porque o Sig. é 0,006, o que é menor do que 0,05. Já foi apresentado o intervalo de confiança para essa diferença e, porque ele não passa pelo zero, podemos afirmar que essa amostra é uma das 95 em 100 que produz

* N. de T.: Mantendo tudo o mais constante, isto é, semelhante.

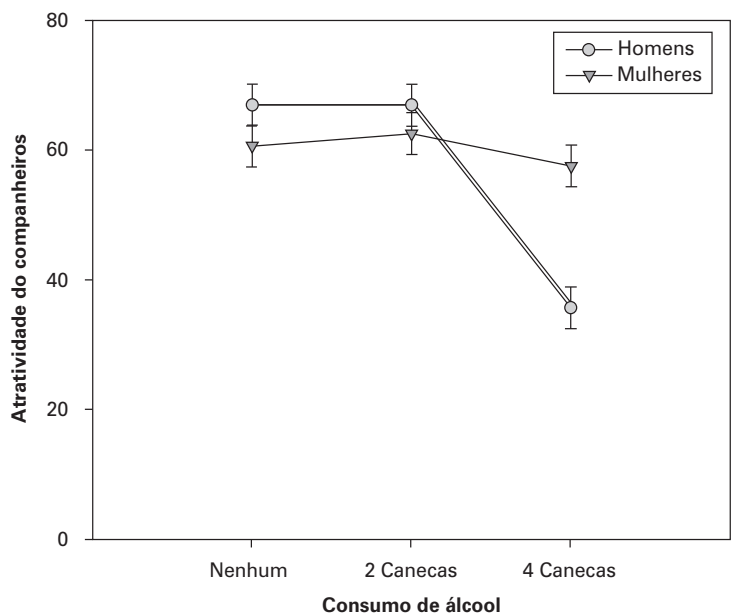


Figura 10.9 Gráfico da interação entre gênero e o consumo de álcool na seleção do companheiro.

Saída do SPSS 10.4

Contrast Results (K Matrix) (Resultados dos Contrastes – Matriz K)

Alcohol consumption (Consumo de álcool) Helmert Contrast (Contraste de Helmert)		Dependent Variable (Variável dependente)
		Attractiveness of Date (Atratividade do companheiro)
Level 1 vs. Later (Nível 1 vs. Outros)	Contrast Estimate (Estimativa do Contraste)	8.125
	Hypothesized Value (Valor Hipotético)	0
	Diference (Estimate – Hypothesized) (Diferença (Estimativa – Hipotético))	8.125
	Std. Error (Erro Padrão)	2.790
	Sig (Sig.)	0.006
	95% Confidante Interval For Diference (Intervalo de Confiança de 95% Para a Diferença)	2.494
	Lower Bound (Limite Inferior) Upper Bound (Limite Superior)	13.756
Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Contrast Estimate (Estimativa do Contraste)	18.125
	Hypothesized Value (Valor Hipotético)	0
	Diference (Estimate – Hypothesized) (Diferença (Estimativa – Hipotético))	18.125
	Std. Error (Erro Padrão)	3.222
	Sig (Sig.)	0.000
	95% Confidante Interval For Diference (Intervalo de Confiança de 95% Para a Diferença)	11.623
	Lower Bound (Limite Inferior) Upper Bound (Limite Superior)	24.627

um intervalo de confiança contendo o valor real da diferença, portanto, a diferença real é maior do que zero (entre 2,49 e 13,76, para ser preciso). Assim, podemos concluir que o efeito do álcool é que qualquer quantidade reduz a atratividade do parceiro selecionado comparado com quando a bebida não é ingerida. É claro que isso é enganoso porque, na verdade, as médias para os grupos sem álcool e 2 canecas são muito parecidas (63,75 e 64,69), assim, 2 canecas de álcool não reduz a atratividade dos parceiros selecionados! A comparação é significativa porque ela está testando o efeito combinado de 2 e 4 canecas, e como 4 canecas tem um efeito muito forte, ele puxa a média para baixo. Isso mostra por que você deve ser cauteloso sobre como interpretar esses contrastes: você precisa observar os contrastes restantes também.

A parte de baixo da tabela mostra o contraste para *Level 2 vs. Level 3* (Nível 2 vs. Nível 3), o que nesse caso significa o grupo de 2 canecas comparado ao grupo dos 4 canecas. Isso testa se a média do grupo de 2 canecas

(64,69) é diferente da média combinada do grupo de 4 canecas (46,56). Isso é uma diferença de 18,13 (64,69 – 46,56) mostrada por ambas *Estimativa de Contraste e Diferença* na tabela. Novamente, o importante a observar é o valor de Sig. que nos mostra se a diferença é significativa. Ela é porque o Sig. é 0,000, o que é menor do que 0,5. Também é mostrado o intervalo de confiança para essa diferença e, por ela não conter zero, podemos estar certos que esse intervalo de confiança conterá o verdadeiro valor da diferença com 95% de precisão, então a verdadeira diferença é maior do que zero (entre 11,62 e 24,63, para ser preciso). Isso nos mostra que tomar 4 canecas reduziu de modo significativo a atratividade nos parceiros selecionados do que com 2 canecas.

10.4.5 Análise *post hoc* ②

Os testes *post hoc* (Saída 10.5 do SPSS) analisam o efeito principal do álcool e podem ser interpretados como se uma ANOVA de um fator independente tivesse sido conduzida na

Dica da Samanta Ferrinho



- A ANOVA de dois fatores independentes compara várias médias quando existem duas variáveis independentes e participantes diferentes foram utilizados em todas as condições experimentais; por exemplo, se quisermos saber se diferentes métodos de ensino funcionam melhor para matérias diferentes. Podemos selecionar estudantes de quatro cursos (Psicologia, Geografia, Administração e Estatística) e designá-los a um ensino baseado em aulas expositivas ou leituras de livros.
- As duas variáveis são curso e método de ensino. A saída pode ser as notas do final do curso (como uma percentagem).
- Teste para homogeneidade da variância usando o teste de Levene. Encontre a tabela com esse rótulo: se o valor na coluna denominada Sig. é menor do que 0,05, a hipótese foi violada.
- Na tabela denominada *Testes dos Efeitos Entre Sujeitos*, observe a coluna denominada Sig. para todos os três efeitos: deve existir um efeito principal para cada variável e um efeito da interação entre duas variáveis; se o valor for menor do que 0,05, o efeito é significativo. Para efeitos principais, consulte os testes *post hoc* para ver quais os grupos diferem, e para a interação, examine um gráfico de interação ou conduza uma análise de efeitos simples. (Quadro 10.2).
- Para os testes *post hoc*, novamente, observe a coluna denominada Sig. a fim de descobrir se suas comparações são significativas (elas serão se o valor de significância for menor do que 0,05).
- Teste as mesmas hipóteses para a ANOVA.

Saída 10.5 do SPSS

Multiple Comparisons (Comparações Múltiplas)Dependent Variable: *Attractiveness of Date* (Variável dependente: Atratividade do parceiro)

	(I) Alcohol Consumption (Consumo de Álcool)	(J) Alcohol Consumption (Consumo de Álcool)	Mean Difference (I - J) (Diferença Média)	Std. Error (Erro Padrão)	Sig.	95% Confidence Interval (Intervalo de 95% de confiança)	
						Lower Bound (Limite inferior)	Upper Bound (Limite Superior)
Bonferroni	None (Sem)	2 pints 4 pints	-0.9375 17.1875*	3.22172 3.22172	1.000 0.000	-8.9714 9.1536	7.0964 25.2214
	2 pints	None (Sem) 4 pints	0.9375 18.1250*	3.22172 3.22172	1.000 0.000	-7.0964 10.0911	8.9714 26.1589
	4 pints	None (Sem) 2 pints	-17.1875* -18.1250*	3.22172 3.22172	0.000 0.000	-25.2214 -26.1589	-9.1536 -10.0911
Games-Howell	None (Sem)	2 pints 4 pints	-0.9375 17.1875*	3.25860 4.16380	0.955 0.001	-8.9809 6.7981	7.1059 27.5769
	2 pints	None (Sem) 4 pints	0.9375 18.1250*	3.25860 4.35860	0.955 0.001	-7.1059 7.3104	8.9809 28.9396
	4 pints	None (Sem) 2 pints	-17.1875* -18.1250*	4.16380 4.35860	0.001 0.001	-27.5769 -28.9396	-6.7981 -7.3104

Com base nas médias observadas.

* A diferença média é significativa ao nível de 0,05.

Obs: Pint = Caneca

Attractiveness of Date (Atratividade do Parceiro)

	Alcohol Consumption (Consumo de Álcool)	N	Subset (Subconjunto)	
			1	2
Ryan-Einot-Gabriel-Welsch Range ^a (Amplitude de Ryan-Einot-Gabriel-Welsch)	4 pints	16	46.5625	
	None (Sem)	16		63.7500
	2 pints	16		64.6875
	Sig.		1.000	0.772

Médias para os grupos em subconjuntos homogêneos são mostradas.

Com base nas somas dos quadrados do Tipo III.

O termo erro é a média dos quadrados (Erro) = 83,036.

a. Alfa = 0,05.

variável **álcool** (isto é, os efeitos relatados para o álcool despencaram com respeito ao gênero). Os testes Bonferroni e Games-Howell mostram o mesmo padrão de resultados: quando os participantes não beberam álcool ou 2 canecas de álcool eles selecionaram, igualmente, parceiros atraentes. Entretanto, depois de consumirem 4 canecas os participantes selecionaram significativamente parceiros menos atraentes do que após 2 canecas ($p < 0,001$) e sem álcool ($p < 0,001$). É interessante notar que as médias da atratividade do parceiro após sem álcool e 2 ca-

necas foram tão semelhantes que a probabilidade da diferença obtida entre aquelas médias é 1 (isto é, completamente provável!) O teste REGWQ confirma que as médias das condições placebo e 2 canecas foram iguais enquanto que a média do grupo de 4 canecas foi diferente. Deve ser notado que esses testes *post hoc* ignoram o efeito interativo de gênero e álcool.

Em resumo, devemos concluir que o álcool tem um efeito na atratividade dos parceiros selecionados. No geral, após uma dose relativamente pequena de álcool (2 canecas)

Quadro 10.2

Análise dos efeitos simples ③



Uma maneira popular de analisar um termo de interação é utilizar uma técnica chamada de *análise dos efeitos simples*. Essa análise basicamente observa os efeitos de uma variável independente em níveis individuais da outra variável independente. Assim, por exemplo, nos dados do efeito álcool (você é que bebeu pouco), pudemos fazer uma análise dos efeitos simples observando o efeito do gênero em cada nível de álcool. Isso significaria pegar a média da atratividade da companhia selecionada pelos homens e compará-la com aquela para as mulheres após o placebo e, então, fazer a mesma comparação para 2 canecas e, finalmente, para 4 canecas. Outra maneira de analisar isso é comparar cada triângulo ao círculo correspondente na Figura 10.9: baseados no gráfico não encontramos diferença entre os grupos sem álcool e 2 canecas (em ambos os casos o triângulo e o círculo estão localizados quase na mesma posição), mas podemos encontrar uma diferença após 4 canecas (porque o círculo e o triângulo estão bem separados). A alternativa seria comparar a média da atratividade após o placebo, 2 canecas e 4 canecas para os homens e, então, uma análise separada para as mulheres. (Isso seria um pouco parecido como fazer a ANOVA de um fator independente no efeito do álcool em homens e, então, fazer uma ANOVA de um fator independente diferente para o efeito do álcool nas mulheres). Você não precisa ver como esses efeitos simples são calculados (embora, se você estiver interessado, está explicado no Apêndice A.6 ou no arquivo **Calculating SimpleEffects.pdf** no site www.artmed.com.br), mas é útil saber como fazê-los no SPSS. Infelizmente, eles não podem ser feitos pela caixa de diálogo e, em vez disso, você deve usar a sintaxe do SPSS (veja a Seção 2.6 para lembrar sobre a janela da sintaxe). A sintaxe que você precisa usar nesse exemplo é:

MANOVA

attract BY gender (0 1) alcohol (1 3)

* Isso inicia a ANOVA especificando a saída ou a variável dependente **attract** (atração) e o comando **BY** é seguido pelas nossas variáveis independentes **gender** e **alcohol** (gênero e álcool) – os números nos parênteses são os códigos mínimos e máximos dos grupos usados para definir essas variáveis.

/DESIGN = gender WITHIN alcohol (1) gender WITHIN alcohol (2) gender WITHIN alcohol (3)

* Isso especifica os efeitos simples. Por exemplo, “gender WITHIN alcohol (1)” solicita ao SPSS a análise do efeito do gênero no nível 1 de álcool (isto é, quando o álcool não foi usado). O número nos parênteses deve se relacionar ao nível da variável que você quer observar (sendo o nível 1 o que tem o código mais baixo no editor de dados). Se quisermos comparar o álcool nos níveis do gênero, escrevemos isso ao contrário:

/DESIGN = alcohol WITHIN gender (1) alcohol WITHIN gender (2)

/PRINT

CELLINFO

SIGNIF (UNIV MULT AVERF HF GG).

* Essas linhas finais apenas solicitam algumas descritivas para cada célula da análise (CELLINFO) e para imprimir a ANOVA principal (SIGNIF). A sintaxe para observar o efeito do gênero nos diferentes níveis de álcool está no arquivo denominado **GogglesSimpleEffects.sps**. Abra esse arquivo (verifique se você tem o **goggles.sav** carregado no editor de dados) e execute a sintaxe. A saída que você consegue será na forma de texto (melhor do que tabelas). Parte do texto irá repetir os resultados da ANOVA principal da saída do SPSS 10.3. Os efeitos simples são apresentados da seguinte maneira:

(Continua)

Quadro 10.2 (Continuação)

* * * * * Analysis of Variance - design 1 * * * * * [Análise da Variância - delineamento 1]

Tests of Significance for ATTRACT using UNIQUE sums of squares [Testes de significância para ATTRACT usando a soma dos quadrados UNIQUE]

Source of Variation

[Fonte da variação]	SS	DF	MS	F	Sig of F
WITHIN + RESIDUAL	6819,79	44	155,00		
GENDER WITHIN ALCOHOL (1)	156,25	1	156,25	1,01	0,321
GENDER WITHIN ALCOHOL (2)	76,56	1	76,56	0,49	0,486
GENDER WITHIN ALCOHOL (3)	1914,06	1	1914,06	12,35	0,001
(Model) [Modelo]	2146,87	3	715,62	4,62	0,007
(Total)	8966,67	47	190,78		
R-Squared (R^2) = [R Quadrado]	0,239				
Adjusted R-Squared [R^2 ajustado]	0,188				

Observando os valores de significância para cada efeito simples, parece que não houve diferença significativa entre homens e mulheres no nível 1 do álcool (isto é, sem álcool) ou no nível 2 do álcool (2 canecas). No entanto, houve uma diferença bem significativa no nível 3 do álcool (o que, julgando pelo gráfico, mostra que a média para os homens é consideravelmente mais baixa do que para as mulheres).

Outra coisa interessante para acompanhar os efeitos de interação é executar os contrastes para os termos de interação. Infelizmente, isso pode ser feito somente usando a sintaxe e é um processo muito envolvente, assim, se você acha que deve fazer isso, olhe no arquivo **ContrastsUsingSyntax.pdf** no site www.artmed.com.br: preparei esse arquivo especialmente para apresentar um exemplo para especificar contrastes por meio de interações.



os humanos continuam com controle dos seus julgamentos e os níveis de atratividade dos companheiros escolhidos são consistentes com o grupo-controle (sem o consumo de álcool). Entretanto, após uma dose maior de álcool, a atratividade das companhias escolhidas diminui significativamente. Esse efeito é chamado de efeito álcool (“efeito você é que bebeu pouco”). A interação também mostra uma diferença de gênero no efeito álcool. Parece que os homens têm mais probabilidade de escolher uma companheira menos atrativa quando bebem. As mulheres, em comparação, conseguem manter seus padrões apesar de estarem bêbadas. O que nós ainda não sabemos é se as mulheres se tornarão mais suscetíveis ao efeito álcool com o aumento da dose.

10.5 INTERPRETANDO GRÁFICOS DE INTERAÇÕES ②

Já observamos um gráfico de interação quando interpretamos a análise neste capítulo. Entretanto, as interações são muito importantes e a chave para o entendimento delas é ser capaz de interpretar os gráficos de interação. No exemplo deste capítulo, usamos a Figura 10.9 para concluir que a interação provavelmente refletiu o fato de que homens e mulheres escolhem, igualmente, companheiros atraentes quando não ingeriram álcool e após 2 canecas, mas com 4 canecas os padrões dos homens caem de forma significativa enquanto o das mulheres se mantém. Imagine que temos o perfil dos resultados apresentado na Figura

10.10; você acha que ainda teríamos um efeito de interação significativo?

O perfil desses dados provavelmente produziria um termo de interação significativo porque embora a atratividade dos acompanhantes dos homens e mulheres seja semelhante antes da ingestão de álcool e após a ingestão de 2 canecas de álcool, existe uma grande diferença após 4 canecas. Isso reflete um cenário onde o efeito álcool é igualmente grande em homens e mulheres após 4 canecas (e não existe sem álcool), mas age mais rapidamente nos homens: a atratividade das suas companhias despenca após 2 canecas enquanto as mulheres mantêm seus padrões até 4 canecas (nesse ponto eles sairiam com qualquer uma – até mesmo com uma macaca). Vamos tentar outro exemplo. Existe uma interação significativa na Figura 10.11?

Para os dados na Figura 10.11, é improvável que exista uma interação significativa porque o efeito do álcool é o mesmo para os homens e mulheres. Assim, para homens e mulheres a atratividade dos seus companheiros antes da ingestão de álcool é muito alta, mas

após 2 canecas eles reduzem de quantidade semelhante (a inclinação das linhas dos homens e mulheres é quase a mesma). Após 4 canecas há uma queda adicional e, novamente, essa queda é quase a mesma em homens e mulheres (as linhas, novamente, se inclinam quase no mesmo ângulo). O fato de que a linha para os homens é mais baixa do que aquela para as mulheres apenas reflete o fato de que, por todas as condições, os homens têm padrões mais baixos do que as mulheres: isso reflete um efeito principal de gênero (isto é, os homens geralmente escolhem companheiras menos atraentes do que as mulheres em todos os níveis de álcool). Dois pontos gerais que podemos concluir disso são:

- As interações significativas são mostradas acima por linhas não-paralelas num gráfico de interação. Entretanto, isso não quer dizer que linhas não-paralelas automaticamente significam que a interação é significativa: a interação ser ou não significativa irá depender do grau em que as linhas não são paralelas!

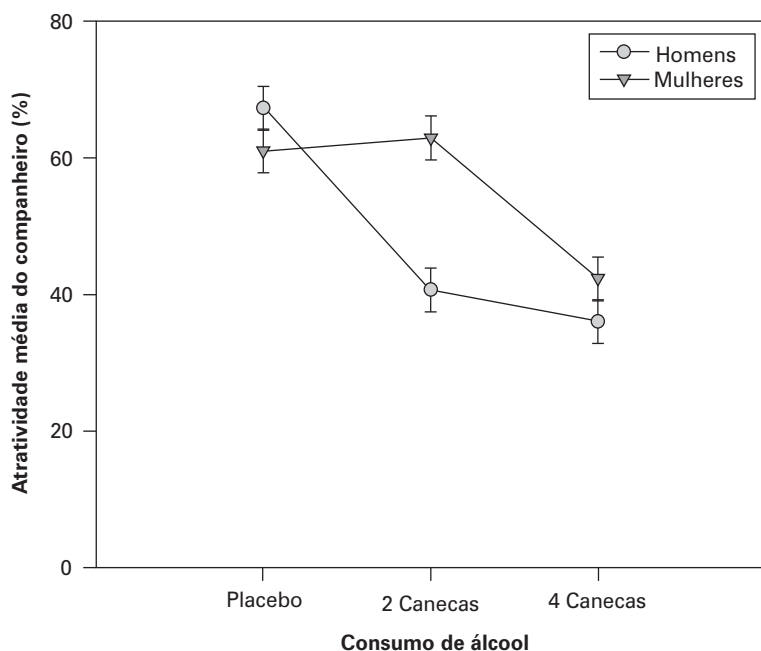


Figura 10.10 Outro gráfico da interação.

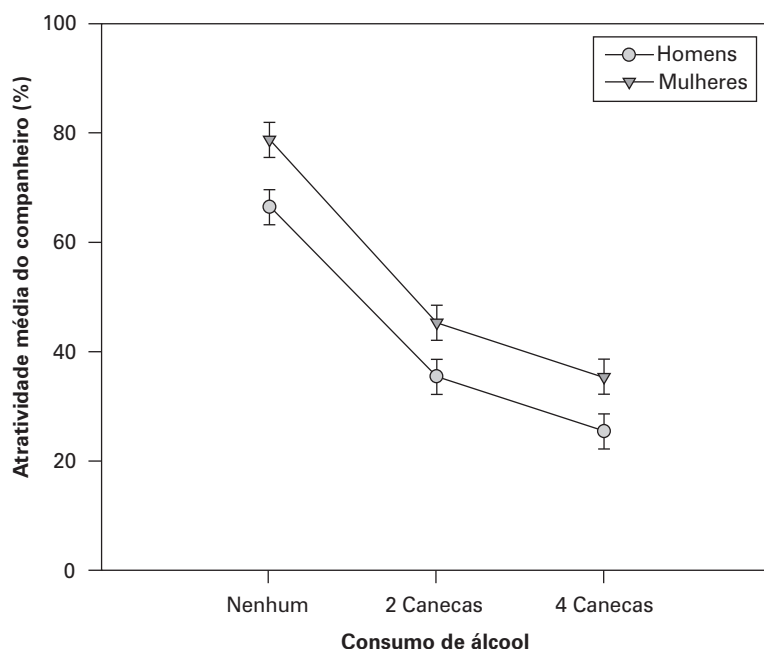


Figura 10.11 Um gráfico “sem” interação.

- Se as linhas num gráfico de interação se cruzam, obviamente elas não são paralelas e isso pode ser uma dica de que você tenha uma possível interação significativa. Entretanto, contrário à crença popular, nem *sempre* quando as linhas do gráfico de interação se cruzam, teremos uma interação significativa.

Outra complicação é que, algumas vezes, as pessoas desenham gráficos de barras em vez de gráficos de linhas. A Figura 10.12 mostra alguns gráficos de barras de interações entre duas variáveis independentes. Os painéis (a) e (b) exibem os dados do exemplo usados neste capítulo (na verdade, por que não traçá-los?). Como você pode ver, existem duas maneiras de apresentar os mesmos dados: o painel (a) mostra os dados quando os níveis de álcool estão localizados ao longo do eixo x e barras com cores diferentes são usadas para mostrar as médias para homens e mulheres e o painel (b) mostra um cenário oposto onde o gênero é traçado no eixo x e cores diferentes distinguem

a dose do álcool. Ambos os gráficos mostram um efeito de interação. O que você está procurando são as diferenças entre as barras coloridas em pontos distintos do eixo x. Assim, para o painel (a) você observa as diferenças entre as barras escuras e claras do grupo sem álcool e, então, analisa a de 2 canecas e pergunta: “A diferença entre as barras é distinta daquela do grupo sem álcool?”. Nesse caso, as barras parecem as mesmas das de sem álcool como as de 2 canecas: porém, sem interação. Entretanto, seguimos e observamos as de 4 canecas e perguntamos novamente: “A diferença entre as barras é diferente do que foi visto em qualquer outra condição?”. Nesse caso, a resposta é sim: para sem álcool e 2 canecas as barras eram quase da mesma altura, mas para os 4 canecas a barra mais clara é muito mais alta do que a barra escura. Isso mostra uma interação: o padrão das respostas muda nas 4 canecas. O painel (b) mostra a mesma coisa, mas traçado de uma maneira inversa. Novamente, você observa o padrão das respostas. Assim, observamos os homens e vemos que

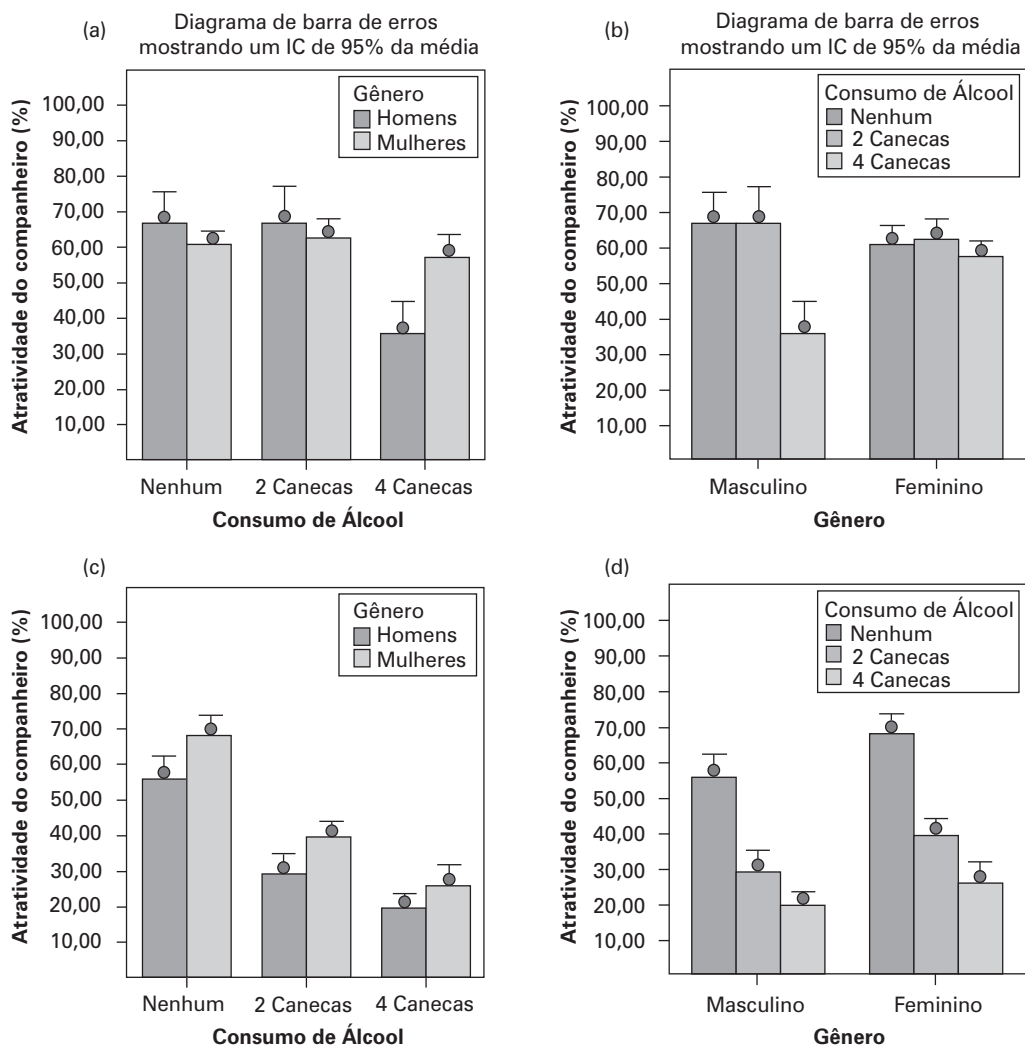


Figura 10.12 Gráfico de barras exibindo interações entre duas variáveis.

o padrão é que as primeiras duas barras têm a mesma altura, mas a última barra é muito mais baixa. O efeito de interação é mostrado acima pelo fato de que para as mulheres há um padrão diferente: todas as três barras são quase do mesmo tamanho.

E os painéis (c) e (d): você acha que existe uma interação? Novamente, eles exibem os mesmos dados de duas maneiras diferentes, mas são dados distintos dos que usamos neste capítulo. Primeiro, vamos observar o painel (c): para

os dados do grupo sem álcool a barra mais clara é um pouco maior do que a mais escura; seguindo para os dados de 2 canecas a barra mais clara é também um pouco mais alta do que a escura e, finalmente, para os dados de 4 canecas, a barra clara é um pouco mais alta do que a escura. Em todas as condições, o mesmo padrão é exibido – a barra clara é um pouco mais alta do que a escura (isto é, as mulheres escolhem parceiros mais atraentes do que os homens apesar do consumo de álcool) – portanto, não há interação.

Observando o painel (d), vemos um resultado semelhante. Para os homens, o padrão é que a taxa de atratividade cai quanto mais álcool é bebido (as barras diminuem em altura) e, então, para as mulheres vemos o mesmo padrão: quanto mais se bebe, mais a taxa cai. Isso novamente é um indicativo de que não há interação: a mudança em atratividade devido ao álcool é similar em homens e mulheres.

10.6 CALCULANDO O TAMANHO DE EFEITO ③



Como vimos nos capítulos anteriores (seções 8.5 e 9.7), podemos fazer o SPSS produzir o eta quadrado, η^2 (o qual é apenas o r^2 com outro nome). Entretanto, você foi orientado, por razões explicadas em outras seções, a usar o ômega quadrado (ω^2). O cálculo do ômega quadrado se torna um pouco mais complicado nos delineamentos fatoriais. Howell (2002), como sempre, explica muito bem as complexidades de tudo (e tem uma ótima tabela resumindo os muitos componentes para uma variedade de situações). Resumindo tudo isso, apenas direi que, em primeiro lugar, temos que calcular o componente da variância para cada um dos efeitos (os dois efeitos principais e o termo de interação) e o erro e, então, usar isso para calcular o tamanho de efeito para cada um. Se denominarmos o efeito principal de A e o segundo efeito principal de B e o efeito de interação AxB, os componentes da variância para cada um deles é baseado nas raízes quadradas de cada efeito e os tamanhos dos efeitos nos quais eles são baseados:

$$\hat{\sigma}_\alpha^2 = \frac{(a-1)(MS_A - MS_R)}{nab}$$

$$\hat{\sigma}_\beta^2 = \frac{(b-1)(MS_B - MS_R)}{nab}$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{(a-1)(b-1)(MS_{A \times B} - MS_R)}{nab}$$

Nessas equações, a é o número de níveis da primeira variável independente, b é o nú-

mero dos níveis da segunda variável independente e n é o número de pessoas por condição. Vamos calcular isso para os nossos dados. Precisamos observar a Saída 10.3 do SPSS para descobrir quadrados médios para cada efeito e para o termo erro. Nossa primeira variável independente foi o álcool. Ela tinha três níveis (portanto, $a = 3$) e tinha uma média dos quadrados de 1666,146. Nossa segunda variável independente era gênero, o qual tinha dois níveis (portanto, $b = 2$) e uma média dos quadrados de 168,75. O número de pessoas em cada grupo era de 8 e a média residual dos quadrados era de 83,036. Portanto, nossas equações ficam:

$$\hat{\sigma}_\alpha^2 = \frac{(3-1)(1666,146 - 83,036)}{8 \times 3 \times 2} = 65,96$$

$$\hat{\sigma}_\beta^2 = \frac{(2-1)(168,75 - 83,036)}{8 \times 3 \times 2} = 1,79$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{(3-1)(2-1)(989,062 - 83,036)}{8 \times 3 \times 2} = 37,75$$

Nós precisamos, também, estimar o total da variabilidade e isso é apenas a soma das outras variáveis mais a média residual dos quadrados:

$$\begin{aligned}\hat{\sigma}_{\text{Total}}^2 &= \hat{\sigma}_\alpha^2 + \hat{\sigma}_\beta^2 + \hat{\sigma}_{\alpha\beta}^2 + MS_R \\ &= 65,96 + 1,79 + 37,75 + 83,04 \\ &= 188,54\end{aligned}$$

O tamanho de efeito é simplesmente a variância estimada para o efeito no qual você está interessado dividido pela variância total estimada:

$$\omega_{\text{Efeito}}^2 = \frac{\hat{\sigma}_{\text{Efeito}}^2}{\hat{\sigma}_{\text{Total}}^2}$$

Para o efeito principal do álcool, temos:

$$\omega_{\text{Álcool}}^2 = \frac{\hat{\sigma}_{\text{Álcool}}^2}{\hat{\sigma}_{\text{Total}}^2} = \frac{65,96}{188,54} = 0,35$$

Para o efeito principal do gênero, temos:

$$\omega_{\text{Gênero}}^2 = \frac{\hat{\sigma}_{\text{Gênero}}^2}{\hat{\sigma}_{\text{Total}}^2} = \frac{1,79}{188,54} = 0,09$$

Para a interação do gênero e álcool, temos:

$$\omega^2_{\text{Álcool} \times \text{Gênero}} = \frac{\hat{\sigma}^2_{\text{Álcool} \times \text{Gênero}}}{\hat{\sigma}^2_{\text{Total}}} = \frac{37,75}{188,54} = 0,20$$



Para tornar esses valores comparáveis ao r , devemos tirar a raiz quadrada, o que nos dá os tamanhos de efeito de 0,59 para o álcool, 0,09 para o gênero e 0,45 para o termo de interação. Assim, os efeitos do álcool e a interação são muito grandes, mas o efeito do gênero, que não foi significativo na análise principal, é realmente muito pequeno (na verdade, próximo de zero).

É possível, também, calcular os tamanhos de efeito para a nossa análise de efeitos simples (se você ler o Quadro 10.2). Esses efeitos têm 1 grau de liberdade para o modelo (o que significa que eles estão comparando somente duas coisas) e nessas situações a F pode ser convertida em r usando a equação seguinte (a qual apenas utiliza a razão F e os graus de liberdade dos resíduos):¹

$$r = \sqrt{\frac{F(1, gl_R)}{F(1, gl_R) + gl_R}}$$

Observando o Quadro 10.2, podemos ver que temos as razões F de 1,01, 0,49 e 12,35 para os efeitos de gênero nos grupos sem álcool, 2 canecas 4 canecas, respectivamente. Para cada um desses, os graus de liberdade eram 1 para o modelo e 44 para o resíduo. Portanto, temos os seguintes tamanhos de efeito:

$$r_{\text{Gênero(Sem Álcool)}} = \sqrt{\frac{1,01}{1,01 + 44}} = 0,15$$

$$r_{\text{Gênero(2Canecas)}} = \sqrt{\frac{0,49}{0,49 + 44}} = 0,10$$

¹ Se o seu F compara mais do que duas coisas, uma equação diferente é necessária (veja Rosenthal, Rosnow e Rubin (2000), p. 44), mas prefiro tentar manter os tamanhos de efeito em situações nas quais somente duas coisas estão sendo comparadas porque a interpretação é mais fácil.

$$r_{\text{Gênero(4Canecas)}} = \sqrt{\frac{12,35}{12,35 + 44}} = 0,47$$

Desse modo, o efeito do gênero é muito pequeno quando não há consumo de álcool e quando são consumidas 2 canecas, mas se torna grande quando são consumidas 4 canecas.

10.7 RELATANDO OS RESULTADOS DA ANOVA DE DOIS FATORES (TWO-WAY) ②

Assim como na outra ANOVA que encontramos, temos que relatar os detalhes da razão F e os graus de liberdade dos quais ela foi calculada. Para os vários efeitos nesses dados as razões F serão baseadas em diferentes graus de liberdade: ela foi derivada pela divisão dos quadrados médios para o efeito pelos quadrados médios dos resíduos. Para os efeitos do álcool e álcool \times interação do gênero, os graus de liberdade do modelo eram 2 ($gl_M = 2$), mas para o efeito do gênero os graus de liberdade eram somente 1 ($gl_M = 1$). Para todos os efeitos os graus de liberdade para os resíduos eram 42 ($gl_R = 84$). Portanto, podemos relatar os três efeitos dessa análise como segue:

- ✓ Havia um efeito principal significativo da quantidade de álcool consumida na boate, na atratividade do parceiro selecionado, $F(2, 42) = 20,07$, $p < 0,001$, $\omega^2 = 0,35$. O teste de *post hoc* de Games-Howell revelou que a atratividade dos parceiros selecionados era significativamente menor após 4 canecas do que tanto para 2 canecas quanto na situação sem álcool (ambos $p < 0,001$). A atratividade dos parceiros após 2 canecas e na situação sem álcool não foi significativamente diferente.
- ✓ Havia um efeito principal não significativo de gênero na atratividade dos acompanhantes selecionados $F(1, 42) = 2,03$, $p = 0,161$, $\omega^2 = 0,009$.
- ✓ Havia um efeito de interação significativo entre a quantidade de álcool consumida e o gênero da pessoa selecionando um companheiro, na atratividade do

companheiro selecionado, $F(2, 42) = 11,91$, $p < 0,001$, $\omega^2 = 0,20$. Isso indica que os gêneros masculinos e femininos foram afetados de maneira diferente pelo álcool. Especificamente, a atratividade dos parceiros foi similar em homens ($M = 66,88$, $SD = 10,33$) e mulheres ($M = 60,63$, $SD = 4,96$) no grupo sem álcool; a atratividade dos parceiros foi similar em homens ($M = 66,88$, $SD = 12,52$) e mulheres ($M = 62,50$, $SD = 6,55$) após 2 canecas; entretanto, a atratividade dos parceiros selecionados pelos homens ($M = 35,63$, $SD = 10,84$) foi significativamente mais baixa do que aqueles selecionados pelas mulheres ($M = 57,50$, $SD = 7,07$) após 4 canecas.

10.8 ANOVA DE DOIS FATORES COMO REGRESSÃO ③

Vimos na Seção 8.2.2 que a ANOVA de um fator independente pode ser conceitualizada como uma equação de regressão (um modelo linear geral). Nesta seção, iremos considerar como estendemos esse modelo linear para incorporar duas variáveis independentes. Para manter as coisas tão simples quanto possível, imagine que temos somente dois níveis da variável álcool em nosso exemplo (sem álcool e 4 canecas). Assim, temos duas variáveis, cada uma com dois níveis. Todos os modelos lineares gerais que consideramos neste livro assumem a forma geral de:

$$\text{Saída}_i = (\text{Modelo}_i) + \text{erro}_i$$

Por exemplo, quando nos deparamos com a regressão múltipla no Capítulo 5, vimos que esse modelo estava escrito como (5.9):

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots b_nX_n) + \varepsilon_i$$

Também, quando nos deparamos com a ANOVA de um fator independente, adaptamos esse modelo de regressão para conceitualizar nosso exemplo do Viagra como (veja equação (8.2)):

$$\text{Libido}_i = (b_0 + b_2\text{Alto}_i + b_1\text{Baixo}_i) + \varepsilon_i$$

Nesse modelo, as variáveis alta e baixa eram variáveis auxiliares (isto é, variáveis que somente podem ter dois valores: 0 e 1). No nosso exemplo atual, temos duas variáveis: gênero (masculino e feminino) e álcool (sem e 4 canecas). Podemos codificar cada uma delas com 0 e 1: por exemplo, podemos codificar o gênero como masculino = 0, feminino = 1; e podemos codificar a variável álcool como 0 = sem, 1 = 4 canecas. Podemos, então, copiar diretamente o modelo que tínhamos na ANOVA de um fator independente:

$$\text{Atratividade}_i = (b_0 + b_2\text{Gênero}_i + b_1\text{Álcool}_i) + \varepsilon_i$$

Agora, os experts dentre vocês podem dizer: “Aonde foi o termo de interação?”. É claro que temos que incluí-lo também e, assim, o modelo simplesmente se estende para se tornar (primeiro expresso no geral e, então, em termos desse exemplo específico):

$$\text{Saída}_i = (b_0 + b_1A_1 + b_2B_i + b_3AB_i) + \varepsilon_i$$

$$\text{Atratividade}_i = (b_0 + b_1\text{Gênero}_i + b_2\text{Álcool}_i + b_3\text{Interação}_i) + \varepsilon_i \quad (10.1)$$

A questão é: como codificamos o termo de interação? O termo de interação representa o efeito combinado de álcool e gênero e, na verdade, para conseguir qualquer termo de interação na regressão você simplesmente multiplica as variáveis envolvidas naquele termo de interação. Essa é a razão pela qual você vê termos de interação escritos como gênero \times álcool, porque em termos de regressão a variável de interação literalmente é as duas variáveis multiplicadas entre si. A Tabela 10.2 mostra as variáveis resultantes para a regressão (note que a variável de interação é simplesmente o valor da variável auxiliar gênero multiplicada pelo valor da variável auxiliar álcool). Assim, por exemplo, um homem recebendo 4 canecas de álcool terá um valor de 0 para a variável de gênero, 1 para a de álcool e 0 para a variável de interação. As médias do grupo para as várias combinações de gênero e álcool estão também incluídas porque elas serão úteis no seu devido tempo.

Tabela 10.2 Esquema de codificação para a ANOVA fatorial

Gênero	Álcool	Auxiliar 1 (Gênero)	Auxiliar 2 (Álcool)	Interação	Média
Masculino	Nenhum	0	0	0	68,875
Masculino	4 canecas	0	1	0	35,625
Feminino	Nenhum	1	0	0	60,625
Feminino	4 canecas	1	1	1	57,500

Para determinar o que os valores b representam nesse modelo, podemos fazer o mesmo que fizemos para o teste t e para a ANOVA de um fator independente; isto é, veja o que acontece quando inserimos valores dos nossos previsores (gênero e álcool)! Para começar, vejamos o que acontece quando observamos os homens que não ingeriram álcool. Nesse caso, o valor de gênero é 0, o valor de álcool é 0 e o valor da interação também é 0. A saída que previmos (como na ANOVA de um fator independente) é a média desse grupo (66,875), assim, nosso modelo fica:

$$\begin{aligned}\text{Atratividade}_i &= b_0 + b_1 \text{Gênero}_i + b_2 \text{Álcool}_i \\ &\quad + b_3 \text{Interação}_i \\ \bar{X}_{\text{Homens, Sem}} &= b_0 + (b_1 \times 0) + (b_2 \times 0) \\ &\quad + (b_3 \times 0) \\ b_0 &= \bar{X}_{\text{Homens, Sem}} \\ b_0 &= 66,875\end{aligned}$$

Assim, a constante b_0 no modelo representa a média do grupo para o qual todas as variáveis estão codificadas como 0. Desse modo, é o valor da média da categoria base (nesse caso, homens que não beberam álcool). Agora, vamos ver o que acontece quando observamos as mulheres que não beberam álcool. Nesse caso, a variável gênero é 1 e as variáveis álcool e interação permanecem 0. Lembre, também, que o b_0 é a média dos homens que não beberam álcool. A saída é a média para as mulheres que não beberam álcool; portanto, a equação fica:

$$\begin{aligned}\text{Atratividade}_i &= b_0 + b_1 \text{Gênero}_i + b_2 \text{Álcool}_i \\ &\quad + b_3 \text{Interação}_i \\ \bar{X}_{\text{Mulheres, Sem}} &= b_0 + (b_1 \times 1) + (b_2 \times 0) \\ &\quad + (b_3 \times 0) \\ \bar{X}_{\text{Mulheres, Sem}} &= \bar{X}_{\text{Homens, Sem}} + b_1\end{aligned}$$

$$\begin{aligned}b_0 &= \bar{X}_{\text{Mulheres, Sem}} - \bar{X}_{\text{Homens, Sem}} \\ &= 60,625 - 66,875 \\ &= -6,25\end{aligned}$$

Assim, b_1 no modelo representa a diferença entre homens e mulheres (para aqueles) que não beberam álcool. De uma forma mais geral podemos dizer que é o efeito do gênero para a categoria base do álcool (a categoria base sendo aquela codificada com 0, nesse caso, sem álcool). Agora, vamos observar os homens que beberam 4 canecas de álcool. Nesse caso, a variável gênero é 0, a variável do álcool é 1 e a variável de interação permanece 0. Podemos, também, trocar o b_0 pela média dos homens que beberam álcool. A saída é a média para os homens que beberam 4 canecas; portanto, a equação fica:

$$\begin{aligned}\bar{X}_{\text{Homens, 4 canecas}} &= b_0 + (b_1 \times 0) + (b_2 \times 1) \\ &\quad + (b_3 \times 0) \\ \bar{X}_{\text{Homens, 4 canecas}} &= b_0 + b_2 \\ \bar{X}_{\text{Homens, 4 canecas}} &= \bar{X}_{\text{Homens, Sem}} + b_2 \\ b_2 &= \bar{X}_{\text{Homens, 4 canecas}} - \bar{X}_{\text{Homens, Sem}} \\ &= 35,625 - 66,875 \\ &= -31,25\end{aligned}$$

Assim, o b_2 no modelo representa a diferença entre sem álcool e 4 canecas para os homens. De uma forma mais geral, é o efeito do álcool na categoria base de gênero (isto é, a categoria de gênero que foi codificada com um 0, nesse caso, homens). Finalmente, podemos observar as mulheres que beberam 4 canecas. Nesse caso, a variável gênero é 1, a variável álcool é 1 e a variável interação é também 1. Podemos, também, substituir b_0 , b_1 pelo o que agora sabemos que eles representam. A saída é a média para as mulheres que beberam 4 canecas. Portanto, a equação fica:

$$\begin{aligned}
 \bar{X}_{\text{Mulheres, 4 canecas}} &= b_0 + (b_1 \times 1) + (b_2 \times 1) \\
 &\quad + (b_3 \times 1) \\
 \bar{X}_{\text{Mulheres, 4 canecas}} &= b_0 + b_1 + b_2 + b_3 \\
 \bar{X}_{\text{Mulheres, 4 canecas}} &= \bar{X}_{\text{Homens, Sem}} + (\bar{X}_{\text{Mulheres, sem}} \\
 &\quad - \bar{X}_{\text{Homens, Sem}}) + (\bar{X}_{\text{Homens, 4 canecas}} \\
 &\quad - \bar{X}_{\text{Homens, Sem}}) + b_3 \\
 \bar{X}_{\text{Mulheres, 4 canecas}} &= \bar{X}_{\text{Mulheres, Sem}} + \bar{X}_{\text{Homens, 4 canecas}} \\
 &\quad - \bar{X}_{\text{Homens, Sem}} + b_3 \\
 b_3 &= \bar{X}_{\text{Homens, Sem}} - \bar{X}_{\text{Mulheres, Sem}} \\
 &\quad + \bar{X}_{\text{Mulheres, 4 canecas}} - \bar{X}_{\text{Homens, 4 canecas}} \\
 &= 66,875 - 60,625 + 57,500 \\
 &\quad - 35,625 \\
 &= 28,125
 \end{aligned}$$

Assim, o b_3 no modelo realmente compara a diferença entre homens e mulheres na condição sem álcool com a diferença entre homens e mulheres na condição 4 canecas. Em outras palavras, ele compara o efeito do gênero após sem álcool ao efeito do gênero após 4 canecas.² Se você pensar em termos de um gráfico de interação, isso faz sentido. Por exemplo, o lado esquerdo do alto da Figura 10.13 mostra o gráfico de interação para esses dados. Agora

² De fato, se você reorganizar os termos na equação, verá que poderá também escrever a interação de forma contrária: ela representa o efeito do álcool nos homens comparados às mulheres.

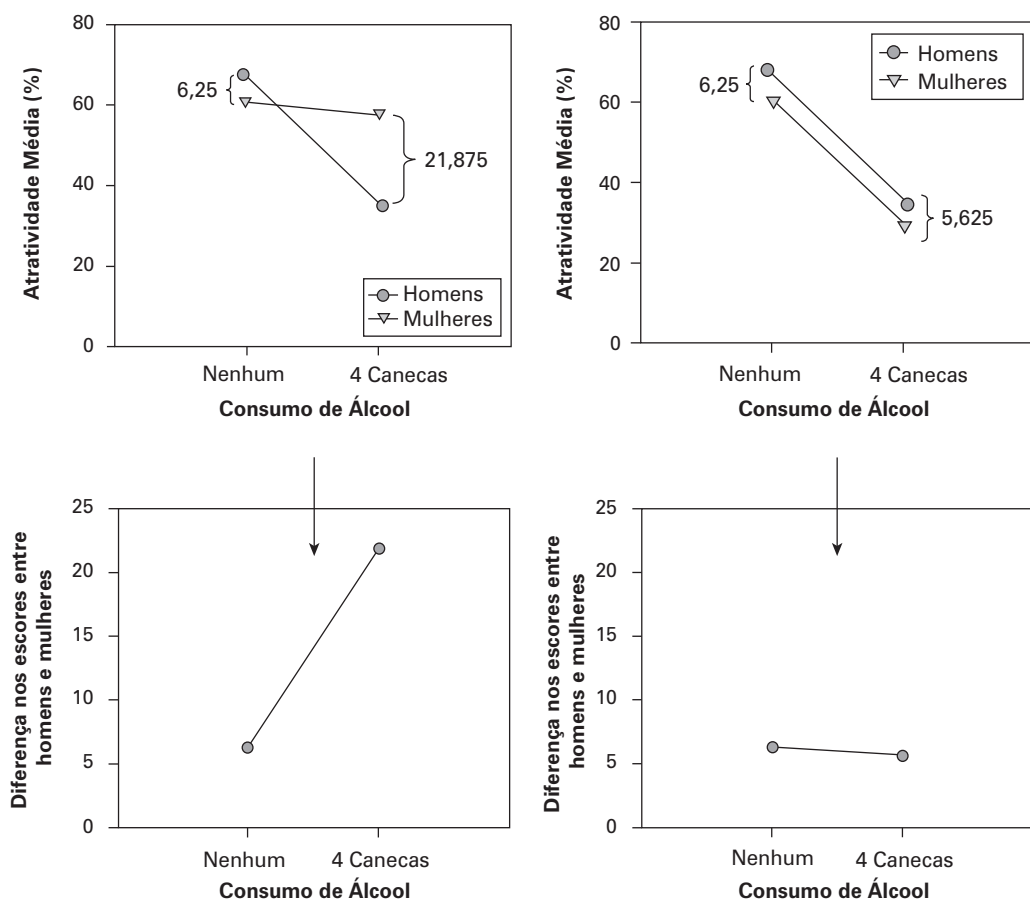


Figura 10.13 Ilustração das situações do exemplo anterior.

imagine que calculamos a diferença entre homens e mulheres para os grupos sem álcool. Isso seria a diferença entre as linhas do gráfico para o grupo sem álcool (a diferença entre as médias do grupo, que é 6,25). Se, então, fizermos o mesmo para o grupo de 4 canecas, encontraremos que a diferença entre homens e mulheres é de 21,875. Se traçarmos esses dois valores como um novo gráfico, conseguiremos uma linha conectando 6,25 a 21,875 (veja a parte de baixo da esquerda da Figura 10.13). Isso reflete a diferença entre o efeito do gênero após sem álcool comparado a após 4 canecas. Sabemos que os valores de beta representam as inclinações (gradientes) das linhas e, de fato, o b_3 no nosso modelo é a inclinação dessa linha! Vejamos, também, o que acontece quando não existe um efeito de interação: o lado direito da Figura 10.13 mostra os mesmos dados, exceto que as médias para as mulheres que beberam 4 canecas mudou para 30. Se calcularmos a diferença entre homens e mulheres após sem álcool, encontraremos o mesmo que antes: 6,25. Se calcularmos a diferença entre homens e mulheres após 4 canecas encontraremos 5,625. Se nós, novamente, traçarmos essas diferenças num novo gráfico encontraremos uma linha praticamente horizontal. Assim, quando não existe interação, as linhas conectando o efeito do gênero após sem álcool e após 4 canecas e o b_3 resultante em nosso modelo seria próximo a 0 (lembre que uma inclinação zero significa uma linha horizontal). Na verdade, seu valor real seria $6,25 - 5,625 = 0,625$.



O arquivo **GogglesRegression.sav** contém as variáveis auxiliares usadas nesse exemplo, e para provar que ele funciona, use esse arquivo e execute uma regressão múltipla nos dados. A tabela dos coeficientes resultantes está na Saída 10.6 do SPSS. É importante notar que os vários valores de beta são os mesmos que acabamos de calcular, o que deveria, espero, convencê-lo de que a ANOVA fatorial é, como em tudo que você viu, apenas uma regressão com uma roupa diferente!

Espero ter mostrado a você nesse exemplo que até mesmo as ANOVAs complexas são apenas formas de regressão (um MLG). Você ficará feliz em saber (eu ficarei, com certeza) que essa é a última vez que falarei sobre a ANOVA como MLG. Espero que você tenha entendido que podemos continuar adicionando variáveis independentes em nosso modelo. O que acontece é que essas novas variáveis são apenas adicionadas em uma equação de regressão múltipla com um valor beta associado. Assim, qualquer ANOVA (não importa quão complexa) é apenas uma forma de regressão múltipla.

10.9 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

Este capítulo foi apenas uma amostra do que é uma ANOVA fatorial. Ainda veremos mais sobre a ANOVA fatorial nos dois próximos capítulos, mas, por enquanto, observamos situações onde existem duas variáveis independentes e diferentes pessoas foram

Saída 10.6 do SPSS

Coefficients^a (Coeficientes)

Model (Modelo)	Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
	B	Std. Error (Erro Padrão)	Beta		
1 (Constant) (Constante)	66.875	3.055		21.890	0.000
Gender (Gênero)	-6.250	4.320	-0.219	-1.447	0.159
Alcohol Consumption (Consumo de Álcool)	-31.250	4.320	-1.094	-7.233	0.000
Interaction (Interação)	28.152	6.110	0.853	4.603	0.000

a. Dependent Variable: Attractiveness of Date (Variável dependente: Atratividade do parceiro)

usadas em todas as condições experimentais. Começamos observando como calcular as várias somas dos quadrados nessa análise, mas o mais importante que vimos foi que conseguimos três efeitos: dois efeitos principais (o efeito para cada variável independente) e um efeito de interação. Aprendemos como essa análise é feita no SPSS e como a saída é interpretada. Muito disso foi similar às ANOVAs que vimos nos capítulos anteriores, mas uma grande diferença foi o termo de interação. Exploramos as interações (e, especialmente, gráficos de interações) para ver como é uma interação e como traçá-la. Os leitores corajosos também descobriram como seguir uma interação com uma análise de efeitos simples e que até mesmo ANOVAs complexas são análises de regressão simples disfarçadas. Finalmente, descobrimos que calcular os tamanhos de efeito em um delineamento fatorial dá uma dor de cabeça tremenda. Até agora nos esquivamos dos delineamentos de medidas repetidas, mas no próximo capítulo, vou ser obrigado a abordar esse assunto. ☹

10.10 TERMOS-CHAVE QUE DESCOBRIMOS

- Efeito álcool (*Beer-Goggles*)
- ANOVA fatorial
- Fatores
- Delineamento fatorial independente
- Efeito de interação
- Gráficos de interação
- Efeito principal
- Delineamentos mistos
- Delineamento fatorial relacionado
- Análises de um único efeito

10.11 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** O gosto musical das pessoas tende a mudar à medida que envelhecem (meus pais, por exemplo, após anos escutando a músicas relativamente

boas quando eu era uma criança nos anos de 1970, quando chegaram aos quarenta

desenvolveram uma obsessão preocupante por música *country* – talvez tenha sido o estresse de me ter como um filho adolescente!). De qualquer modo, isso me preocupa muito porque temo que no futuro eu também escute as músicas de Garth Brooks e pense: “Menino, será que subestimei o imenso talento de Garth quando eu estava na casa dos 20?”. Assim, pensei em fazer uma pesquisa para descobrir se minha sorte está realmente selada ou se é possível ser velho e gostar de boa música. Primeiro, peguei dois grupos de pessoas (45 pessoas em cada grupo): um grupo continha pessoas jovens (o que eu arbitrariamente classifiquei como menos de 40 anos) e o outro grupo continha indivíduos mais maduros (acima de 40 anos). Essa é minha primeira variável independente, **age** (idade), e ela tem dois níveis (menos do que ou mais do que 40 anos). Então, dividi cada um desses grupos de 45 em três grupos menores de 15 e dei a tarefa de escutar Fugazi³, ABBA ou Barf Grooks (que é um músico *country* e *western* menos conhecido e não deve ser confundido com outra pessoa que tenha um nome parecido). Essa é a minha segunda variável independente, **music** (música) e tem três níveis (Fugazi, ABBA ou Barf Grooks). Havia diferentes participantes em todas as condições, o que significa que dos 45 abaixo de 40, 15 escutaram Fugazi, 15 escutaram ABBA e 15 escutaram a Barf Grooks; da mesma forma, dos 45 acima dos 40, 15 escutaram a Fugazi, 15 escutaram ABBA e 15 escutaram Barf Grooks. Após escutar uma música, cada pessoa avaliou-a numa escala variando de –100 (eu odeio essa música horrível) passando por 0 (eu sou completamente indiferente) até +100 (eu gosto muito dessa música). Essa variável é denominada **liking** (gostar). Os dados estão no arquivo **Fugazi.sav**. Execute uma ANOVA de dois fatores independentes com esses dados. ②

³ Veja <http://www.dischord.com>

- **Tarefa 2:** Usando o quadro 10.2, mude a sintaxe no **GoogleSimpleEffects.sps** para observar os efeitos do álcool nos diferentes níveis do gênero. ③
- As respostas estão no arquivo **Answers (Chapter10).pdf** e a tarefa 1 é um exemplo de Field e Hole (2003) logo, há mais detalhes sobre ela lá.

10.12 LEITURAS COMPLEMENTARES

HOWELL, D. C. *Statistical methods for psychology* (5th edition). Belmont (CA): Duxbury, 2002. Capítulo 13.

ROSENTHAL, R., ROSNOW, R. L., RUBIN, D. B. *Contrasts and effect sizes in behavioral research: a correlational approach*. Cambridge (MA): Cambridge University Press, 2000. Esse livro é bem avançado, mas é o melhor sobre contrastes e estimativas do tamanho de efeito.

ROSNOW, R. L., ROSENTHAL, R. *Beginning behavioral research: a conceptual primer* (5th edition). Englewood Cliffs (NJ): Pearson/Prentice Hall, 2005. Tem alguns capítulos maravilhosos sobre a ANOVA com um foco específico na estimativa do tamanho de efeito e alguns ótimos comentários sobre o que as interações realmente significam.

DELINEAMENTOS DE MEDIDAS REPETIDAS (MLG 4)

11.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②

Nos últimos três capítulos, aprendemos sobre um procedimento denominado ANOVA, usado para testar diferenças entre várias médias. Até agora nos concentramos em situações em que pessoas diferentes contribuem para médias diferentes; em outras palavras, pessoas diferentes participam de condições experimentais diferentes. Na verdade, não é preciso que sejam pessoas diferentes (tenho a tendência de escrever pessoas porque sou um psicólogo, assim, passo minha vida torturando, quero dizer, testando crianças em nome da ciência), podem ser plantas diferentes, companhias diferentes, diferentes lotes de terra, diferentes grupos virais, diferentes bodes ou até mesmo diferentes ornitorrincos. De qualquer modo, o ponto é que eu ignorei completamente as situações nas quais as mesmas pessoas (plantas, bodes, *hamsters*, líderes galácticos verdes do espaço com sete olhos, ou seja lá o que for) contribuem para as diferentes médias. Agora finalmente veremos o que acontece quando executamos uma ANOVA com dados de medidas repetidas. Primeiro, iremos observar detalhadamente uma hipótese adicional que precisamos considerar quando usamos os

delineamentos de medidas repetidas (*esfericidade*). Depois, iremos observar a teoria por trás da ANOVA de medidas repetidas quando temos uma variável independente simples. Iremos examinar um exemplo no SPSS antes de verificar situações onde mensuramos duas variáveis independentes. Ao concluir este capítulo, você saberá quase tudo o que precisa sobre a ANOVA de medidas repetidas.

11.2 INTRODUÇÃO AOS DELINEAMENTOS DE MEDIDAS REPETIDAS ②

“Medidas repetidas” é um termo usado quando as mesmas pessoas participam em todas as condições de um experimento. Por exemplo, você pode testar os efeitos do álcool na diversão de uma festa. Algumas pessoas podem beber muito álcool sem realmente sentir as consequências, enquanto outras, como eu, precisam tomar somente um gole de cerveja para sentir seus efeitos. Portanto, é importante controlar, para as diferenças individuais, a tolerância ao álcool. Para controlar essas diferenças individuais, podemos testar as mesmas pessoas em todas as condições do experimento: testaríamos cada pessoa após o consumo de 1 caneca, 2 canecas, 3 canecas e 4 canecas de cerveja. Seria

dado ao participante, após cada copo, um questionário avaliando sua diversão na festa. Cada participante forneceria um escore representando sua diversão antes da manipulação experimental (sem o consumo de álcool), após 1 caneca, 2 canecas e assim por diante. Esse delineamento foi planejado para o uso de medidas repetidas.

Esse tipo de delineamento tem muitas vantagens. Mais importante, ele reduz a variabilidade não-sistemática no delineamento (veja o Capítulo 7) e, assim, fornece um poder maior para detectar efeitos. As medidas repetidas também são mais econômicas porque menos participantes são exigidos. Mas existe uma desvantagem: na ANOVA entre grupos a precisão do teste F depende da hipótese de que escores em condições diferentes são independentes (Scariano e Davenport, 1987, documentaram algumas das consequências da violação dessa hipótese). Quando medidas repetidas são usadas, essa hipótese é violada: escores tirados de diferentes condições experimentais provavelmente estão relacionados porque eles vêm do mesmo participante. Desse modo, o teste F convencional não será preciso. O relacionamento entre escores em diferentes condições de tratamento significa que uma hipótese adicional tem que ser feita e, simplificando, presumimos que o relacionamento entre pares de condições experimentais seja similar (isto é, o nível de dependência entre condições experimentais é aproximadamente igual). Essa hipótese é denominada hipótese da *esfericidade*, nome difícil de pronunciar quando você está dando aulas de estatística às 9 horas da manhã.

11.2.1 A suposição de esfericidade ②



Todos aprendemos (especialmente se você leu este livro) que é crucial termos homogeneidade de variância entre condições quando analisamos dados de participantes *diferentes*, mas geralmente presumimos que esse problema “desaparece” nos delineamentos de medidas repetidas. Isso não é verdade, e a hipótese de esfericidade pode ser

igualada à hipótese de homogeneidade da variância na ANOVA entre grupos. A esfericidade (representada por ϵ e algumas vezes referida como *circularidade*) é uma condição mais geral da *simetria composta*. A simetria composta se mantém verdadeira quando ambas as variâncias entre condições são iguais (isso é o mesmo que a homogeneidade da hipótese da variância em delineamentos entre grupos) e as covariâncias entre pares de condições são iguais. Assim, presumimos que a variação dentro das condições experimentais é muito parecida e que duas condições não são mais dependentes do que quaisquer outras duas. Embora a simetria composta tenha se mostrado uma condição suficiente para a ANOVA usando dados de medidas repetidas, ela não é uma condição necessária. A esfericidade é uma forma menos restritiva da simetria composta (de fato, as primeiras pesquisas sobre ANOVA de medidas repetidas confundiam a simetria composta com esfericidade). A esfericidade se refere à igualdade das variâncias das *diferenças* entre níveis de tratamento. Assim, se você for tomar cada par dos níveis de tratamento e calcular as diferenças entre cada par dos escores, é necessário que essas diferenças tenham variâncias iguais. Assim, **você precisa, pelo menos, de três condições para que a esfericidade exista.**

11.2.2 Como a esfericidade é medida? ②

A maneira mais simples de verificar se a hipótese de esfericidade foi confirmada é calcular as diferenças entre os pares dos escores em todas as combinações dos níveis de tratamento. Uma vez que isso foi feito, você pode calcular a variância dessas diferenças. A Tabela 11.1 mostra os dados de um experimento com três condições. As diferenças entre pares de escores são computadas para cada participante e a variância para cada conjunto de diferenças é calculado. Vimos acima que a esfericidade é confirmada quando essas variâncias são aproximadamente iguais. Para esses dados, a esfericidade irá se manter quando:

$$\text{variância}_{A-B} \approx \text{variância}_{A-C} \approx \text{variância}_{B-C}$$

Tabela 11.1 Dados hipotéticos para ilustrar os cálculos da variância das diferenças entre condições

Grupo A	Grupo B	Grupo C	A – B	A – C	B – C
9	12	7	–3	2	5
15	15	12	0	3	3
25	30	20	–5	5	10
35	30	28	5	7	2
30	27	20	3	10	7
Variância			17,0	10,3	10,3

Nesses dados existe um desvio da esfericidade porque a variância das diferenças entre as condições A e B (17,0) é maior do que a variância das diferenças entre A e C e entre B e C (10,3). Entretanto, esses dados têm *circularidade local* (ou esfericidade local) porque duas das variâncias das diferenças são iguais. Portanto, a hipótese da esfericidade foi confirmada para quaisquer comparações múltiplas envolvendo essas condições (para uma discussão sobre circularidade local, veja Rouanet e Lépine, 1970). O desvio da esfericidade nos dados da Tabela 11.1 não parece muito forte (todas as variâncias são aproximadamente iguais), mas podemos avaliar se o desvio é forte o suficiente para exigir uma ação?

11.2.3 Interpretando as violações da hipótese de esfericidade ②

O SPSS produz o *teste de Mauchly*, o qual testa a hipótese de que as variâncias das diferenças entre condições são iguais. Portanto, se o teste estatístico de Mauchly é significativo (isto é, tem um valor de probabilidade menor do que 0,05), devemos concluir que existem diferenças significativas entre as variâncias das diferenças; portanto, a condição de esfericidade foi violada. Se, entretanto, o teste estatístico de Mauchly não é significativo (isto é, $p > 0,05$), é razoável concluir que as variâncias das diferenças não são significativamente diferentes (isto é, elas são aproximadamente iguais). Resumindo, se o teste de Mauchly é significativo, precisamos ser cautelosos com as razões F produzidas pelo computador.

11.2.4 Qual é o efeito de violar a hipótese de esfericidade? ③



Rouanet e Lépine (1970) forneceram um relato detalhado da validade da razão F sob as violações da hipótese da esfericidade. Eles argumentaram que existem duas razões F diferentes que podem ser usadas para validar comparações de tratamento denominadas F' e F'' , respectivamente. F' se refere à razão F derivada da média dos quadrados da comparação em questão e o termo erro específico para a comparação de interesse – essa é a razão F normalmente usada. F'' é derivada não do quadrado médio do erro específico, mas dos quadrados médios do erro total para *todas* as comparações de medidas repetidas. Rouanet e Lépine (1970) argumentaram que F' tem menos poder do que F'' , assim, talvez esse teste estatístico não detecte os efeitos genuínos. Além disso, eles mostraram que para F' ser válida, a esfericidade deve se manter para a *comparação específica em questão* (veja também Mendoza, Toothaker e Nicewander, 1974). F'' exige somente uma circularidade *geral* (isto é, todo o conjunto de dados deve ser circular), mas por causa da natureza não-recíproca da circularidade e simetria composta, a F'' não exige simetria composta, embora a F' exija. Assim, dado que F' é a estatística geralmente usada, o efeito da violação da esfericidade é a perda de poder (comparada quando F'' é usado) e um teste estatístico (razão F), o qual não pode ser comparado aos valores tabulados da distribuição F (para mais detalhes veja Field, 1998a).



11.2.5 O que fazer se a hipótese de esfericidade for violada? ②

Se os dados violam a hipótese de esfericidade, existem várias correções que podem ser aplicadas para produzir uma razão F válida. O SPSS produz três correções baseadas nas estimativas da esfericidade defendidas por Greenhouse e Geisser (1959) e Huynh e Feldt (1976). Ambas estimativas utilizam um fator de correção aplicada aos graus de liberdade usados para avaliar a razão F observada. O cálculo dessas estimativas está além do escopo deste livro (leitores interessados devem consultar Girden, 1992); precisamos saber somente que as três estimativas diferem. A *correção de Greenhouse e Geisser* (usualmente denotada como $\hat{\epsilon}$) varia entre $1/(k - 1)$ (onde k é o número das condições das medidas repetidas) e 1. Quanto mais próximo $\hat{\epsilon}$ estiver de 1,00, mais homogêneas serão as variâncias das diferenças e portanto os dados estarão mais próximos da esfericidade. Por exemplo, numa situação na qual existem cinco condições, o limite mais baixo de $\hat{\epsilon}$ será de $1/(5 - 1)$ ou 0,25 (conhecido como a estimativa limite inferior da esfericidade).



Huynh e Feldt (1976) relataram que quando a estimativa de Greenhouse e Geisser é maior do que 0,75, muitas hipóteses nulas falsas não são rejeitadas (isto é, a correção é muito conservadora) e Collier, Baker, Mandeville e Hayes (1967) mostraram que isso também aconteceu com uma estimativa de esfericidade de 0,90. Huynh e Feldt, portanto, propuseram sua própria correção menos conservadora (usualmente denotada como $\tilde{\epsilon}$). Entretanto, Maxwell e Delaney (1990) relatam que $\tilde{\epsilon}$ superestima a esfericidade. Stevens (1992), portanto, recomenda calcular a média das duas e ajustar o gl para esse valor médio. Girden (1992) recomenda que quando a estimativa da esfericidade é maior do que 0,75, a *correção de Huynh-Feldt* deve

ser usada, mas quando a estimativa da esfericidade for menor do que 0,75 ou nada é conhecido sobre a esfericidade, a correção Greenhouse e Geisser deve ser utilizada. Veremos como usar esses valores no devido tempo.

Uma última opção para dados que violam a esfericidade é usar a estatística teste multivariada (MANOVA – veja o Capítulo 14), porque ela não depende da hipótese da esfericidade (veja O'Brien e Kaiser, 1985). A MANOVA evita a hipótese de esfericidade (e todas as considerações correspondentes sobre as razões F apropriadas e correções) usando um termo erro específico para contrastes com 1 gl e, portanto, cada contraste é somente associado com seu termo erro específico (em vez dos termos erro combinados usados na ANOVA). A MANOVA é tratada em profundidade no Capítulo 14, mas o procedimento de medidas repetidas do SPSS produz automaticamente estatísticas teste multivariadas.



Existe uma troca no poder do teste entre as abordagens univariada e multivariada (embora alguns autores argumentem que isso pode ser resolvido com um domínio adequado das técnicas – O'Brien e Kaiser, 1985). Davidson (1972) comparou o poder das técnicas univariadas ajustadas com as do T^2 do Hotelling (um teste estatístico MANOVA) e descobriu que a técnica univariada era relativamente menos poderosa para detectar pequenas mudanças confiáveis entre condições altamente correlacionadas quando outras condições menos correlacionadas estavam também presentes. Mendoza e colaboradores (1974) conduziram um estudo Monte Carlo comparando as técnicas univariadas e multivariadas sob violações de simetria e normalidade composta e descobriram que “à medida que o grau de violação da simetria composta aumenta, o poder empírico para os testes multivariados também aumenta. Em contraste, o poder para os testes univariados geralmente diminui” (p.174). Maxwell e Delaney (1990) observaram que o teste univariado é relativamente mais poderoso do que o teste multivariado à medida que n diminui e propôs

que “a abordagem multivariada não deve provavelmente ser usada se o n é menor do que $a + 10$ (a é o número de níveis para as medidas repetidas)” (p.602). Como uma regra, parece que quando temos uma grande violação na esfericidade ($\epsilon < 0,7$) e o tamanho da amostra é maior do que ($a + 10$) os procedimentos da multivariada são mais poderosos, enquanto com tamanhos da amostra pequenos ou quando a esfericidade existe ($\epsilon > 0,7$), a abordagem univariada é preferível (Stevens, 1992). Vale a pena notar, também, que o poder da MANOVA aumenta ou diminui como uma função das correlações entre as variáveis dependentes (Cole, Maxwell, Arvey e Salas, 1994), assim, o relacionamento entre as condições de tratamento também deve ser considerado.



11.3 TEORIA DA ANOVA DE MEDIDAS REPETIDAS DE UM FATOR ②

Na ANOVA de medidas repetidas, o efeito do nosso experimento aparece na variância entre participantes (em vez da variância entre grupos). Lembre que na ANOVA independente (Seção 8.2), a variância dentre participantes é a nossa variância residual (SS_R); ela é a variância criada por diferenças individuais no desempenho. Essa variância não está contaminada pelo efeito experimental porque todas as manipulações foram feitas em pessoas diferentes. Entretanto, quando realizamos nossa manipulação experimental com as mesmas pessoas, a variância dentre participantes será constituída de duas coisas: do efeito da nossa manipulação e, como anteriormente, de diferenças individuais no desempenho. Assim, algumas variações dentre sujeitos vêm dos efeitos da nossa manipulação experimental: fizemos coisas diferentes em cada manipulação experimental com os participantes e, assim, a variância nos escores de um indivíduo será, parcialmente, devido a essas manipulações. Por exemplo, se todos tiverem escores mais altos em uma condição do que na outra, é razoável presumir que isso não aconteceu por acaso, mas porque fizemos algo diferente com os participantes em uma

das condições comparada com outra qualquer. *Porque* fizemos a *mesma* coisa com todos em uma condição particular, qualquer variação que não possa ser explicada pelas manipulações que realizamos deve ser devido a fatores aleatórios fora do nosso controle, não-relacionados com nossas manipulações experimentais (podemos chamar isso de “erro”). Como na ANOVA independente, usamos uma razão F que compara o tamanho da variação devido às nossas manipulações experimentais com o tamanho da variação devido a fatores aleatórios, a única diferença sendo como calculamos essas variâncias. Se a variância, devido a nossas manipulações é grande relativa à variância devido a fatores aleatórios, conseguimos um valor alto de F e podemos concluir que os resultados observados são pouco prováveis de terem aparecido por acaso.

A Figura 11.1 mostra como a variância é dividida em uma ANOVA de medidas repetidas. É importante notar que temos os mesmos tipos de variâncias que na ANOVA independente: temos a soma dos quadrados total (SS_T), a soma dos quadrados do modelo (SS_M) e a soma dos quadrados residual (SS_R). A *única* diferença entre medidas repetidas e a ANOVA independente é de onde as somas dos quadrados vêm: na ANOVA de medidas repetidas, a soma dos quadrados do modelo e a residual são partes da variância dentre participantes. Vamos ver um exemplo.

Em geral, os estudantes se preocupam com a consistência das notas atribuídas por diferentes professores. É comum que os professores tenham reputação de ser “rígidos” ou “flexíveis” com as notas (ou usando a terminologia dos estudantes “manifestações diabólicas das entranhas de Belzebu” e “boa gente”), mas geralmente não há provas para confirmar a reputação. Um grupo de estudantes investigou a consistência das notas submetendo os mesmos trabalhos a quatro professores diferentes. A nota que cada professor atribui a cada trabalho foi registrada para cada um de oito trabalhos diferentes. Era importante que todos os trabalhos fossem avaliados por todos os professores porque

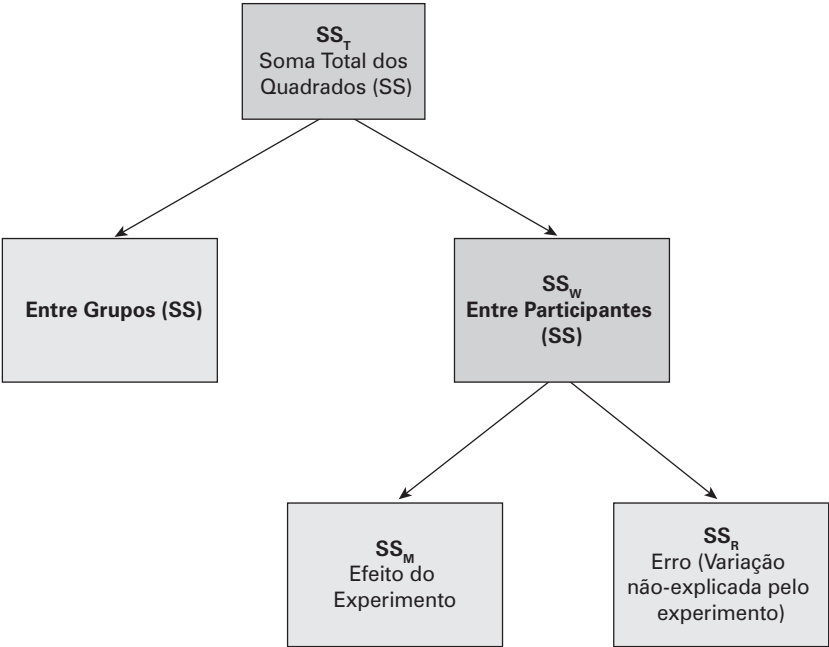


Figura 11.1 Partição da variância para a ANOVA de medidas repetidas.

isso eliminava quaisquer diferenças individuais no padrão da avaliação do trabalho que cada professor pontuou. Esse delineamento é de medidas repetidas porque cada professor deu uma nota para cada trabalho. A variável independente foi o professor que avaliou o trabalho e a variável dependente foi a nota atribuída ao trabalho.

A Tabela 11.2 mostra os dados para esse exemplo. Havia oito trabalhos, cada um avaliado por quatro professores diferentes. As notas estão apresentadas na tabela. Além disso, a média das notas dadas por cada professor é mostrada na tabela e, também, a média das notas que cada trabalho recebeu e a variância das pontuações para um trabalho específico.

Tabela 11.2 Dados para o exemplo das notas de trabalhos

Trabalho	Professor 1 (Dr. Field)	Professor 2 (Dr. Ferreira)	Professor 3 (Dr. Scrote)	Professor 4 (Dr. Morte)	Média	S ²
1	62	58	63	64	61,75	6,92
2	63	60	68	65	64,00	11,33
3	65	61	72	65	65,75	20,92
4	68	64	58	61	62,75	18,25
5	69	65	54	59	61,75	43,58
6	71	67	65	50	63,25	84,25
7	78	66	67	50	65,25	132,92
8	75	73	75	45	67,00	216,00
Média	68,875	64,25	65,25	57,375		

Agora, a variância total dentre os trabalhos irá, em parte, ser causada pelo fato de que professores diferentes são mais rígidos ou flexíveis na nota (a manipulação) e irá, em parte, ser causada pelo fato de que os próprios trabalhos irão diferir em qualidade (diferenças individuais).

11.3.1 A Soma dos Quadrados Total (SS_T) ②



Lembre da ANOVA de um fator independente que a SS_T é calculada usando a seguinte equação (veja a equação (8.4)):

$$SS_T = s_{\text{Total}}^2(N - 1)$$

Nos delineamentos de medidas repetidas, a soma dos quadrados total é calculada exatamente da mesma maneira. A variância total na equação é simplesmente a variância de todos os escores quando ignoramos o grupo ao qual eles pertencem. Assim, se tratarmos os dados como um grande grupo eles ficarão como segue:

62	58	63	64
63	60	68	65
65	61	72	65
68	64	58	61
69	65	54	59
71	67	65	50
78	66	67	50
75	73	75	45

Média Geral = 63,9375

A variância desses escores é de 55,028 (tente fazer isso com a sua calculadora). Usamos 32 escores para gerar esse valor e, assim, N é 32. A equação fica:

$$\begin{aligned} SS_T &= s_{\text{Total}}^2(N - 1) \\ &= 55,028(32 - 1) \\ &= 1705,868 \end{aligned}$$

Os graus de liberdade para essa soma dos quadrados, como na ANOVA independente, serão $N - 1$ ou 31.

11.3.2 A variação dentre participantes (SS_W) ②

A variação crucial nesse delineamento é que existe um componente da variância denominado variância dentre participantes (ela surge porque manipulamos nossa variável independente dentre cada participante). Ela é calculada usando a soma dos quadrados. Em termos gerais, quando calculamos qualquer soma de quadrados observamos a diferença ao quadrado entre a média e escores individuais. Isso pode ser expresso em termos da variância através de um número de escores e o número de escores no qual a variância é baseada. Por exemplo, quando calculamos a soma dos quadrados residual na ANOVA independente (SS_R) usamos a seguinte equação (veja a equação (8.7)):

$$\begin{aligned} SS_R &= \sum (x_{ik} - \bar{x}_k)^2 \\ SS_R &= \sum s_k^2(n_k - 1) \end{aligned}$$

Essa equação nos dá a variância entre indivíduos dentro de um grupo particular e, assim, é uma estimativa das diferenças individuais dentro desse grupo. Portanto, para conseguir o valor total das diferenças individuais temos que calcular a soma dos quadrados dentre cada grupo e, então, somá-las:

$$\begin{aligned} SS_R &= s_{\text{Grupo1}}^2(n_1 - 1) + s_{\text{Grupo2}}^2(n_2 - 1) \\ &\quad + s_{\text{Grupo3}}^2(n_3 - 1) \end{aligned}$$

Isso funciona bem quando temos pessoas diferentes em cada grupo, mas em delineamentos de medidas repetidas submetemos as pessoas a mais de uma condição experimental e, portanto, estamos interessados na variação não dentro de um grupo de pessoas (como na ANOVA independente), mas em uma só pessoa (a variabilidade dos escores de uma mesma pessoa). Para descobrir isso usamos, na verdade, a mesma equação, mas a adaptamos para pessoas e não grupos. Assim, se chamarmos isso de soma dos quadrados SS_W (para a SS dentre participantes), podemos escrevê-la como:

$$\begin{aligned} SS_W &= s_{\text{Pessoa1}}^2(n_1 - 1) + s_{\text{Pessoa2}}^2(n_2 - 1) \\ &\quad + s_{\text{Pessoa3}}^2(n_3 - 1) + \dots + s_{\text{Pessoa n}}^2(n_n - 1) \end{aligned}$$

Essa equação simplesmente significa que estamos observando a variação nos escores dos indivíduos e, então, acrescentando essas variâncias para todas as pessoas do estudo. Alguns de vocês podem ter notado que, no nosso exemplo, estamos usando trabalhos em vez de pessoas; para ser detalhista escrevemos assim:

$$SS_W = s^2_{\text{Trabalho1}}(n_1 - 1) + s^2_{\text{Trabalho2}}(n_2 - 1) + s^2_{\text{Trabalho3}}(n_3 - 1) + \dots + s^2_{\text{Trabalho}n}(n_n - 1)$$

Os n_s simplesmente representam o número de escores nos quais as variâncias estão baseadas (isto é, o número de condições experimentais ou, nesse caso, o número de professores). Todas as variâncias que precisamos estão na Tabela 11.2, portanto, podemos calcular o valor do SS_W como:

$$\begin{aligned} SS_W &= s^2_{\text{Trabalho1}}(n_1 - 1) + s^2_{\text{Trabalho2}}(n_2 - 1) + s^2_{\text{Trabalho3}}(n_3 - 1) + \dots + s^2_{\text{Trabalho}n}(n_n - 1) \\ &= (6,92)(4 - 1) + (11,33)(4 - 1) + (20,92)(4 - 1) + (18,25)(4 - 1) + (43,58)(4 - 1) + (84,25)(4 - 1) + (132,92)(4 - 1) + (216)(4 - 1) \\ &= 20,76 + 34 + 62,75 + 54,75 + 130,75 + 252,75 + 398,75 + 648 \\ &= 1602,5 \end{aligned}$$

Os graus de liberdade para cada pessoa são $n - 1$ (isto é, o número de condições menos 1). Para conseguir o total dos graus de liberdade somamos o gl para todos os participantes. Assim, com oito participantes (trabalhos) e quatro condições (isto é, $n = 4$) temos $8 \times 3 = 24$ graus de liberdade.

11.3.3 A Soma dos Quadrados do Modelo (SS_M) ②

Até agora sabemos que o total da variação entre os dados é de 1705,868 unidades. Sabemos, também, que 1602,5 dessas unidades são explicadas pela variância criada pelos desempenhos (trabalhos) dos indivíduos sob diferentes condições. Algumas dessas variações são resultados da nossa manipulação experimental

e algumas são simplesmente flutuações aleatórias. O próximo passo é analisar quanta variância é explicada pela nossa manipulação e quanta variância não é explicada.

Na ANOVA independente, analisamos quanta variação poderia ser explicada pelo nosso experimento (o modelo SS) observando todas as médias para cada grupo e as comparando com a média geral. Assim, mensuramos a variância resultante das diferenças entre as médias dos grupos e a média geral (veja a equação (8.5)). Fazemos o mesmo com o delineamento de medidas repetidas. Em primeiro lugar, calculamos a média para cada nível da variável independente (nesse caso, a média das pontuações dadas por cada professor) e comparamos esses valores com a média total de todas as pontuações.

Assim, calculamos esse SS da mesma maneira do que para a ANOVA independente:

1. Calcule as diferenças entre as médias para cada grupo e a média total.
2. Eleve ao quadrado cada uma dessas diferenças.
3. Multiplique cada resultado pelo número de entidades daquele grupo (n_i)
4. Adicione os valores de cada um dos grupos.

$$SS_M = \sum n_k(\bar{x}_k - \bar{x}_{\text{Geral}})^2$$

Usando as médias para os dados dos trabalhos (veja a Tabela 11.2), podemos calcular o SS_M como segue:

$$\begin{aligned} SS_M &= 8(68,875 - 63,9375)^2 + 8(64,25 - 63,9375)^2 + 8(65,25 - 63,9375)^2 + 8(57,375 - 63,9375)^2 \\ &= 8(4,9375)^2 + 8(0,3125)^2 + 8(1,3125)^2 + 8(-6,5625)^2 \\ &= 554,125 \end{aligned}$$

Para o SS_M os graus de liberdade (gl_M) são, novamente, 1 a menos do que o número das coisas usadas para calcular essa soma dos quadrados. Para a soma dos quadrados do modelo, calculamos a soma dos erros ao quadrado entre as quatro médias e a média total. Portanto, usa-

mos quatro coisas para calcular essas somas dos quadrados. Assim, os graus de liberdade serão 3. Assim como com a ANOVA independente, os graus de liberdade do modelo são sempre iguais ao número de condições (k) menos 1:

$$gl_M = k - 1 = 3$$

11.3.4 A Soma dos Quadrados dos Resíduos (SS_R) ②

Sabemos agora que existem 1706 unidades de variação a serem explicadas nos nossos dados e que a variação ao longo de nossas condições é responsável por 1602 unidades. Dessas 1602 unidades, nossa manipulação experimental pode explicar 554 unidades. A soma final dos quadrados é a soma dos quadrados dos resíduos (SS_R) o que nos diz quanta variação não pode ser explicada pelo modelo. Esse valor é a quantidade de variação causada por fatores estranhos fora do controle experimental (tal como a variação natural na qualidade dos trabalhos). Conhecendo os SS_W e SS_M , a maneira mais simples de calcular o SS_R é subtrair o SS_M do SS_W ($SS_R = SS_W - SS_M$):

$$\begin{aligned} SS_R &= SS_W - SS_M \\ &= 1602,5 - 554,125 \\ &= 1048,375 \end{aligned}$$

Os graus de liberdade são calculados de maneira similar:

$$\begin{aligned} gl_R &= gl_W - gl_M \\ &= 24 - 3 \\ &= 21 \end{aligned}$$

11.3.5 A média dos quadrados ②

O SS_M nos informa quanta variação o modelo (por exemplo, a manipulação experimental) explica e o SS_R nos diz quanta variação é devida a fatores estranhos. Entretanto, porque esses dois valores são valores somados, o número de escores somados os influencia. Como com a ANOVA independente, eliminamos essa tendenciosidade calculando a média da

soma dos quadrados (conhecida como *média dos quadrados*, MS) o que simplesmente é a soma dos quadrados dividida pelo número de graus de liberdade:

$$\begin{aligned} MS_M &= \frac{SS_M}{gl_M} = \frac{554,125}{3} = 184,708 \\ MS_R &= \frac{SS_R}{gl_R} = \frac{1048,375}{21} = 49,923 \end{aligned}$$

O MS_M representa a quantidade média de variação explicada pelo modelo (por exemplo, a variação sistemática), enquanto o MS_R é a medida da quantidade média de variação explicada pelas variáveis estranhas (a variação não-sistemática).

11.3.6 A razão F ②

A razão F é uma medida da razão da variação explicada pelo modelo e a variação explicada pelos fatores não-sistemáticos. Ela pode ser calculada dividindo a média dos quadrados do modelo pela média dos quadrados dos resíduos. Você deve lembrar que isso é exatamente o mesmo para a ANOVA independente:

$$F = \frac{MS_M}{MS_R}$$

Assim, como com a ANOVA independente, a razão F é ainda a razão entre a variação sistemática e a variação não-sistemática. Como tal, ela é a razão do efeito experimental para o efeito no desempenho dos fatores não-explicados. Para os dados das notas a razão F é:

$$F = \frac{MS_M}{MS_R} = \frac{184,708}{49,923} = 3,70$$

Esse valor é maior do que 1, o que indica que a manipulação experimental tem algum efeito sobre as notas além do efeito de fatores estranhos. Assim como com a ANOVA independente, esse valor pode ser comparado com um valor crítico baseado nos seus graus de liberdade (gl_M e gl_R que são 3 e 21 nesse caso).

11.4 ANOVA DE MEDIDAS REPETIDAS DE UM FATOR UTILIZANDO O SPSS ②

11.4.1 A análise principal ②

Continuando com o exemplo das notas dos trabalhos, no Capítulo 2, nos deparamos com a regra básica do editor de dados: **cada linha representa os dados de um dos participantes e cada coluna representa um nível de uma variável**. Portanto, colunas separadas representam níveis de uma variável de medidas repetidas. Assim, não há necessidade de uma variável código (como nos delineamentos entre grupos). Os dados estão na Tabela 11.2 e podem ser digitados no editor de dados do SPSS no mesmo formato que estão na tabela (você não precisa incluir as colunas denominadas trabalho, média ou s^2 porque elas foram incluídas somente para esclarecer que os professores pontuaram os mesmos trabalhos e para ajudar na explicação de como essa ANOVA é calculada). Para começar, crie uma variável denominada **tutor1** (professor 1) e use a caixa de diálogo **labels** (rótulos) para dar a essa variável um título completo de “Dr. Field”. Na próxima coluna, crie uma variável denominada de **tutor2** (professor 2) e dê a essa variável um título completo de “Dr. Ferreira”. Faça o mesmo para criar as variáveis restantes denominadas **tutor3** e **tutor4**. Esses dados também podem ser encontrados no arquivo **TutorMarks.sav**.

Para executar uma ANOVA usando o delineamento de medidas repetidas, selecione a caixa de diálogo **define factors** (defina fatores) seguindo o caminho do menu

Analyze⇒General Linear Model⇒Repeated Measures... (Analisar⇒Modelo Linear Generalizado⇒Medidas Repetidas...) Na caixa de diálogo **define factors** (Figura 11.2,) é solicitado que você dê um nome para a variável dentre participantes (medidas repetidas). Nesse caso, a variável de medidas repetidas foi o professor que pediu o trabalho, assim, a palavra **factor1** (fator 1) deve ser trocada pela palavra **tutor** (professor). O nome que você der para a variável de medidas repetidas está restrito a oito caracteres. Quando você tiver dado um nome ao fator de medidas repetidas, você deve dizer ao computador quantos níveis havia para aquela variável (isto é, quantas condições experimentais havia). Nesse caso, há quatro professores, assim, temos que entrar com o número 4 no quadro denominado **Number of Levels** (Número de Níveis). Clique em **Add** (Adicionar) para acrescentar essa variável à lista de variáveis de medidas repetidas. Essa variável irá agora aparecer na caixa branca na parte de baixo da caixa de diálogo e aparece como **tutor(4)**. Se o seu delineamento tem diversas variáveis de sujeitos, você pode acrescentar mais fatores à lista. Quando você entrou com todos os fatores de sujeitos que foram mensurados clique em **Define** (Definir) para ir à caixa de diálogo principal.

A caixa de diálogo principal (Figura 11.3) tem um espaço denominado **Within-Subjects Variables** (variáveis dentre participantes) que contém uma lista com quatro pontos de interrogação seguidos por um número. Esses pontos de interrogação são para as variáveis representando os quatro níveis da variável independente. As variáveis corres-

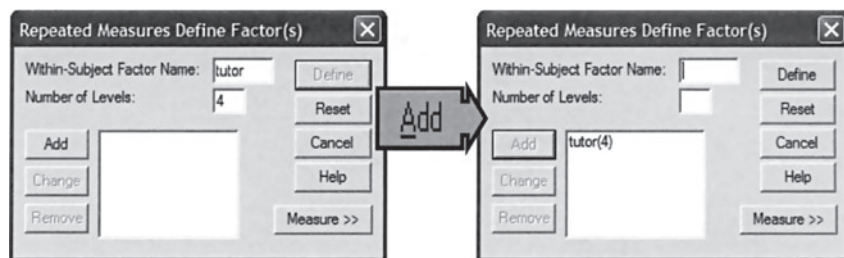


Figura 11.2 Caixa de diálogo para definir fatores para a ANOVA de medidas repetidas.

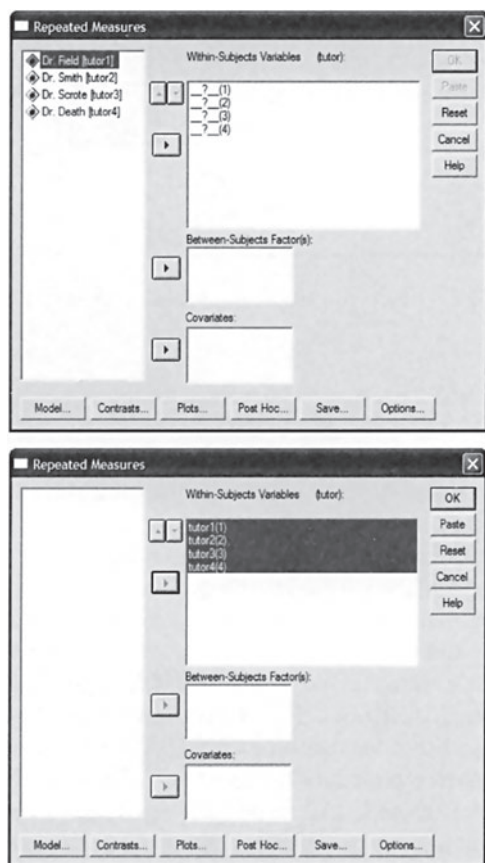


Figura 11.3 Caixa de diálogo principal para a ANOVA de medidas repetidas (antes e após a finalização).

pondentes a esses níveis devem ser selecionadas e colocadas no lugar apropriado. Temos somente quatro variáveis no editor de dados, assim, é possível selecionar todas as quatro variáveis de uma vez (clicando na variável do topo, pressionando o mouse e arrastando-o sobre das outras variáveis). As variáveis selecionadas podem ser transferidas clicando em . Quando todas as variáveis foram transferidas, você pode selecionar várias opções para a análise. Existem várias opções que podem ser acessadas com botões na parte de baixo da caixa de diálogo principal. Essas opções são semelhantes às aquelas que acabamos de encontrar.

11.4.2 Definindo contrastes para medidas repetidas ②

Não é possível especificar comparações planejadas definidas pelo usuário para os delineamentos de medidas repetidas no SPSS.¹ Entretanto, existe a opção de conduzir um dos muitos contrastes padrão que vimos anteriormente (veja a Seção 9.3.3 para detalhes de como mudar contrastes). Se você clicar em **Contrasts...** (Contrastes), na caixa de diálogo principal, pode acessar a caixa de diálogo **contrasts** (Figura 11.4). O contraste padrão é o polinomial; para mudar esse padrão, selecione a variável no quadro denominada **Factors** (Fatores), clique em próximo da caixa **Contrast**, (Contrastes), selecione um contraste da lista e clique em **Change** (Alterar). Se você escolher realizar um contraste simples, pode especificar se você gostaria de comparar os grupos contra a primeira ou a última categoria. A primeira categoria seria aquela que você entrou como (1) na caixa de diálogo e, para esses dados, a última categoria seria a que entrou como (4). Portanto, a ordem na qual você entra as variáveis na caixa de diálogo principal é importante para os contrastes que você escolher.

Não existe um contraste particularmente bom para os dados que temos (o contraste simples não é muito útil porque não temos uma categoria de controle), assim, sugiro usar o contraste **repeated** (repetido), que irá comparar cada professor com o anterior. Esse contraste pode ser útil nos delineamentos de medidas repetidas nos quais os níveis da variável independente tem uma ordem significativa. Um exemplo é se você mensurou a variável dependente em pontos sucessivos no tempo ou administrou doses crescentes de um remédio. Quando você tiver selecionado esse contraste, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

¹ Na verdade, como mencionei no capítulo anterior, você pode, mas apenas com a sintaxe do SPSS. Aqueles que não estão com vontade de enfiar a cabeça numa máquina de moer podem ler o arquivo **ContrastsUsingSyntax.pdf** (Contrastes usando a sintaxe) no site www.artmed.com.br. Aqueles que estão com vontade de enfiar cabeça na máquina de moer podem ler o arquivo, também: ele vai ter o mesmo efeito (pelo menos para mim teve)!

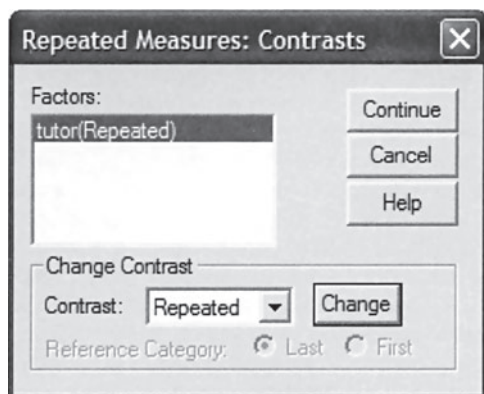


Figura 11.4 Contrastes de medidas repetidas.

11.4.3 Testes *post hoc* e opções adicionais ³



Existem importantes considerações quando pensamos em qual teste *post hoc* iremos utilizar. A violação da esfericidade tem implicações para as comparações múltiplas. Boik (1981) forneceu um importante relato dos efeitos da não-esfericidade nos testes *post hoc* em delineamentos de medidas repetidas e concluiu que mesmo pequenos desvios da esfericidade produzem grande tendenciosidade na estatística *F*. Ele recomenda usar esses testes para contrastes de medidas repetidas. Quando os termos erro experimentais são pequenos, o poder de detectar efeitos relativamente fortes pode ser tão baixo quanto 0,05 (com a esfericidade = 0,80). Boik argumenta que a situação para comparações *múltiplas* não pode ser melhorada e conclui recomendando a análoga multivariada. Mitzel e Games (1981) descobriram que quando a esfericidade não se mantém ($\epsilon < 1$), o erro de conjunto convencionalmente empregado em comparações entre pares resultou em diferenças não-significativas entre duas médias declaradas como significativas (isto é, uma taxa tolerante de erro Tipo I) ou diferenças não-detectadas (uma taxa conservadora de erro Tipo I). Mitzel e Games, portanto, recomendaram o uso separado dos termos erro para cada com-

paração. Maxwell (1980) testou sistematicamente o poder e níveis alfa para cinco testes *post hoc* sob condições de medidas repetidas. Os testes avaliaram a diferença totalmente significativa de Tukey (WSD) que usa o termo erro de conjunto; o procedimento de Tukey, mas com um termo erro separado ou com $(n - 1) gl$ (rotulado de SEP1) ou $(n - 1)(k - 1) gl$ (rotulado de SEP2); o procedimento de Bonferroni (BON); uma abordagem multivariada – o intervalo de confiança simultâneo de Roy-Bose (SCI). Maxwell testou esses procedimentos *a priori* variando o tamanho da amostra, o número de níveis do fator repetido e o afastamento da esfericidade. Ele descobriu que a abordagem multivariada era sempre “muito conservadora para usos práticos” (p. 277) e isso era mais extremo quando n (o número de sujeitos) é pequeno em relação a k (o número das condições). O teste de Tukey inflacionou o valor de alfa de forma inaceitável com afastamentos crescentes da esfericidade até mesmo quando um termo erro separado era usado (SEP1 e SEP2). O método Bonferroni, entretanto, foi extremamente robusto (embora levemente conservador) e controlava bem os níveis de alfa apesar da manipulação. Portanto, em termos de taxas de erro Tipo I, o teste de Bonferroni foi o melhor.

Em termos de poder de teste (a taxa de erro do Tipo II) para uma amostra pequena ($n = 8$), Maxwell descobriu que o WSD é o mais poderoso sob as condições de não-esfericidade, mas essa vantagem era muito reduzida quando $n = 15$. Keselman e Keselman (1988) estenderam o trabalho de Maxwell para deli-



neamentos não-balanceados. Eles também usaram o WSD de Tukey, um WSD modificado (com a variância do erro não-agrupada), a estatística *t* de Bonferroni e uma abordagem multivariada e descobriram que quando médias não-ponderadas são usadas (com delineamentos não-balanceados) nenhum dos quatro testes poderia controlar a taxa de erro do Tipo I. Quando médias ponderadas foram usadas, somente os testes multivariados podiam limitar as taxas





de alfa, embora a estatística t de Bonferroni fosse consideravelmente melhor do que os dois métodos de Tukey. Em termos de poder, Keselman e Keselman concluíram que “à medida que o número dos níveis das medidas repetidas aumenta, BON é substancialmente mais poderoso do que SCI” (p. 223).

Em resumo, quando a esfericidade é violada, o método Bonferroni parece, geralmente, ser a mais robusta das técnicas univariadas, especialmente em termos de poder e controle da taxa do erro do Tipo I. Quando a esfericidade definitivamente não for violada, o teste de Tukey pode ser usado. Em ambos os casos, o procedimento Games-Howell, que utiliza o termo erro de conjunto, é preferível ao teste de Tukey.

Para os leitores que usam as versões do SPSS anteriores à versão 7.0, essa discussão é acadêmica, porque esses leitores irão descobrir que não existe facilidade para produzir os testes *post hoc* para os delineamentos de medidas repetidas nessas versões! Então, por que incluí essa discussão de quais técnicas são as melhores? Bem, para começar, é possível executar a análise novamente como um delineamento entre grupos e usar os procedimentos *post hoc*. Entretanto, como foi dito, muitos procedimentos tem um desempenho ruim com dados relacionados (especialmente se a esfericidade foi violada), assim, eu não recomendo essa abordagem. Entretanto, existem arquivos de sintaxe disponíveis para a execução de testes *post hoc* para medidas repetidas (disponíveis em <http://www.spss.com/tech/macros/>). Dessas macros o método de Dunn-Sidak é provavelmente o melhor porque é menos conservador do que as comparações corrigidas de Bonferroni.

Para os leitores que não tem acesso à internet ou não gostam da janela de sintaxe, podemos aplicar as comparações de Bonferroni usando o procedimento do teste t emparelhado. Realize testes t em todos os pares dos níveis da variável independente e, então, aplique a correção de Bonferroni para a probabilidade na qual você aceita qualquer um desses testes. Essa correção é alcançada dividindo o valor da probabilidade (0,05) pelo número de

testes t realizados. O valor da probabilidade resultante deve ser usado como o critério da significância estatística. Assim, por exemplo, se compararmos todos os níveis da variável independente dos dados do trabalho, faríamos seis comparações ao todo, e o nível de significância apropriado seria $0,05/6 = 0,0083$. Portanto, aceitaríamos testes t que tenham o valor de significância menor do que 0,0083. Uma maneira de recuperar parte do poder desse tipo de procedimento é comparar somente grupos entre os quais você espera que apareçam diferenças (em vez de comparar todos os pares dos níveis de tratamento). Quanto menos testes você executar, menos você terá que corrigir o nível de significância e mais poder você irá manter.

A boa notícia para as pessoas usando o SPSS versão 7.5 ou posteriores é que alguns procedimentos *post hoc* estão disponíveis para medidas repetidas. Entretanto, eles não são acessados pela caixa de diálogo dos testes *post hoc*, mas podem ser encontrados como parte das opções adicionais. Essas opções podem ser acessadas clicando em  (Opções) na caixa de diálogo principal para abrir a caixa de diálogo **medidas repetidas MLG: opções** (veja a Figura 11.5). Para especificar os testes *post hoc*, selecione a variável de medidas repetidas (nesse caso, **tutor** (professor)) no quadro denominado **Estimated Marginal Means: Factor(s) and Factor Interactions** (Médias Marginais Estimadas: Fator(es) e interações de fatores) e transfira-a para o quadro **Display Means** (Exibir Médias) clicando em . Uma vez que a variável foi transferida, a caixa denominada **Compare main effects** (Compare efeitos principais) se torna ativa e você deve selecionar essa opção . Se essa opção é selecionada, a caixa denominada **Confidence Interval adjustments** (Ajustes do intervalo de confiança) se torna ativa e você pode clicar em  para ver a escolha de três níveis de ajustes. O padrão é não ter ajuste e, simplesmente, executar o teste *post hoc* LSD de Tukey (isso não é recomendado). A segunda opção é a correção de Bonferroni (recomendada pelas razões mencionadas acima).

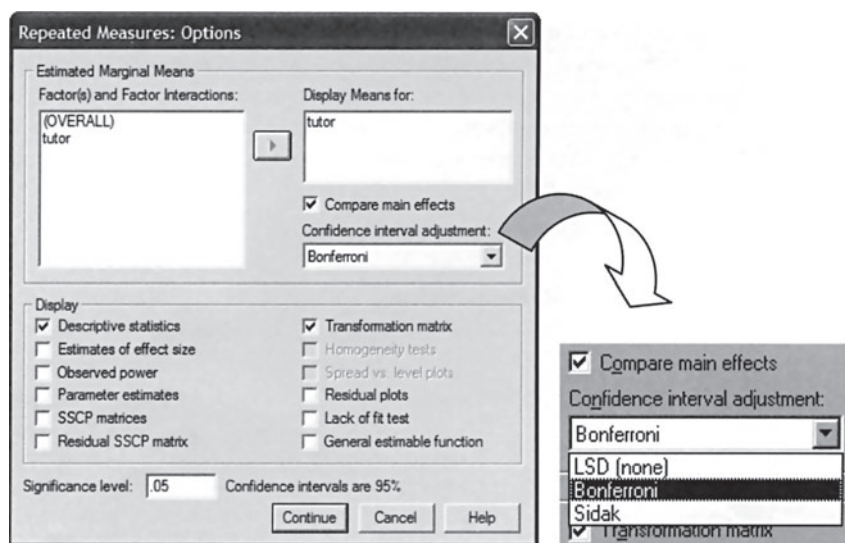


Figura 11.5 Caixa de diálogo *opções*.

e a opção final é a correção Sidak que deve ser selecionada se você está preocupado com a perda de poder associada aos valores corrigidos de Bonferroni.

A caixa de diálogo **options** (Opções) (Figura 11.5) tem, também, outras opções úteis. Você pode pedir por estatísticas descritivas, que irá fornecer as médias, desvios padrão e o número de participantes para cada nível da variável independente. Você também pode solicitar uma matriz de transformação, que fornece os valores de código para qualquer contraste selecionado na caixa de diálogos **contrasts** (contrastes) (Figura 11.4) e é muito útil para a interpretação dos contrastes em delineamentos mais complexos. O SPSS também pode ser solicitado a imprimir as hipóteses, erros e a soma dos quadrados residual e matrizes do produto cruzado (SSCPs) e aprenderemos sobre a importância dessas matrizes no Capítulo 14. A opção para a homogeneidade dos testes da variância será ativada somente quando houver um fator entre grupos (delineamentos mistos – veja o próximo capítulo). Você pode, também, mudar o nível de significância para testar qualquer teste *post hoc*; geralmente, o nível 0,05 é aceitável. Depois de selecionar as

opções de interesse, clique em **Continue** para retornar à caixa de diálogo principal e clique em **OK** para executar a análise.

11.5 SAÍDAS PARA A ANOVA DE MEDIDAS REPETIDAS DE UM FATOR ②

11.5.1 Descritivas e outros diagnósticos ①

A Saída 11.1 do SPSS mostra as estatísticas de diagnóstico iniciais. Em primeiro lugar elas nos informam as variáveis que representam cada nível da variável independente. Esse resultado é útil para verificar que as variáveis foram colocadas na ordem correta. A próxima tabela fornece estatísticas descritivas básicas para os quatro níveis da variável independente. A partir dessa tabela podemos ver que, em média, o Dr. Field deu as notas mais altas para os trabalhos (isso porque eu sou legal, sabe... ou pode ser porque eu sou um estúpido e, assim, meus padrões acadêmicos são muito baixos!). O Dr. Morte, por outro lado, deu notas muito baixas. Esses valores médios são úteis para interpretar qualquer efeito que possa surgir da análise principal.

Saída 11.1 do SPSS

Within-Subjects Factors
(Fatores entre participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

TUTOR (Professor)	Dependent Variable (Variável Dependente)
1	TUTOR1
2	TUTOR2
3	TUTOR3
4	TUTOR4

Descriptive Statistics (Estatísticas Descritivas)

	Mean (Média)	Std. Deviation (Desvio Padrão)	N
Dr. Field (Dr. Field)	68.8750	5.6426	8
Dr. Smith (Dr. Ferreira)	64.2500	4.7132	8
Dr. Scrote (Dr. Scrots)	65.2500	6.9230	8
Dr. Death (Dr. Morte)	57.3750	7.9091	8

11.5.2 Interpretando e corrigindo a esfericidade ②

Na Seção 11.2.3, foi afirmado que o SPSS testa se os dados violam a hipótese de esfericidade. A próxima parte da saída contém informações sobre esse teste. O teste de Mauchly não deve ser significativo se assumirmos que a condição de esfericidade foi satisfeita. A Saída 11.2 do SPSS mostra o teste de Mauchly para as notas dos professores e a coluna importante é a que contém o valor da significância. O valor de significância (0,043) é menor do que o valor crítico de 0,05, assim, aceitamos que as variâncias das diferenças entre os níveis são significativamente diferentes. Em outras palavras, a hipótese de esfericidade foi violada. Sabendo que violamos essa hipótese a pergunta pertinente é: como devemos proceder?

Descobrimos, na Seção 11.2.5, que o SPSS produz três correções baseadas nas estimativas de esfericidade apoiado por Greenhouse e

Geisser (1959) e Huynh e Feldt (1976). Ambas estimativas fornecem um fator de correção que é aplicado aos graus de liberdade usados para determinar a razão *F* observada. A correção de *Greenhouse* e *Geiser* varia entre $1/(k - 1)$ (onde *k* é o número de condições de medidas repetidas) e 1. Quanto mais próximo esse $\hat{\epsilon}$ está de 1,00, mais homogêneas serão as diferenças das variâncias e, portanto, mais próximos os dados estão de serem esféricos. Numa situação em que existem quatro condições (como nos nossos dados), o limite mais baixo de $\hat{\epsilon}$ será $1/(4 - 1)$ ou 0,33 (conhecido como estimativa limite inferior (*lower-bound*) da esfericidade). A Saída do SPSS 11.2 mostra que o valor calculado de $\hat{\epsilon}$ é 0,558. Isso é mais próximo ao limite mais baixo de 0,33 do que o limite superior de 1 e ele, portanto, representa um desvio substancial da esfericidade. Veremos como esses valores são usados na próxima seção.

Saída do SPSS 11.2

Mauchly's Test of Sphericity^a (Teste de Mauchly da Esfericidade)

Measure: Measure_1 (Medida: Medida_1)

Within-Subjects Effect (Efeito dentre participantes)	Mauchly's W (W de Mauchly)	Approx. Chi-Square (Qui-Quadrado Aproximado)	df (gl)	Sig. (Sig.)	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound (Limite Inferior)
TUTOR (Professor)	0.131	11.628	5	0.043	0.558	0.712	0.333

Testa a hipótese nula de que a matriz de covariância dos erros das variáveis dependentes transformadas e ortonormalizadas é proporcional a uma matriz identidade.

a Delineamento: Intercepto dentre participantes: TUTOR

b Pode ser utilizado para ajustar os graus de liberdade dos testes de significância para médias. Os testes corrigidos são mostrados em camadas (por padrão) na tabela dos testes de efeitos dentre participantes.

11.5.3 A ANOVA principal ②

A Saída 11.3 do SPSS mostra os resultados da ANOVA para a variável entre participantes. Essa tabela pode ser lida da mesma maneira que a da ANOVA de um fator entre grupos (veja o Capítulo 8). Existe uma soma dos quadrados para o efeito de medidas repetidas **tutor** (professor) que nos diz quanto da variabilidade total é explicada pelo efeito experimental. Note que o valor é 554,125, a soma dos quadrados do modelo(SS_M) calculado na seção 11.3.3. Existe também um termo erro que é a variação não-explicada entre as condições da variável de medidas repetidas. Essa é a soma dos quadrados residual (SS_R) que foi calculada na Seção 11.3.4; note que o valor é 1048,375 (que é o mesmo valor que foi calculado). Como expliquei anteriormente, essas somas dos quadrados são convertidas em médias dos quadrados pela divisão

pelos graus de liberdade. Como vimos antes, o gl para o efeito **professor** é simplesmente $k - 1$, onde k é o número de níveis da variável independente. O gl do erro é $(n - 1)(k - 1)$, onde n é o número de participantes (ou, nesse caso, o número de trabalhos) e k é como antes. A razão F é obtida dividindo a média dos quadrados para o efeito experimental (184,708) pelas do erro (49,923). Assim como com a ANOVA entre grupos, essa estatística representa a razão entre a variância sistemática e a variância não-sistemática. O valor de F ($184,71/49,92 = 3,70$) é, então, comparado com o valor crítico para 3 e 21 graus de liberdade. O SPSS mostra o valor da significância exata para a razão F . A significância de F é 0,028, o que é significativa, porque é menor do que o valor crítico de 0,05. Podemos, portanto, concluir que existe uma diferença significativa entre as notas dadas pelos quatro professores. Entretanto, esse teste

Saída 11.3 do SPSS ANOVA de medidas repetidas para as versões 8.0 e posteriores (tabela de cima) e versões 7.0 e 7.5 (tabela de baixo)

Tests of Within-Subjects Effects (Testes de efeitos entre participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)		Type III Sum of Squares (Soma dos Quadrados do Tipo III)	df (gl)	Mean Square (Média dos Quadrados)	F	Sig.
TUTOR	Sphericity Assumed (Esfericidade Assumida)	554.125	3	184.708	3.700	0.028
	Greenhouse-Geisser	554.125	1.673	331.245	3.700	0.063
	Huynh-Feldt	554.125	2.137	259.329	3.700	0.047
	Lower-bound (Limite-inferior)	554.125	1.000	554.125	3.700	0.096
Error(TUTOR) (Tutor(Erro))	Sphericity Assumed (Esfericidade Assumida)	1048.375	21	49.923		
	Greenhouse-Geisser	1048.375	11.710	89.528		
	Huynh-Feldt	1048.375	14.957	70.091		
	Lower-bound (Limite-inferior)	1048.375	7.000	149.768		

a. Calculado utilizando alfa = 0,05.

Tests of Within-Subjects Effects (Testes de efeitos entre participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Sphericity Assumed (Esfericidade Assumida)

Source (Fonte)	Type III Sum of Squares (Soma dos Quadrados do Tipo III)	df (gl)	Mean Square (Média dos Quadrados)	F	Sig.
TUTOR	564.125	3	184.708	3.700	0.028
Error(TUTOR) (Tutor(Erro))	1048.375	21	49.923		

a. Calculado utilizando alfa = 0,05.

principal não nos informa quais professores diferiram entre si nas suas notas.

Embora esse resultado seja bem plausível, aprendemos que as violações da hipótese da esfericidade tornam o teste F impreciso. Sabemos pela Saída 11.2 do SPSS que esses dados não eram esféricos e, assim, precisamos levar em consideração essa violação. A tabela na Saída 11.3 do SPSS mostra a razão F e os graus de liberdade associados quando a esfericidade é assumida e a estatística F significativa indica diferença(s) entre as notas dadas pelos quatro professores. Em versões do SPSS posteriores a 8, essa tabela também contém linhas adicionais dando os valores corrigidos de F e para os três diferentes tipos de ajustamentos (Greenhouse-Geisser, Huynh-Feldt e limite inferior). Nas versões 7.0 e 7.5 você deve ajustar a tabela da ANOVA para ver esses valores corrigidos. Embora muitas pessoas não tenham essas versões, vale a pena notar que se você usar o mouse para clicar duas vezes na tabela principal da ANOVA, a tabela irá abrir para a edição e um novo conjunto de rótulos de menu deve aparecer no alto da janela. Um desses menus é denominado **Pivot**. Clique nesse menu e selecione a opção denominada **Move Layers to Rows** (Mova Camadas para Linhas). Os valores de F corrigidos devem aparecer.

A Saída 11.4 do SPSS mostra a tabela expandida da ANOVA com os valores corrigidos

para cada uma das três estimativas de esfericidade e ela parece com a tabela das versões mais recentes do SPSS (como na Saída 11.3 do SPSS). Note que em todos os casos as razões F permanecem as mesmas; são os graus de liberdade que mudam (e, portanto, o valor crítico com o qual a estatística F é comparada). Os graus de liberdade foram ajustados usando as estimativas de esfericidade calculadas na Saída 11.2 do SPSS. Os ajustes são feitos multiplicando os graus de liberdade pela estimativa da esfericidade (veja Field, 1998a).² Os novos graus de liberdade são, então, usados para averiguar a significância de F . Para esses dados, as correções resultam no F observado sendo não-significativo quando usada a correção de Greenhouse e Geisser (porque $p > 0,05$). Entretanto, foi notado anteriormente que essa correção é bem conservadora e, assim, pode perder efeitos que genuinamente existem. Ela é, portanto, útil para consultar a estatística do F corrigido de Huynh-Feldt. Usando essa correção, o valor do F continua

² Por exemplo, a estimativa de esfericidade de Greenhouse e Geisser era 0,558. Os graus de liberdade originais para o modelo eram 3; esse valor é corrigido multiplicando pela estimativa da esfericidade ($3 \times 0,558 = 1,674$). Da mesma forma, o gl do erro era 21; esse valor é corrigido da mesma maneira ($21 \times 0,558 = 11,718$). A razão F é, então, testada contra o valor crítico com esses novos graus de liberdade (1,674, 11,718). As outras correções são aplicadas da mesma maneira.

Saída 11.4 do SPSS

Tests of Within-Subjects Effects (Testes de efeitos dentre participantes)

<i>Measure</i> (Medida)		<i>Source</i> (Fonte)	<i>Type III Sum of Squares</i> (Soma dos Quadrados do Tipo III)	<i>df</i> (gl)	<i>Mean Square</i> (Média dos Quadrados)	<i>F</i>	<i>Sig.</i>
<i>MEASURE_1</i> (Medida_1)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	TUTOR Error(TUTOR)	554.125 1048.375	3 21	184.708 49.923	3.700	0.028
	Greenhouse-Geisser	TUTOR Error(TUTOR)	554.125 1048.375	1.673 11.710	331.245 89.528	3.700	0.063
	Huynh-Feldt	TUTOR Error(TUTOR)	554.125 1048.375	2.137 14.957	259.329 70.091	3.700	0.047
	<i>Lower-bound</i> (Limite inferior)	TUTOR Error(TUTOR)	554.125 1048.375	1.000 7.000	554.125 149.768	3.700	0.096

a. Calculado utilizando alfa = 0,05.

significativo porque o valor de probabilidade de 0,047 está abaixo do valor crítico de 0,05. Assim, para essa correção podemos aceitar a hipótese de que os professores diferem nas suas notas. Entretanto, foi também notado que essa correção é bem liberal, assim, tende a aceitar valores como significativos quando, na realidade, eles não são significativos.

Isso nos deixa num dilema desconcertante se aceitamos ou não essa estatística *F* como significativa. Mencionei anteriormente que Stevens (1992) recomenda tirar a média das duas estimativas e, certamente, quando duas correções dão resultados diferentes (como nesse caso), esse é um conselho sábio. Se as duas correções dão origem à mesma conclusão, faz pouca diferença qual delas você escolhe para relatar (embora se você aceitar a estatística *F* como significativa é melhor relatar a estimativa conservadora de Greenhouse-Geisser para evitar críticas!). Embora seja fácil calcular a média dos dois fatores de correção e corrigir os graus de liberdade de modo apropriado, não é tão fácil calcular a probabilidade exata para esses graus de liberdade. Portanto, se você tiver que enfrentar essa situação (que, para ser honesto, é muito improvável), recomendo tirar a média dos dois valores de significância para ter uma ideia aproximada da correção que está dando a resposta mais precisa. Nesse caso, a média dos dois valores de *p* é $(0,063 + 0,047) / 2 = 0,055$. Portanto, provavelmente devemos usar a correção de Greenhouse e Geisser e concluir que a razão *F* não é significativa.

Os dados ilustram a importância de usar um valor crítico válido de *F*: ele pode significar

a diferença entre um resultado estatisticamente significativo e um resultado não-significativo. Mais importante ainda, ele pode significar a diferença entre fazer um erro do Tipo I ou não. Se não tivéssemos usado as correções para a esfericidade teríamos concluído, erroneamente, que os professores deram notas significativamente diferentes. Entretanto, devo quantificar essa afirmação dizendo que esse exemplo também destaca quão arbitrário é usar um nível de significância de 0,05. Essas duas correções produzem valores de significância ligeiramente menor ou maior do que 0,05 e, mesmo assim, eles nos levam a conclusões completamente opostas! Assim, devemos observar o tamanho de efeito para ver se ele é substantivo apesar da sua significância.

Vimos também anteriormente que a opção final quando você tem dados que violam a esfericidade é usar estatísticas teste multivariadas porque elas não fazem essa hipótese (veja O'Brien e Kaiser, 1985). A Saída 11.5 do SPSS mostra as estatísticas teste multivariadas para esse exemplo (detalhes dessas estatísticas teste podem ser encontrados na Seção 14.3.4). A coluna exibindo os valores de significância claramente mostra que os testes multivariados não são significativos (porque *p* é 0,063, o que é maior do que o valor crítico de 0,05). Tendo em mente a perda do poder nesses testes (veja a Seção 11.2.5), esse resultado suporta a decisão de aceitar a hipótese nula e conclui que não há diferenças significativas entre as pontuações dadas por diferentes professores. A interpretação desses resultados deveria parar agora porque os efeitos principais não são significativos.

Saída do SPSS 11.5

Multivariate Tests^a (Testes Multivariados)

Effect (Efeito)		Value (Valor)	<i>F</i>	Hypothesis df (gl da Hipótese)	Error df (gl do Erro)	Sig. (Sig.)
TUTOR (Professor)	Pillai's Trace (Traço de Pillai)	0.741	4.760 ^a	3.000	5.000	0.063
	Wilks' Lambda (Lambda de Wilks)	0.259	4.760 ^a	3.000	5.000	0.063
	Hotelling's Trace (Traço de Hotelling)	2.856	4.760 ^a	3.000	5.000	0.063
	Roy's Largest Root (Maior Raiz de Roy)	2.856	4.760 ^a	3.000	5.000	0.063

a. Delineamento: Intercepto Dentre Participantes: Professor.
b. Calculado utilizando alfa = 0,05.
c. Estatística exata.

Entretanto, iremos observar a saída para os contrastes para ilustrar como esses testes estão apresentados no visualizador do SPSS.

11.5.4 Contrastes ②

A matriz de transformação solicitada nas opções é apresentada na Saída 11.6 do SPSS e precisamos lembrar como esquematizamos os contrastes (veja o Capítulo 8) para interpretar essa tabela. Primeiro, deve-se lembrar que um código 0 significa que o grupo não está incluído num contraste. Portanto, contraste 1 (denominado **Level 1 vs. Level 2** (Nível 1 vs. Nível 2)) ignora o Dr. Scrote e o Dr. Morte. A próxima coisa a ser lembrada é que os grupos com uma ponderação negativa são comparados aos grupos com uma ponderação positiva. Nesse caso, significa que o primeiro contraste compara o Dr. Field com o Dr. Ferreira. Usando a mesma lógica, o contraste 2 (denominado **Level 2 vs. Level 3** (Nível 2 vs. Nível 3)) ignora o Dr. Field e o Dr. Morte e compara o Dr. Ferreira com o Dr. Scrote. Finalmente, o contraste três (**Level 3 vs. Level 4** (Nível 3 vs. Nível 4)) compara o Dr. Morte com o Dr. Scrote. Esse padrão de contrastes é consistente com o que esperamos conseguir dos contrastes repetidos (isto é, todos os grupos, com exceção do primeiro, são

Saída do SPSS 11.6

TUTOR^a (Professor)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Dependent Variable (Variável Dependente)	TUTOR (Professor)		
	Level 1 vs. Level 2 (Nível 1 vs. Nível 2)	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 3 vs. Level 4 (Nível 3 vs. Nível 4)
Dr. Field (Dr. Field)	1	0	0
Dr. Smith (Dr. Ferreira)	-1	1	0
Dr. Scrote (Dr. Scrote)	0	-1	1
Dr. Death (Dr. Morte)	0	0	-1

a. Os contrastes para os fatores dentre participantes são: Professor: Contraste repetido.

comparados com a categoria precedente). A matriz de transformação, que aparece na parte de baixo da saída, é usada principalmente para confirmar o que cada contraste representa.

Acima da matriz de transformação, devemos encontrar uma tabela resumo dos contrastes (Saída 11.7 do SPSS). Cada contraste está listado por sua vez e, como com os contrastes entre grupos, um teste *F* é realizado comparando as duas partes da variação. Assim, observando os valores de significância da tabela, podemos di-

Saída 11.7 do SPSS

Tests of Within-Subjects Contrasts (Testes dos contrastes dentre participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)	TUTOR (professor)	Type III Sum of Squares (Soma dos Quadrados do Tipo III)	df (gl)	Mean Square (Média dos Quadrados)	F	Sig.
TUTOR	Level 1 vs. Level 2 (Nível 1 vs. Nível 2)	171.125	1	171.125	18.184	0.004
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	8.000	1	8.000	0.152	0.708
	Level 3 vs. Level 4 (Nível 3 vs. Nível 4)	496.125	1	496.125	3.436	0.106
Error(TUTOR)	Level 1 vs. Level 2 (Nível 1 vs. Nível 2)	65.875	7	9.411		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	368.000	7	52.571		
	Level 3 vs. Level 4 (Nível 3 vs. Nível 4)	1010.875	7	144.411		

zer o Dr. Field atribuiu notas significativamente mais altas do que o Dr. Ferreira (**Level 1 vs. Level 2**), e as notas do Dr. Scrote foram próximos às do Dr. Morte (**Level 3 vs. Level 4**). Entretanto, o contraste significativo deve ser ignorado por causa do efeito principal não-significativo (lembre-se de que os dados não obedecem à esfericidade). Um ponto importante a ser notado é que a esfericidade nos nossos dados nos levou a algumas questões importantes sobre a correção dos fatores e sobre aplicação de algum discernimento nos seus dados (é reconfortante saber que o computador não tem todas as respostas – mas isso significa que temos que, realmente, saber algumas das respostas). Nesse exemplo, teríamos que concluir que existem diferenças não-significativas entre as notas dadas por diferentes professores. Entretanto, a ambiguidade dos nossos dados pode nos fazer considerar a execução de um estudo similar com um número maior de trabalhos sendo avaliados.

11.5.5 Testes *post hoc* ②

Se você selecionou os testes *post hoc* para a variável de medidas repetidas na caixa de diálogo **options** (opções) (veja a Seção

11.4.3), a tabela Saída 11.8 do SPSS será apresentada na janela visualizadora.

A disposição da tabela na Saída 11.8 do SPSS é similar a da tabela produzida para os testes *post hoc* entre grupos: as médias das diferenças entre grupos são exibidas assim como o erro padrão, o valor de significância e o intervalo de confiança para a diferença entre as médias. Observando os valores de significância podemos ver que a única diferença entre as médias do grupo é entre o Dr. Field e o Dr. Ferreira. Observando as médias desses grupos (Saída 11.1 do SPSS), podemos ver que eu dou notas mais altas do que o Dr. Ferreira. Entretanto, existe um resultado um tanto anômalo nisso, pois não existe diferença significativa entre as notas dadas pelo Dr. Morte e as minhas, embora a diferença da média entre as pontuações seja mais alta (11,5) do que a diferença entre mim e o Dr. Ferreira (4,6). A razão desse resultado é a esfericidade nos dados. Você pode considerar a execução de algumas correlações entre as notas dos quatro professores. Você irá descobrir que existe uma correlação positiva bem alta entre as pontuações dadas pelo Dr. Ferreira e eu (indicando

Saída 11.8 do SPSS

Pairwise Comparisons (Comparações aos pares)

Measure: MEASURE_1 (Medida: MEDIDA_1)

(I) TUTOR	(j) TUTOR	Mean Difference (I – J) (Diferença Média)	Std. Error (Erro Padrão)	Sig. ^a	95% Confidence Interval for Difference (Intervalo de 95% de Confiança para a Diferença)	
					Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	2	4.625*	1.085	0.022	0.682	8.568
	3	3.625	2.841	1.000	-6.703	13.953
	4	11.500	4.675	0.261	-5.498	28.498
2	1	-4.625*	1.085	0.022	-8.568	-0.682
	3	-1.000	2.563	1.000	-10.320	8.320
	4	6.875	4.377	0.961	-9.039	22.789
3	1	-3.625	2.841	1.000	-13.953	6.703
	2	1.000	2.563	1.000	-8.320	10.320
	4	7.875	4.249	0.637	-7.572	23.322
4	1	-11.500	4.675	0.261	-28.498	5.498
	2	-6.875	4.377	0.961	-22.789	9.039
	3	-7.875	4.249	0.637	-23.322	7.572

Baseando nas médias marginais estimadas.
* A diferença média é significativa ao nível de 0,05.
a. Ajuste para comparações múltiplas: Bonferroni.

Dica da Samanta Ferrinho



- A ANOVA de medidas repetidas de um fator compara várias médias quando essas médias vieram dos mesmos participantes; por exemplo, se você mensurou a habilidade estatística das pessoas a cada mês durante um ano.
- Na ANOVA de medidas repetidas, há uma hipótese adicional: a esfericidade. Ela precisa ser considerada somente quando você tem três ou mais condições de medidas repetidas. Teste a esfericidade usando o teste de Mauchly. Encontre a tabela com essa denominação: se o valor na coluna denominada Sig. é menor do que 0,05, a hipótese foi violada. Se a significância do teste de Mauchly é maior do que 0,05, a hipótese de esfericidade foi confirmada.
- A tabela denominada *Tests of Within-Subjects Effects* (Testes dos Efeitos dentre Sujeitos) mostra o resultado principal para nossa ANOVA. Se a hipótese de esfericidade foi confirmada, observe a linha denominada *Sphericity Assumed* (Esfericidade Assumida). Se a hipótese foi violada, leia a linha denominada *Greenhouse-Geisser* (você pode também olhar em Huynh-Feldt, mas você deve ler este capítulo para descobrir as vantagens dos dois procedimentos). Tendo selecionado a linha apropriada, observe a coluna denominada Sig.: se o valor for menor do que 0,05, as médias dos grupos são significativamente diferentes.
- Para contrastes e testes *post hoc*, novamente, observe as colunas denominadas Sig. para descobrir se suas comparações são significantes (elas serão se o valor de significância for menor do que 0,05).

um nível baixo de variabilidade nos nossos dados). Contudo, existe uma correlação baixa entre as notas dadas pelo Dr. Morte e por mim (indicando um alto nível de variabilidade entre nossas pontuações). É essa grande variabilidade entre o Dr. Morte e eu que produziu o resultado não-significativo, apesar da média das notas ser muito diferente (essa observação é também evidente a partir dos erros padrão).

11.6 TAMANHO DE EFEITO PARA ANOVA DE MEDIDAS REPETIDAS ③



Como com a ANOVA independente, a melhor medida para o tamanho de efeito total é o ômega ao quadrado (ω^2).

Contudo, só para tornar a vida mais complicada do que já é, as equações que anteriormente usamos para o ômega ao quadrado não podem ser usadas para os dados de medidas repetidas! Mas se você realmente usar as mesmas equações nos dados de medidas repetidas irá superestimar leve-

mente o tamanho de efeito. Para simplificar, algumas pessoas usam a mesma equação para a ANOVA de um fator independente e para a de medidas repetidas (e eu sou culpado disso em outro livro), mas acho que neste livro nós não vamos optar pela simplicidade, mas vamos agarrar a complexidade como se fosse nosso irmão gêmeo separado de nós por muitos anos.

Na ANOVA de medidas repetidas, a equação para ômega ao quadrado é (segurem-se firme...):

$$\omega^2 = \frac{\left[\frac{k-1}{nk} (MS_M - MS_R) \right]}{MS_R + \frac{MS_{BG} - MS_R}{k} + \left[\frac{k-1}{nk} (MS_M - MS_R) \right]} \quad (11.1)$$

OK, isso parece infernal, e é. Mas vamos examiná-la. Primeiro, existem algumas médias dos quadrados que já vimos (e calculamos) antes. Existe o quadrado médio para o modelo (MS_M) e quadrado médio residual (MS_R) e ambos podem ser obtidos da tabela da ANOVA que o SPSS produz (Saída 11.3 do SPSS). Tem

também o k , o número de condições no experimento, que para esses dados seria 4 (são quatro professores) e há n , o número de participantes (nesse caso, o número de trabalhos, 8). O problema principal é o termo MS_{BG} . No começo da Seção 11.3 (Figura 11.1), mencionei que a variação total é dividida em uma variação entre participantes e uma variação entre grupos. Em todos os cálculos subsequentes, a variação entre grupos ficou um pouco esquecida (porque nós, na verdade, não precisamos calcular a razão F). Agora ela aparece na equação do ômega ao quadrado. A maneira mais fácil de calcular esse termo é pela subtração, porque sabemos pela Figura 11.1 que:

$$SS_T = SS_{BG} + SS_W$$

Agora, o SPSS não nos fornece o SS_W na saída, mas nós sabemos que ele é formado pelo SS_M e pelo SS_R que foram dados. Substituindo esses termos e reorganizando essa equação, temos:

$$SS_T = SS_{BG} + SS_M + SS_R$$

$$SS_{BG} = SS_T - SS_M - SS_R$$

O próximo problema é que o SPSS, que claramente está tentando nos atrapalhar a cada passo, não nos fornece o SS_T e eu receio (a não ser que eu não o tenha visto) que você irá precisar calcular manualmente (veja a Seção 11.3.1). Dos valores que calculamos anteriormente, temos:

$$\begin{aligned} SS_{BG} &= 1705,868 - 554,125 - 1048,375 \\ &= 103,37 \end{aligned}$$

O próximo passo é converter isso em um quadrado médio dividindo os graus de liberdade que, nesse caso, são os números de pessoas no experimento menos 1 ($n - 1$):

$$\begin{aligned} MS_{BG} &= \frac{SS_{BG}}{gl_{BG}} = \frac{SS_{BG}}{n - 1} \\ &= \frac{103,37}{8 - 1} \\ &= 14,77 \end{aligned}$$

Tendo feito tudo isso, e provavelmente morto de tédio nesse processo, devemos agora ressurgir com renovado vigor para a equação do tamanho de efeito, que fica:

$$\begin{aligned} \omega^2 &= \frac{\left[\frac{4-1}{8 \times 4} (184,71 - 49,92) \right]}{49,92 + \frac{14,77 - 49,92}{4} + \left[\frac{4-1}{8 \times 4} (184,71 - 49,92) \right]} \\ &= \frac{12,64}{53,77} \\ &= 0,24 \end{aligned}$$

Assim, temos um ômega ao quadrado de 0,24. Se você calculá-lo da mesma maneira que a ANOVA independente (veja a Seção 8.5), deve ter uma resposta ligeiramente maior (0,25, na verdade).



Mencionei em outras oportunidades que é mais útil ter medidas do tamanho de efeito para comparações focadas (em vez da ANOVA principal) e, assim, uma abordagem um pouco mais fácil para calcular os tamanhos de efeito é calculá-los para o contraste que fizemos (veja a Saída 11.7 do SPSS). Para isso, podemos usar a equação que vimos antes para converter os valores de F (porque todos eles têm 1 grau de liberdade para o modelo) para r :

$$r = \sqrt{\frac{F(1, gl_R)}{F(1, gl_R) + gl_R}}$$

Para as três comparações que fizemos, vamos ter:

$$r_{\text{Field vs. Ferreira}} = \sqrt{\frac{18,18}{18,18 + 7}} = 0,85$$

$$r_{\text{Ferreira vs. Scrote}} = \sqrt{\frac{0,15}{0,15 + 7}} = 0,14$$

$$r_{\text{Scrote vs. Morte}} = \sqrt{\frac{3,44}{3,44 + 7}} = 0,57$$

Portanto, as diferenças entre os Drs. Field e Ferreira, e Scrote e Morte eram, ambas, efeitos grandes, mas as diferenças entre os Drs. Ferreira e Scrote eram pequenas.

11.7 RELATANDO A ANOVA DE MEDIDAS REPETIDAS COM UM FATOR ②

Quando relatamos a ANOVA de medidas repetidas, apresentamos os mesmos detalhes do que com a ANOVA independente. O único fator adicional que devemos nos preocupar é relatar os graus de liberdade corrigidos se a esfericidade foi violada. Pessoalmente, eu também gosto de relatar os resultados dos testes da esfericidade. Assim como a ANOVA independente, os graus de liberdade usados para avaliar a razão F são os graus de liberdade para o efeito do modelo ($gl_M = 1,67$) e os graus de liberdade para o resíduo do modelo ($gl_R = 11,71$). Lembre que, nesse exemplo, nós corrigimos ambos usando a correção das estimativas da esfericidade de Greenhouse e Geisser. Portanto, podemos relatar as principais descobertas como:

- ✓ Os resultados mostram que a nota de um trabalho não foi significativamente afetada pelo professor que a deu $F(1,67, 11,71) = 3,70, p > 0,05$.

Se você preferir escolher relatar o teste de esfericidade, você deve relatar a aproximação do qui-quadrado, seus graus de liberdade e o seu valor de significância. É bom, também, relatar o grau de esfericidade com o valor do ϵ -silon. Iremos, igualmente, relatar o tamanho de efeito na versão melhorada:

- ✓ O teste de Mauchly indicou que a hipótese de esfericidade foi violada ($\chi^2(5) = 11,63, p < 0,05$); portanto, os graus de liberdade foram corrigidos usando as estimativas de esfericidade de Greenhouse e Geisser ($\epsilon = 0,56$). Os resultados mostram que a nota de um trabalho não era significativamente afetada pelo professor que a deu $F(1,67, 11,71) = 3,70, p > 0,05, \omega^2 = 0,24$.

Lembre que em virtude de a ANOVA não ser significativa, não devemos relatar qualquer análise posterior.

11.8 MEDIDAS REPETIDAS COM VÁRIAS VARIÁVEIS INDEPENDENTES ②

Já vimos que delineamentos simples entre grupos podem ser estendidos para incorporar uma segunda (ou terceira) variável independente. É igualmente fácil incorporar uma segunda, terceira ou até mesmo uma quarta variável independente em uma análise de medidas repetidas. Como um exemplo, alguns cientistas sociais foram solicitados a pesquisar se as imagens poderiam influenciar atitudes públicas em relação ao álcool. Existem evidências de que atitudes em relação a estímulos podem ser mudadas usando imagens positivas e negativas (por exemplo, Stuart, Shimp e Engle, 1987; mas veja Field e Davey, 1999) e esses pesquisadores estavam interessados em responder duas perguntas. Por um lado, o governo os financiou para observar se imagens negativas em propagandas poderiam ser usadas para mudar atitudes em relação ao álcool. Por outro lado, uma companhia de álcool patrocinou uma pesquisa para ver se imagens positivas poderiam ser usadas para melhorar as atitudes em relação ao álcool. Os cientistas projetaram um estudo para abranger ambas as questões. A Tabela 11.3 ilustra o delineamento experimental e contém os dados para esse exemplo (cada linha representa um único participante).

Os participantes assistiram um total de nove anúncios fictícios em três sessões. Em uma sessão eles assistiram três anúncios: (1) uma marca de cerveja (Brain Death) apresentada com uma imagem negativa (um cadáver com o *slogan* “beber Brain Death faz o seu fígado explodir”); (2) uma marca de vinho (Dangleberry) apresentada num contexto de uma imagem positiva (um homem ou uma mulher *sexy* nus – dependendo do gênero do participante – e o *slogan* “beber vinho Dangleberry o torna um garanhão tesudo”) e (3) uma marca de água (Puritan) apresentada com uma imagem neutra (uma pessoa assistindo televisão acompanhada pelo *slogan* “beber a água Puritan faz você ter um comportamento completamente normal”). Numa segunda

sessão (uma semana depois), os participantes viram as mesmas três marcas, mas dessa vez Brain Death estava acompanhada pela imagem positiva, Dangleberry pela imagem neutra e Puritan pela imagem negativa. Numa terceira sessão, os participantes viram Brain Death acompanhada pela imagem neutra, Dangleberry pela imagem negativa e Puritan pela positiva. Após cada anúncio foi pedido aos participantes para avaliar as bebidas numa escala variando de – 100 (detesto) passando por 0 (neutro) a 100 (gosto muito). A ordem dos comerciais foi aleatória assim como a ordem na qual as pessoas participaram das três sessões. Esse delineamento é bem complexo. Existem duas variáveis independentes: o tipo de bebida (cerveja, vinho ou água) e o tipo de imagem usado (positiva, negativa ou neutra). Essas duas variáveis se cruzam completamente produzindo nove condições experimentais.

11.8.1 A análise principal ②

Para entrar com esses dados no SPSS, precisamos recapitular a regra básica do editor de dados, que declara que cada linha representa os dados de um único participante. Se a pessoa participa em todas as condições experimentais (nesse caso eles assistem todos os tipos de estímulos com todos os tipos de imagens), cada condição experimental deve ser representada por uma coluna no editor de dados. Nesse experimento existem nove condições experimentais, assim, os dados precisam entrar em nove colunas (o formato é idêntico ao da Tabela 11.3). Você deve criar as nove variáveis seguintes no editor de dados com os nomes como foi dado. Para cada uma você deve, também, entrar com uma variável com um nome completo (veja a Seção 2.4.2) para clareza na saída.

Tabela 11.3 Dados de **Attitude.sav**

Bebida Imagem	Cerveja			Vinho			Água		
	vê+	vê–	Neutro	vê+	vê–	Neutro	vê+	vê–	Neutro
Homem	1	6	5	38	–5	4	10	–14	–2
	43	30	8	20	–12	4	9	–10	–13
	15	15	12	20	–15	6	6	–16	1
	40	30	19	28	–4	0	20	–10	2
	8	12	8	11	–2	6	27	5	–5
	17	17	15	17	–6	6	9	–6	–13
	30	21	21	15	–2	16	19	–20	3
	34	23	28	27	–7	7	12	–12	2
	34	20	26	24	–10	12	12	–9	4
	26	27	27	23	–15	14	21	–6	0
Mulher	1	–19	–10	28	–13	13	33	–2	9
	7	–18	6	26	–16	19	23	–17	5
	22	–8	4	34	–23	14	21	–19	0
	30	–6	3	32	–22	21	17	–11	4
	40	–6	0	24	–9	19	15	–10	2
	15	–9	4	29	–18	7	13	–17	8
	20	–17	9	30	–17	12	16	–4	10
	9	–12	–5	24	–15	18	17	–4	8
	14	–11	7	34	–14	20	19	–1	12
	15	–6	13	23	–15	15	29	–1	10

beerpos	Cerveja	+	Pessoa sexy
beerneg	Cerveja	+	Cadáver
beerneut	Cerveja	+	Pessoa sentada
winepos	Vinho	+	Pessoa sexy
wineneg	Vinho	+	Cadáver
wine neut	Vinho	+	Pessoa sentada
waterpos	Água	+	Pessoa sexy
waterneg	Água	+	Cadáver
waterneu	Água	+	Pessoa sentada

Depois que essas variáveis foram criadas, entre com os dados como na Tabela 11.3. Se você tiver problemas com a entrada dos dados, use o arquivo **Attitude.sav**. Para acessar a caixa de diálogo **Define factors** (Definir fatores) use o menu **Analyze⇒General Linear Model⇒Repeated Measures...** (Análise⇒Modelo Linear Geral⇒Medidas Repetidas...). Na caixa de diálogo *Define factors*, você é solicitado a fornecer um nome para a variável entre sujeitos (medidas repetidas). Nesse caso, existem dois fatores dentre sujeitos: **drink** (bebida) (cerveja, vinho ou água) e **imagery** (imagem) (positiva, negativa e neutra). Troque a palavra **factor 1** pela palavra **drink** (bebida). Quando você tiver dado um nome a esse fator de medidas repetidas, deve dizer ao computador quantos níveis a variável apresenta. Nesse caso, há três tipos de bebidas, assim, temos que entrar com o número três no espaço denomi-

nado **Number of Levels** (Número de níveis). Clique em **Add** (Adicionar) para acrescentar essa variável à lista de variáveis de medidas repetidas. Essa variável irá agora aparecer no quadro em branco na parte de baixo da caixa de diálogo e aparecerá como **drink** (bebida) (3). Temos, agora, que repetir esse processo para a segunda variável independente. Entre com a palavra **imagery** (imagem) no espaço denominado **Within-Subject Factor Name** (Nome do fator entre sujeitos) e, então, porque há três níveis dessa variável, entre com o número 3 no espaço denominado **Number of Levels** (Número de Níveis). Clique em **Add** (Adicionar) para incluir essa variável na lista dos fatores; ela irá aparecer como **imagery** (3) (imagem). A caixa de diálogo final é mostrada na Figura 11.6. Quando você tiver entrado com ambos os fatores dentre sujeitos clique em **Define** (Definir) a fim de ir para a caixa de diálogo principal.

A caixa de diálogo principal é essencialmente a mesma do que quando existe somente uma variável independente, com excessão de que existem agora nove pontos de interrogação (Figura 11.7). No topo da caixa **Within Subjects Variables** (Variáveis entre sujeitos), o SPSS declara que existem dois fatores: **drink** e **imagery**. Na caixa abaixo existe uma série de pontos de interrogações seguidos por

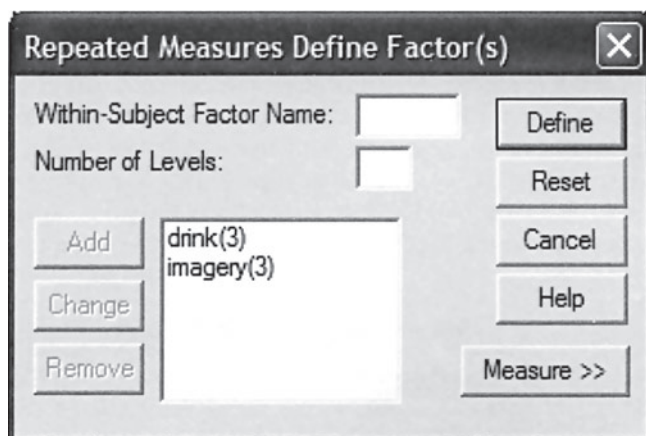


Figura 11.6 Caixa de diálogo **Define factors** (Definir fatores) para a ANOVA fatorial de medidas repetidas.

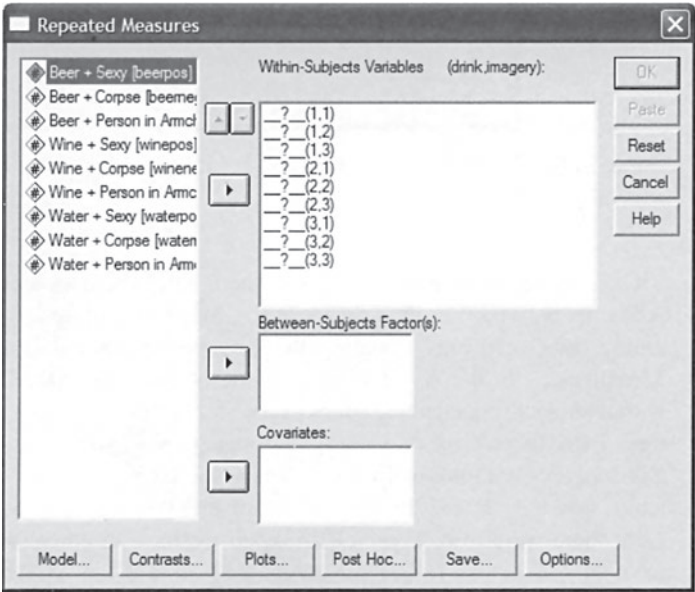


Figura 11.7 Caixa de diálogo principal para a ANOVA de medidas repetidas (antes e após a finalização).

números entre parênteses. Os números nos parênteses representam os níveis dos fatores (variáveis independentes):










?(1, 1)	Variável representando o nível 1 de bebida (drink) e o nível 1 de imagem (imagery)
?(1, 2)	Variável representando o nível 1 de bebida (drink) e o nível 2 de imagem (imagery)
?(1, 3)	Variável representando o nível 1 de bebida (drink) e o nível 3 de imagem (imagery)
?(2, 1)	Variável representando o nível 2 de bebida (drink) e o nível 1 de imagem (imagery)
?(2, 2)	Variável representando o nível 2 de bebida (drink) e o nível 2 de imagem (imagery)
?(2, 3)	Variável representando o nível 2 de bebida (drink) e o nível 3 de imagem (imagery)
?(3, 1)	Variável representando o nível 3 de bebida (drink) e o nível 1 de imagem (imagery)
?(3, 2)	Variável representando o nível 3 de bebida (drink) e o nível 2 de imagem (imagery)
?(3, 3)	Variável representando o nível 3 de bebida (drink) e o nível 3 de imagem (imagery)

Nesse exemplo, existem duas variáveis independentes e, assim, existem dois números nos parênteses. O primeiro número se refere aos níveis do primeiro fator listado acima do quadro (nesse caso, **drink**). O segundo número nos parênteses se refere aos níveis do

segundo fator listado acima do quadro (nesse caso, **imagery**). Assim como com a ANOVA independente de medidas repetidas, você é solicitado a substituir esses pontos de interrogação pelas variáveis da lista do lado esquerdo da caixa de diálogo. Com delineamentos entre grupos, em que códigos de variáveis são utilizados, os níveis de um fator particular são especificados pelos códigos designados a eles no editor de dados. Entretanto, em delineamentos de medidas repetidas, esse código não é usado e, assim, determinamos quais condições devemos atribuir a um determinado nível. Por exemplo, se entrarmos com **beerpos** na lista primeiro, o SPSS irá tratar a cerveja como o primeiro nível de **drink** e imagem positiva como o primeiro nível da variável **imagery**. Entretanto, se entrarmos primeiro com **wineneg** na lista, o SPSS irá considerar o vinho como o primeiro nível de **drink** e imagem negativa como o primeiro nível de **imagery**. Por essa razão, é imperativo que pensemos sobre o tipo de contrastes que queremos fazer *antes* de entrar com as variáveis nessa caixa de diálogo. Nesse delineamento, se obser-



vamos a primeira variável, **drink**, existiam três condições, duas das quais envolvem bebidas alcoólicas. De certo modo, a condição da água age como um controle para ver se os efeitos da imagem são específicos ao álcool. Portanto, para essa variável vamos comparar as condições da cerveja e do vinho com a condição da água. Essa comparação pode ser feita especificando um contraste simples (veja a Tabela 8.6), no qual as condições da cerveja e do vinho são comparadas com a da água, ou usando um contraste de diferença no qual ambas as condições do álcool são comparadas com a condição da água antes de serem comparadas entre si. Em qualquer um dos casos é essencial que a condição da água entre ou como o primeiro ou como o último nível da variável independente **drink** (porque você não pode especificar o nível do meio como a categoria referência num contraste simples). Agora, vamos pensar sobre o segundo fator. O fator imagem também tem uma categoria de controle que não era esperado que mudasse atitudes (imagens neutras). Como anteriormente, podemos estar interessados em usar essa categoria como uma categoria referência num contraste simples³ e assim é importante que essa categoria neutra entre ou como primeiro ou como último nível.

Com base no que foi discutido sobre o uso de contrastes, não faz sentido ter água como nível 3 do fator **drink** e neutra como o terceiro nível do fator imagens. Os níveis restantes podem ser decididos arbitrariamente. Escolhi cerveja como nível 1 e vinho como nível 2 do fator **drink** (bebida). Para a variável **imagery** (imagem) escolhi positiva como nível 1 e negativa como nível 2. Essas decisões significam que as variáveis devem entrar da seguinte forma:

beerpos		_?(1, 1)
beerneg		_?(1, 2)
beerneut		_?(1, 3)
winepos		_?(2, 1)
wineneg		_?(2, 2)
wineneut		_?(2, 3)
waterpos		_?(3, 1)
waterneg		_?(3, 2)
waterneu		_?(3, 3)

Coincidentemente, essa é a ordem na qual as variáveis estão listadas no editor de dados; a coincidência aconteceu simplesmente porque eu me antecipei e pensei o que os contrastes fariam e, então, entrei com as variáveis na ordem apropriada! Quando essas variáveis forem transferidas, a caixa de diálogo deve ficar exatamente igual a da Figura 11.8. Os botões na parte de baixo da tela já foram descritos para o caso da variável independente e, assim, irei descrever somente a mais relevante.

11.8.2 Contrastes ②

Depois da análise principal é interessante comparar níveis das variáveis independentes para ver se elas diferem. Como vimos, há dificuldade para entrar com códigos dos contrastes (a não ser que você use sintaxe), assim, temos que confiar nos contraste padrão disponíveis (veja a Tabela 8.6). A Figura 11.9 mostra a caixa de diálogo para a condução de contrastes e é obtida clicando em  (Contrastes) na caixa de diálogo principal. Na seção anterior, descrevi por que deve ser interessante usar água e condições neutras como categorias base para os fatores **drink** (bebida) e **imagery** (imagem), respectivamente. Usamos a caixa de diálogo **contrasts** (Contrastes) antes, nas seções 9.3.3 e 11.4.2 e, assim, você deve selecionar um contraste simples para cada variável independente. Para ambas as variáveis independentes, entramos com as variáveis de maneira que a categoria de controle fosse a última; portanto, não precisamos trocar a categoria de referência pelo contraste simples. Uma vez que os contrastes foram selecionados, clique em .

³ Esperamos que as imagens positivas melhorem as atitudes, e que as imagens negativas tornem as atitudes mais negativas. Portanto, não faz sentido fazer um contraste Helmert ou contraste de diferença para esse fator porque os efeitos das duas condições experimentais irão cancelar a si mesmas.

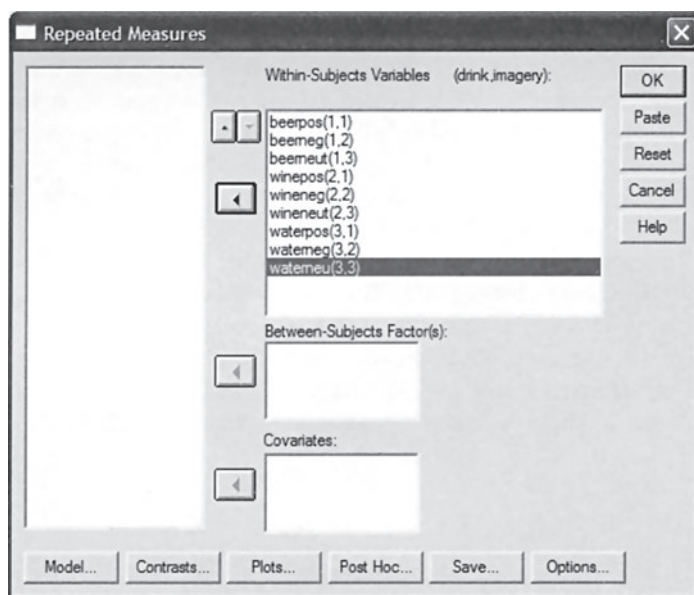


Figura 11.8 Caixa de diálogo principal final para a análise dentre participantes de medidas repetidas

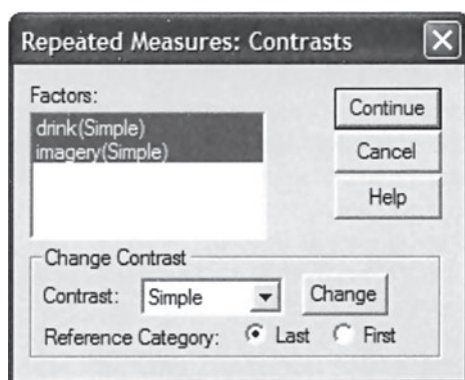


Figura 11.9 Contrastes de medidas repetidas.

(Continuar) para retornar à caixa de diálogo principal. Uma alternativa aos contrastes disponíveis aqui é fazer uma análise dos efeitos simples (veja o Quadro 11.1).

11.8.3 Ilustrando interações ②

Quando tínhamos somente uma variável independente, ignoramos a caixa de diálogo

plots (desenhar); contudo, se existem dois ou mais fatores, a caixa de diálogo é uma maneira conveniente para traçar as médias para cada nível dos fatores. Para acessar essa caixa de diálogo, clique em **Plots...** (Diagramas). Selecione **drink** da lista de variáveis no lado esquerdo da caixa de diálogo e transfira-a para o espaço denominado **Horizontal Axis** (Eixo Horizontal) clicando em **►**. No espaço denominado **Separate Lines** (Linhas Separadas) precisamos colocar a variável independente restante: **imagery**. Como antes, fica a seu critério de que maneira traçar o gráfico. Quando você tiver movido as duas variáveis independentes para a caixa apropriada, clique em **Add** (Adicionar) e esse gráfico de interação será acrescentado à lista na parte de baixo da caixa (veja a Figura 11.10). Quando você terminar com a especificação dos gráficos, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

11.8.4 Outras opções ②

Você deve notar que os testes *post hoc* estão desabilitados somente para os delinea-

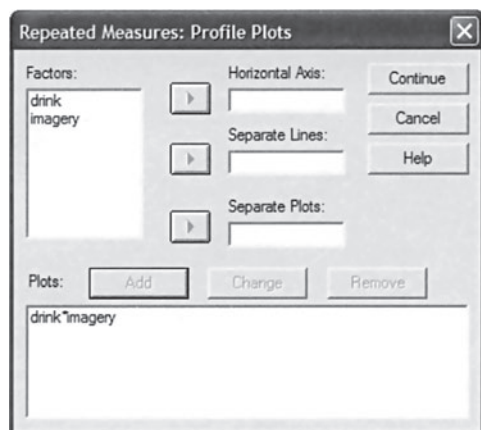


Figura 11.10 Caixa de diálogo “diagramas” para a ANOVA de medidas repetidas.

mentos de medidas repetidas. Portanto, as únicas opções restantes estão na caixa de diálogo **options**, que é acessada clicando em **Options...** (opções). As opções aqui são as mesmas que para a ANOVA de um fator. Recomendo selecionar algumas estatísticas descritivas, e você pode querer selecionar algumas comparações múltiplas, selecionando todos os fatores na caixa denominada de **Factor(s) and Factor Interactions** (Fator(es) e Interações de Fatores) e transfira-os para a caixa denominada **Display Means for** (Exibir Médias) clicando em **►** (veja a Figura 11.1). Tendo selecionado essas variáveis, você deve marcar a caixa denominada **Compare main effects** (Compare efeitos principais) ☒ **Compare main effects** e selecione uma

Quadro 11.1

Análise dos efeitos simples ③



Vimos no capítulo anterior que outra maneira de decompor um termo de interação é usar uma técnica chamada de análise dos “efeitos simples”. Essa análise observa o efeito de uma variável independente em níveis individuais da outra variável independente. Assim, para esse exemplo, podemos observar o efeito da bebida para imagens positivas, para imagens negativas e, então, para imagens neutras. Também poderíamos analisar o efeito das imagens separadamente para cerveja, vinho e água. Com delineamentos de medidas repetidas podemos, ainda, conduzir efeitos simples por meio da sintaxe do SPSS, mas a sintaxe que usamos é ligeiramente diferente. A sintaxe que você precisa usar é:

MANOVA

```
beerpos beerneg beerneut winepos wineneg wineneut waterpos waterneg waterneu
/WSFACTORS drink (3) imagery (3)
```

* Isso inicia a ANOVA especificando as variáveis no editor de dados que se referem aos níveis das nossas variáveis de medidas repetidas. O comando /WSFACTORS define as duas variáveis de medidas repetidas que temos. A ordem na qual listamos as variáveis do editor de dados é importante. Assim, porque definimos que bebida(3) imagem(3) o SPSS inicia no nível 1 de bebida e porque especificamos os três níveis de imagens, ele usa as primeiras três variáveis listadas como os níveis de imagens no nível 1 de bebida. Ele continua no nível 2 de bebida e novamente observa as próximas três variáveis na lista para serem relevantes níveis de imagens. Finalmente, ele continua no nível 3 de bebida e usa as próximas três variáveis (as últimas três, nesse caso) para serem os níveis de imagens. Isso é difícil de explicar, mas observe a ordem das variáveis e veja que as primeiras três se referem à cerveja (e diferem de acordo com imagens), as próximas três são vinho e os três níveis de imagens e as últimas três são água, ordenadas novamente de acordo com imagens. Porque nós as ordenamos dessa maneira, temos que definir bebida(3) e, então, imagens(3). (Seria igualmente válido escrever /WSFACTORS imagery(3) drink(3), mas somente se inicialmente tivéssemos ordenado as variáveis: beerpos winepos waterpos beerneg wineneg waterneg beerneut wineneut waterneu.)

(Continua)

Quadro 11.1 (Continuação)

`/WSDESIGN = MWITHIN drink(1) MWITHIN drink(2) MWITHIN drink(3)`

* Isso especifica os efeitos simples. Por exemplo, “MWITHIN drink(1)” solicita ao SPSS analisar o efeito de imagem no nível 1 de bebida (isto é, quando a cerveja foi usada). Se você quiser comparar bebida aos níveis de imagens, escreveríamos isso ao contrário:

`/WSDESIGN = MWITHIN imagery(1) MWITHIN imagery (2) MWITHIN imagery (3)`

`/PRINT`

`SIGNIF (UNIV MULT AVERF HF GG)`

* Essas linhas finais apenas solicitam que a ANOVA principal seja impressa (SIGNIF). A sintaxe para olhar o efeito de imagens em diferentes níveis de bebida está armazenada no arquivo denominado **SimpleEffectsAttitude.sps** para você olhar. Abra esse arquivo (confira se você também tem **Attitude.sav** carregado no editor de dados) e execute a sintaxe. A saída que você irá conseguir será na forma de texto (em vez de tabelas bonitas). Parte dela irá duplicar os resultados da ANOVA principal. Os efeitos simples são apresentados assim:

*****A n a l y s i s o f V a r i a n c e - d e s i g n 1 *****
(Análise de Variância - delineamento 1)

Tests involving "MWITHIN DRINK(1) Within-Subject Effect.

(Testes envolvendo "MWITHIN DRINK(1) Efeito Entre Participantes)

Tests of Significance for T1 using UNIQUE sums of squares

(Testes de significância para T1 utilizando soma dos quadrados UNIQUE)

Source of Variation	SS	DF	MS	F	Sig of F
(Fonte de Variação)	SS	GL	MS	F	Sig de F)
WITHIN+RESIDUAL	7826,67	19	412,09		
(Entre + Resíduo)					
MWITHIN DRINK(1)	8401,67	1	8401,67	20,39	0,000
(Entre Bebida(1))					

*****A n a l y s i s o f V a r i a n c e - d e s i g n 1 *****
(Análise de Variância - delineamento 1)

Tests involving "MWITHIN DRINK(2) Within-Subject Effect.

(Testes envolvendo "MWITHIN DRINK(2) Efeito Entre Participantes)

Tests of Significance for T2 using UNIQUE sums of squares

(Testes de significância para T2 utilizando soma dos quadrados UNIQUE)

Source of Variation	SS	DF	MS	F	Sig of F
(Fonte de Variação)	SS	GL	MS	F	Sig de F)
WITHIN+RESIDUAL	376,00	19	19,79		
(Entre + Resíduo)					
MWITHIN DRINK(2)	4166,67	1	4166,67	210,55	0,000
(Entre Bebida(2))					

(Continua)

Quadro 11.1 (Continuação)

```
*****A n a l y s i s o f V a r i a n c e - d e s i g n 1 *****
(Análise de Variância - delineamento 1)
Tests involving "MWITHIN DRINK(3) Within-Subject Effect.
(Testes envolvendo "MWITHIN DRINK(3) Efeito Entre Participantes)
Tests of Significance for T3 using UNIQUE sums of squares
(Testes de significância para T3 utilizando soma dos quadrados UNIQUE)
Source of Variation      SS      DF      MS      F      Sig of F
(Fonte de Variação)      SS      GL      MS      F      Sig de F)
WITHIN+RESIDUAL          1500,32   19      78,96
(Entre + Resíduo)
MWITHIN DRINK(3)         742,02    1      742,02    9,40    0,006
(Entre Bebida(3))
```

A tabela denominada "MWITHIN DRINK(1)" fornece uma ANOVA do efeito das imagens para cerveja e as tabelas subsequentes são para vinho e água, respectivamente. Observando os valores de significância para cada efeito simples parece que existem efeitos significativos de imagens em todos os níveis da bebida!

Podemos, também, complementar os efeitos com contrastes especialmente definidos para o termo de interação. Infelizmente, isso somente pode ser feito usando a sintaxe e é um processo bem envolvente, assim, se isso é algo que você queira fazer, veja o arquivo **ContrastsUsingSyntax.pdf** no site www.artmed.com.br.

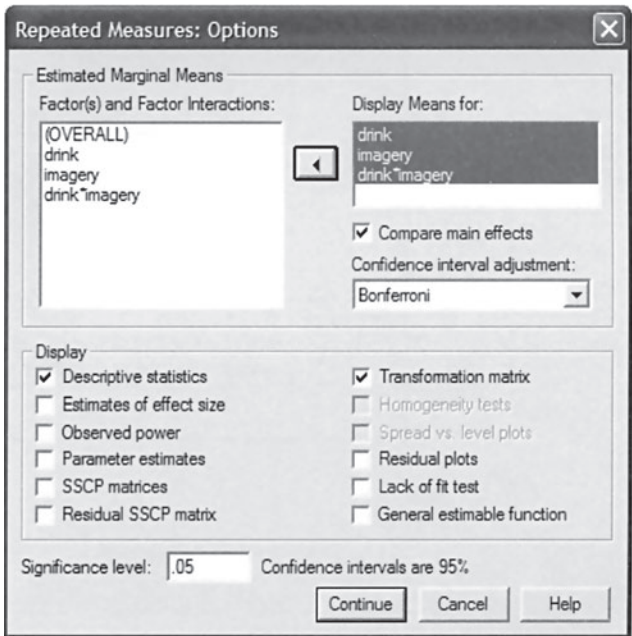


Figura 11.11 Caixa de diálogo “opções” para a ANOVA de medidas repetidas.

correção apropriada (eu escolhi Bonferroni). A única opção restante de especial interesse é selecionar a opção **Transformation matrix** (Matriz de Transformação). Essa opção produz muitas saídas extras, mas é importante para interpretar a saída dos contrastes.

11.9 SAÍDAS PARA A ANOVA FATORIAL DE MEDIDAS REPETIDAS ②

11.9.1 Descritivas e análise principal ②

A saída do SPSS 11.9 mostra a saída inicial dessa ANOVA. A primeira tabela apenas lista as variáveis que foram incluídas no editor de dados e o nível de cada variável independente que elas representam. Essa tabela é mais importante do que pode parecer porque ela permite verificar que você entrou com as variáveis na ordem correta para as comparações que você deseja fazer. A segunda tabela é a de descritivas e fornece a média e o desvio padrão para cada uma das nove condições. Os nomes nessa tabela são os que forneci para as variáveis no editor de dados (portanto, se você não deu a essas variáveis nomes completos, ela será um pouco diferente).

As descritivas são interessantes porque elas nos dizem que a variabilidade entre es-

cores foi maior quando a cerveja foi usada como um produto (compare os desvios padrão da variável cerveja contras as outras). Da mesma forma, quando a imagem de um cadáver foi usada, os índices dados aos produtos foram negativos (como esperado) para vinho e água, mas não para cerveja (assim, por alguma razão, as imagens negativas parecem não funcionar quando a cerveja foi utilizada como um estímulo). Os valores nessa tabela nos ajudarão a interpretar os efeitos principais da análise.

A Saída 11.10 do SPSS mostra os resultados do teste de esfericidade de Mauchly (veja a Seção 11.2.3) para cada um dos três efeitos no modelo (dois efeitos principais e uma interação). Os valores de significância desses testes indicam que ambos os efeitos principais de **bebida** e **imagens** violaram essa hipótese e, assim, os valores de *F* devem ser corrigidos (veja a Seção 11.5.2). Para a interação, a hipótese de esfericidade é encontrada (porque $p > 0,05$) e, assim, não precisamos corrigir a razão *F* para esse efeito.

A Saída 11.11 do SPSS mostra os resultados para a ANOVA (com valores *F* corrigidos – se alguns de vocês estiverem usando a versão 7.5 ou anterior, expanda a tabela produzida usando o método descrito na Seção 11.5.3).

Saída 11.9 do SPSS

Within-Subjects Factors
(Fatores entre Participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

<i>DRINK</i> (Bebida)	<i>IMAGERY</i> (Imagem)	<i>Dependent Variable</i> (Variável Dependente)
1	1	BEERPOS
	2	BEERNEG
	3	BEERNEUT
2	1	WINEPOS
	2	WINENEG
	3	WINENEUT
3	1	WATERPOS
	2	WATERNEG
	3	WATERNEU

Descriptive Statistics (Estatísticas Descritivas)

	<i>Mean</i> (Média)	<i>Std. Deviation</i> (Desvio Padrão)	<i>N</i>
<i>Beer + Sexy</i> (Cerveja + Sexy)	21.0500	13.0080	20
<i>Beer + Corpse</i> (Cerveja + Cadáver)	4.4500	17.3037	20
<i>Beer + Person in Armchair</i> (Cerveja + Pessoa Sentada)	10.0000	10.2956	20
<i>Wine + Sexy</i> (Vinho + Sexy)	23.3500	6.7378	20
<i>Wine + Corpse</i> (Vinho + Cadáver)	-12.0000	6.1815	20
<i>Wine + Person in Armchair</i> (Vinho + Pessoa Sentada)	11.6500	6.2431	20
<i>Water + Sexy</i> (Água + Sexy)	17.4000	7.0740	20
<i>Water + Corpse</i> (Água + Cadáver)	-9.2000	6.8025	20
<i>Water + Person in Armchair</i> (Água + Pessoa Sentada)	2.3500	6.8386	20

Saída do SPSS 11.10

Mauchly's Test of Sphericity^b (Teste de Mauchly da Esfericidade)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Within-Subjects Effect (Efeito entre Participantes)	Mauchly's W (W de Mauchly)	Approx. Chi-Square (Qui-Quadrado Aproximado)	df (gl)	Sig. (Sig.)	Epsilon ^a		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound (Limite Inferior)
DRINK (Bebida)	0.267	23.753	2	0.000	0.577	0.591	0.500
IMAGERY (Imagem)	0.662	7.422	2	0.024	0.747	0.797	0.500
DRINK * IMAGERY (Bebida * Imagem)	0.595	9.041	9	0.436	0.798	0.979	0.250

Testa a hipótese nula de que a matriz de covariância dos erros das variáveis dependentes transformadas e ortonormalizadas é proporcional a uma matriz identidade.

a Pode ser utilizado para ajustar os graus de liberdade dos testes de significância para médias. Os testes corrigidos são mostrados em camadas (por padrão) na tabela dos testes de efeitos entre participantes.

b Delineamento: Intercepto dentre participantes: BEBIDA-IMAGEM + BEBIDA*IMAGEM.

A saída é dividida em seções que se referem a cada um dos efeitos no modelo e os termos erro associados com esses efeitos. Observando os valores de significância está claro que existe um efeito significativo do tipo de bebida usado como um estímulo, um efeito principal significativo do tipo de imagens usadas e uma interação significativa entre essas duas variáveis. Irei examinar cada um desses efeitos em turnos.

11.9.2 O efeito da bebida ②

A primeira parte da Saída 11.11 do SPSS nos fala do efeito do tipo da bebida usada no comercial. Para esse efeito devemos observar um dos valores de significância corrigidos porque a esfericidade foi violada. Todos os valores corrigidos são significativos e, assim, devemos relatar os valores conservadores dos graus de liberdade corrigidos de Greenhouse e Geisser. Esse efeito nos diz que se ignorarmos o tipo de imagens que foram usadas os participantes ainda avaliaram alguns tipos de bebidas de forma significativamente diferente.

Na Seção 11.8.4, solicitamos ao SPSS que mostrasse as médias de todos os efeitos no modelo (antes de conduzir os testes *post hoc*) e se você der uma olhada por toda a sua saída, deverá encontrar a tabela da Saída 11.12 do SPSS na seção denominada **Estimated Mar-**

ginal Means⁴ (Médias Marginais Estimadas). A Saída 11.12 do SPSS é uma tabela de médias para o efeito principal de bebida com os erros padrão associados. Os níveis dessa variável são denominados 1, 2 e 3 e, assim, devemos lembrar como entramos a variável para ver qual linha da tabela se relaciona a que condição. Entramos essa variável com a condição cerveja primeiro e a condição água por último. A Figura 11.12 usa essa transformação para mostrar as médias para cada condição. Fica claro a partir desse gráfico que cerveja e vinho têm, naturalmente, uma avaliação mais alta do que a água (tendo a cerveja a avaliação mais alta). Para ver a natureza desse efeito podemos observar os testes *post hoc* (veja a seguir) e os contrastes (veja a Seção 11.9.5).

A Saída 11.13 do SPSS mostra comparações de pares para o efeito principal corrigido de bebida usando o ajuste de Bonferroni. Essa tabela indica que o efeito principal significativo reflete uma diferença significativa ($p < 0,01$) entre os níveis 2 e 3 (vinho e água).

⁴ Essas médias são obtidas calculando as médias das médias na Saída 11.9 do SPSS para uma condição dada. Por exemplo, a média para a condição da cerveja (ignorando imagens) é:

$$\begin{aligned}\bar{X}_{\text{Cerveja}} &= (\bar{X}_{\text{Cerveja+Sexy}} + \bar{X}_{\text{Cerveja+Cadáver}} + \bar{X}_{\text{Cerveja+Neutro}})/3 \\ &= (21,05 + 4,45 + 10,00)/3 = 11,83\end{aligned}$$

Saída 11.11 do SPSS

Tests of Within-Subjects Effects (Testes de efeitos entre participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)		Type III Sum of Squares (Soma dos quadrados do tipo III)	df (gl)	Mean Square (Média dos quadrados)	F	Sig.
DRINK (Bebida)	Sphericity Assumed (Esfericidade Assumida)	2092.344	2	1046.172	5.106	0.011
	Greenhouse-Geisser	2092.344	1.154	1812.764	5.106	0.030
	Huynh-Feldt	2092.344	1.181	1770.939	5.106	0.029
	Lower-bound (Limite inferior)	2092.344	1.000	2092.344	5.106	0.036
Error(DRINK) (Erro(Bebida))	Sphericity Assumed (Esfericidade Assumida)	7785.878	38	204.892		
	Greenhouse-Geisser	7785.878	21.930	355.028		
	Huynh-Feldt	7785.878	22.448	346.836		
	Lower-bound (Limite inferior)	7785.878	19.000	409.783		
IMAGERY (Imagem)	Sphericity Assumed (Esfericidade Assumida)	21628.678	2	10814.339	122.565	0.000
	Greenhouse-Geisser	21628.678	1.495	14468.490	122.565	0.000
	Huynh-Feldt	21628.678	1.594	13571.496	122.565	0.000
	Lower-bound (Limite inferior)	21628.678	1.000	21628.678	122.565	0.000
Error(IMAGERY) (Erro(Imagem))	Sphericity Assumed (Esfericidade Assumida)	3352.878	38	88.234		
	Greenhouse-Geisser	3352.878	28.403	118.048		
	Huynh-Feldt	3352.878	30.280	110.729		
	Lower-bound (Limite inferior)	3352.878	19.000	176.467		
DRINK*IMAGERY (Bebida*Imagem)	Sphericity Assumed (Esfericidade Assumida)	2624.422	4	656.106	17.155	0.000
	Greenhouse-Geisser	2624.422	3.194	821.778	17.155	0.000
	Huynh-Feldt	2624.422	3.914	670.462	17.155	0.000
	Lower-bound (Limite inferior)	2624.422	1.000	2624.422	17.155	0.001
Error(DRINK*IMAGERY) (Erro(Bebida*Imagem))	Sphericity Assumed (Esfericidade Assumida)	2906.689	76	38.246		
	Greenhouse-Geisser	2906.689	60.678	47.903		
	Huynh-Feldt	2906.689	74.373	39.083		
	Lower-bound (Limite inferior)	2906.689	19.000	152.984		

Saída 11.12 do SPSS

Estimates (Estimativas)

Measure: MEASURE_1 (Medida: MEDIDA_1)

DRINK (Bebida)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	11.833	2.621	6.348	17.319
2	8.333	0.574	7.131	9.535
3	3.517	1.147	1.116	5.918

Curiosamente, a diferença entre as condições cerveja e água é maior do que para as de vinho e água, mas esse efeito não é significativo ($p > 0,05$). Essa inconsistência pode ser explicada observando o erro padrão na condição da cerveja comparada com a condição do vinho. O erro padrão para a condição do vinho é incrivelmente menor e, assim, a diferença entre as médias é relativamente grande (veja o Capítulo 6). Tente executar novamente os testes *post hoc*, mas selecione os valores não corrigidos (LSD) na caixa de diálogo **options** (veja a Seção 11.8.4). Você deve descobrir que a diferença entre cerveja e água agora é significativa ($p = 0,02$). Essa descoberta

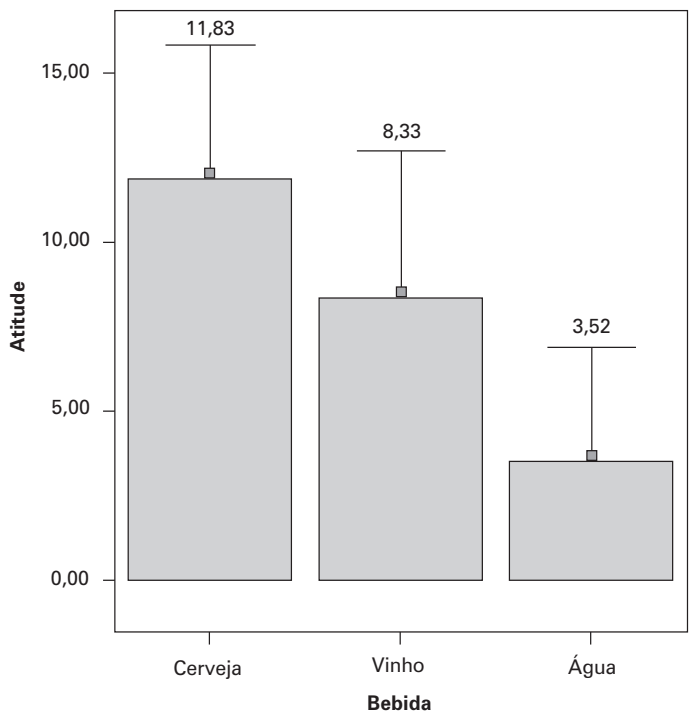


Figura 11.12 Gráfico de barras de erro mostrando a média de cada condição.

Saída 11.13 do SPSS

Pairwise Comparisons (Comparações aos Pares)

Measure: MEASURE_1 (Medida: MEDIDA_1)

(I) DRINK (II) Bebida	(J) DRINK (J) Bebida	Mean Difference (I - J) (Diferença Média)	Std. Error (Erro Padrão)	Sig. ^a	95% Confidence Interval for difference ^a (Intervalo de 95% de confiança para diferença)	
					Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	2	3.500	2.849	0.703	-3.980	10.980
	3	8.317	3.335	0.066	-0.438	17.072
2	1	-3.500	2.849	0.703	-10.980	3.980
	3	4.817*	1.116	0.001	1.886	7.747
3	1	-8.317	3.335	0.066	-17.072	0.438
	2	-4.817*	1.116	0.001	-7.747	-1.886

Com base nas médias marginais estimadas)
* A diferença média é significativa ao nível de 0,05.
a Ajuste para comparações múltiplas: Bonferroni.

ressalta a importância do controle da taxa do erro usando a correção de Bonferroni. Se não tivéssemos usado essa correção poderíamos

ter concluído erroneamente que a cerveja teve avaliação significativamente mais alta do que a água.

11.9.3 O efeito da imagem ②

A Saída 11.11 do SPSS também indica que o efeito do tipo de imagens usadas no anúncio teve uma influência significativa nas avaliações dos estímulos dos participantes. Novamente, devemos observar um dos valores de significância corrigidos porque a esfericidade foi violada (veja anteriormente). Todos os valores corrigidos são altamente significativos e, assim, podemos novamente relatar os valores corrigidos dos graus de liberdade da estatística de Greenhouse e Geisser. Esse efeito nos diz que se ignorarmos o tipo de bebida que foi usado, as avaliações dos participantes dessas bebidas foram diferentes de acordo com o tipo de imagens usado. Na Seção 11.8.4, solicitamos médias para todos os efeitos no modelo e se você olhar as saídas deverá encontrar a Tabela 11.14 (após a comparação dos pares para o efeito principal de bebida). Essa saída do SPSS

é uma tabela de médias para o efeito principal de imagens com os erros padrão associados. Os níveis dessa variável estão rotulados como 1, 2 e 3 e, assim, devemos nos lembrar como entramos com a variável para ver qual linha da tabela se relaciona com qual condição. Nós en-

Saída 11.14 do SPSS
Estimates (Estimativas)

Measure: MEASURE_1 (Medida: MEDIDA_1)

IMAGERY (Imagem)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	21.267	0.977	19.222	23.312
2	-5.583	1.653	-9.043	-2.124
3	8.000	0.969	5.972	10.028

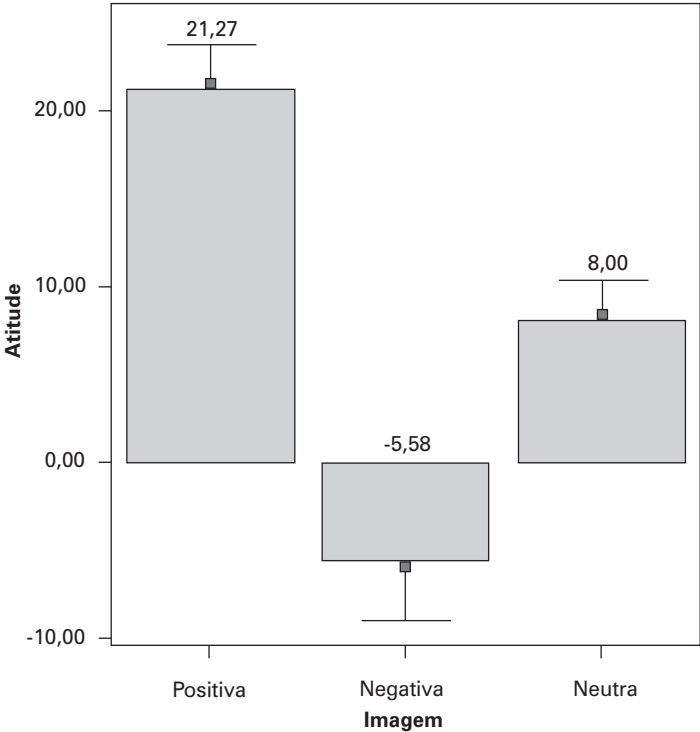


Figura 11.13 Gráfico de barras de erro mostrando o efeito da imagem em cada condição.

Saída 11.15 do SPSS

Pairwise Comparisons (Comparações aos Pares)

Measure: MEASURE_1 (Medida: MEDIDA_1)

(I) IMAGERY (II) Imagem	(J) IMAGERY (J) Imagem	Mean Difference (I - J) (Diferença Média)	Std. Error (Erro Padrão)	Sig. ^a	95% Confidence Interval for Difference ^a (Intervalo de 95% de Confiança para a Diferença)	
					Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	2	26.850*	1.915	0.000	21.824	31.876
	3	13.267*	1.113	0.000	10.346	16.187
2	1	-26.850*	1.915	0.000	-31.876	-21.824
	3	-13.582*	1.980	0.000	-18.781	-8.386
3	1	-13.267*	1.113	0.000	-16.187	-10.346
	2	13.583*	1.980	0.000	8.386	18.781

Based on estimated marginal means (Com base nas médias marginais estimadas)
* The mean difference is significant at the .05 level (A diferença média é significativa ao nível de 0,05)
a Adjustment for multiple comparisons: Bonferroni (Ajuste para comparações múltiplas: Bonferroni)

tramos essa variável com a condição positiva primeiro e a condição neutra por último. A Figura 11.13 usa essa informação para ilustrar as médias para cada condição. Fica claro, a partir desse gráfico, que a imagem positiva resultou em uma avaliação muito positiva (comparada com a imagem neutra) e a imagem negativa resultou em uma avaliação negativa (especialmente comparada ao efeito de imagem neutra). Para ver a natureza desse efeito podemos olhar aos testes *post hoc* (veja anteriormente) e os contrastes (veja a Seção 11.9.5).

A Saída 11.15 do SPSS mostra a comparação dos pares para o efeito principal corrigido de imagens usando o ajuste de Bonferroni. Essa tabela indica que o efeito principal significativo reflete diferenças significativas (todos $p < 0,01$) entre níveis 1 e 2 (positiva e negativa), entre níveis 1 e 3 (positiva e neutra) e entre níveis 2 e 3 (negativa e neutra).

11.9.4 O efeito interação (bebida × imagem) ②

A Saída 11.11 do SPSS indicava que as imagens interagiam de alguma maneira com o tipo de bebida usada como estímulo. Daquela tabela devemos relatar que havia uma interação significativa entre o tipo de bebida usada e imagens associadas a ela, $F(4, 76) = 17,16, p$

$< 0,001$. Esse efeito nos diz que o tipo de imagens usadas tem efeito diferente dependendo do tipo de bebida com que ela foi apresentada junto. Como antes, podemos usar as médias que solicitamos na Seção 11.8.4 para determinar a natureza dessa interação (essa tabela deve estar abaixo das comparações dos pares para imagens e está exibida na Saída 11.16 do SPSS). A tabela das médias na Saída 11.16 do SPSS é essencialmente a mesma das estatísticas descritivas iniciais na Saída 11.9 do SPSS exceto que os erros padrão estão exibidos em vez dos desvios padrão.

As médias na Saída 11.16 do SPSS são usadas para criar o gráfico que solicitamos na Seção 11.8.3 e esse gráfico é essencial para a interpretação da interação. A Figura 11.14 mostra o gráfico de interação (levemente modificado para ficar mais bonito!) e estamos procurando por linhas não-paralelas. O gráfico mostra que o padrão das respostas ao longo das bebidas foi similar quando imagens positivas e neutras foram usadas. Isto é, as avaliações foram positivas para cerveja, levemente altas para vinho e baixaram levemente para água. O fato de que a linha representando imagens positivas é mais alta do que a linha neutra indica que imagens positivas aumentam as avaliações mais do que imagens neutras para todas as bebidas. A linha mais embaixo (re-

Saída 11.16 do SPSS

Estimates (Estimativas)

Measure: MEASURE_1 (Medida: MEDIDA_1)

DRINK (Bebida)	IMAGERY (Imagem)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de Confiança)	
				Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	1	21.050	2.909	14.962	27.138
	2	4.450	3.869	-3.648	12.548
	3	10.000	2.302	5.181	14.819
2	1	25.350	1.507	22.197	28.503
	2	-12.000	1.382	-14.893	-9.107
	3	11.650	1.396	8.728	14.572
3	1	17.400	1.582	14.089	20.711
	2	-9.200	1.521	-12.384	-6.016
	3	2.350	1.529	-0.851	5.551

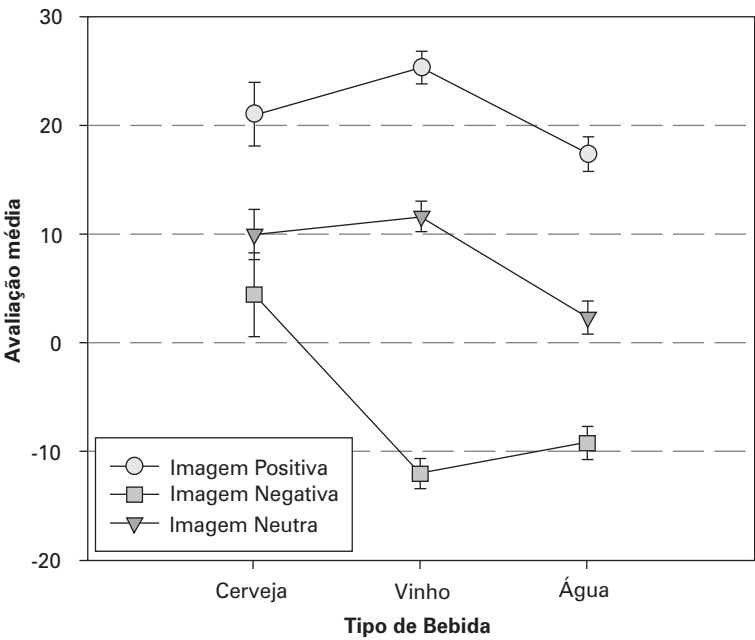


Figura 11.14 Gráfico de interação para Attitude.sav. O tipo de imagens é representado pelas três linhas: imagens positivas (círculos), imagens negativas (quadrados) e imagens neutras (triângulos).

presentando imagens negativas) mostra um efeito diferente: as avaliações foram mais baixas para vinho e água, mas não para cerveja. Portanto, imagens negativas tiveram o efeito desejado em atitudes em relação ao vinho e água, mas, por alguma razão, atitudes em re-

lação à cerveja permaneceram praticamente neutras. Portanto, a interação provavelmente reflete o fato de que imagens negativas têm efeito diferente em ambas as imagens, positiva e neutra (porque ela diminui as avaliações em vez de aumentá-las). Essa interação está

completamente de acordo com as previsões experimentais. Para verificar a interpretação do efeito de interação precisamos observar os contrastes que solicitamos na Seção 11.8.2.

11.9.5 Contrastes para variáveis de medidas repetidas ②

Na Seção 11.8.2, solicitamos contrastes simples para a variável **drink** (para qual a água foi usada como a categoria controle) e

para a categoria **imagery** (para qual a categoria imagem neutra foi usada como controle). A Saída 11.17 do SPSS mostra um resumo dos resultados para esses contrastes. A tabela é dividida em efeitos principais e interações e cada efeito é dividido nos componentes do contraste. Assim, para o efeito principal da bebida, o primeiro contraste compara o nível 1 (cerveja) com a categoria base (nesse caso, a última categoria: água). Se você está confuso sobre os níveis, lembre que na Saída 11.9 do

Saída 11.17 do SPSS

Tests of Within-Subjects Contrasts (Testes dos contrastes entre participantes)

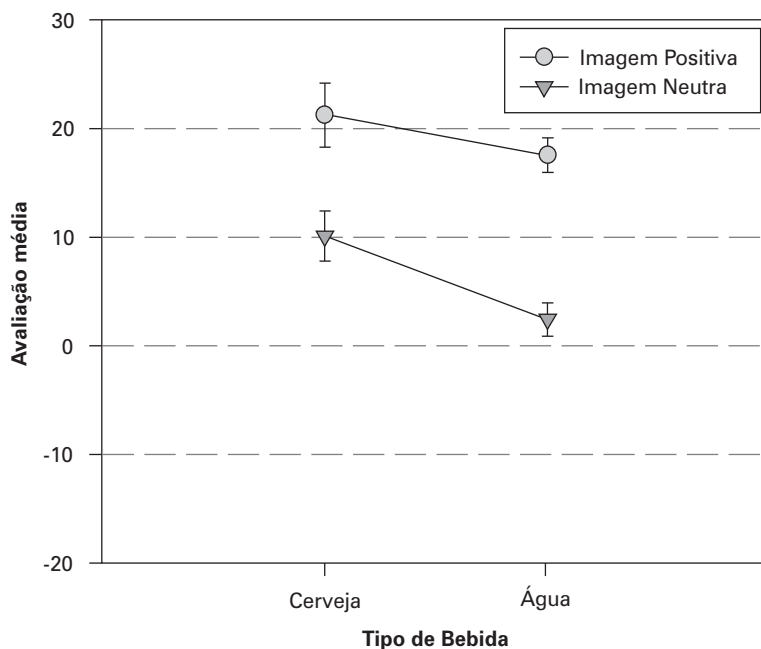
Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)	DRINK (Bebida)	IMAGERY (Imagem)	Type III Sum of Squares (Soma dos quadrados do tipo III)	df (gl)	Mean Square (Média dos quadrados)	F	Sig.
DRINK (Bebida)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		1383.339	1	1383.339	6.218	0.022
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)		464.006	1	464.006	18.613	0.000
Error(DRINK) (Erro(Bebida))	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		4226.772	19	222.462		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)		473.661	19	24.930		
IMAGERY (Imagem)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		3520.089	1	3520.089	142.194	0.000
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)		3690.139	1	3690.139	47.070	0.000
Error(IMAGERY) (Erro(Imagem))	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		470.356	19	24.756		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)		1489.528	19	78.396		
DRINK*IMAGERY (Bebida*Imagem)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	320.000	1	320.000	1.576	0.225
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	720.000	1	720.000	6.752	0.018
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	36.450	1	36.450	0.235	0.633
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	2928.200	1	2928.200	26.906	0.000
Error(DRINK*IMAGERY) (Erro(Bebida*Imagem))	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	3858.000	19	203.053		
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	2026.000	19	106.632		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	2946.550	19	155.082		
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	2067.800	19	108.832		

SPSS existe uma listagem desses valores. Esse resultado é significativo, $F(1, 19) = 6,22$, $p < 0,05$, o que contradiz o que foi encontrado usando os testes *post hoc* (veja a Saída 11.13 do SPSS). Por que isso acontece? O próximo contraste compara o nível 2 (vinho) com a categoria base (água) e confirma a diferença significativa encontrada com os testes *post hoc*, $F(1, 19) = 18,61$, $p < 0,001$. Para o efeito principal da imagem, o primeiro contraste compara o nível 1 (positiva) com a categoria base (a última categoria: neutra) e verifica a diferença significativa encontrada com os testes *post hoc*, $F(1, 19) = 142,19$, $p < 0,001$. O segundo contraste confirma a diferença significativa nas avaliações encontradas na condição da imagem negativa comparada com a neutra, $F(1, 19) = 47,07$, $p < 0,001$. Esses contrastes estão todos bem, mas eles nos dizem somente o que já sabemos (contudo, note o crescente poder estatístico com esses testes mostrado pelos altos valores de significância). Os contrastes se tornam mais interessantes quando observamos o termo de interação.

11.9.5.1 Cerveja versus água, percepção positiva versus neutra ②

O primeiro termo da interação compara o nível 1 de bebida (cerveja) com o nível 3 (água) quando imagens positivas (nível 1) são utilizadas em comparação com imagens neutras (nível 3). Esse contraste não é significativo. Esse resultado nos diz que o aumento da simpatia encontrado quando imagens positivas são usadas (comparadas com imagens neutras) é o mesmo tanto para cerveja ou água. Em termos do gráfico de interação (Figura 11.14) significa que a distância entre o círculo e o triângulo na condição da cerveja é a mesma do que a distância entre o círculo e do triângulo na condição da água. Se apenas traçarmos essa seção do gráfico de interação, é fácil ver que as linhas são aproximadamente paralelas, indicando que não há efeito de interação. Podemos concluir que o aumento das avaliações devido às imagens positivas comparadas às neutras não é afetada se as pessoas estão avaliando a cerveja ou a água.

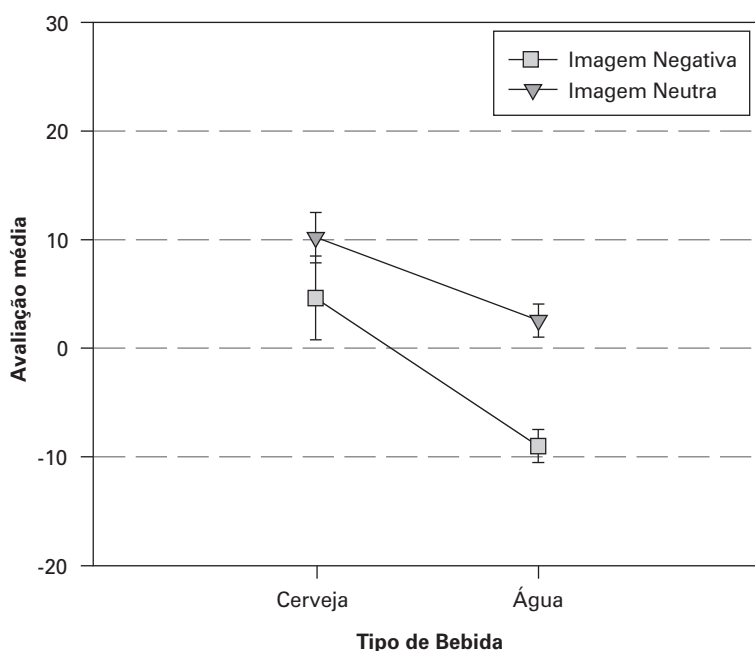


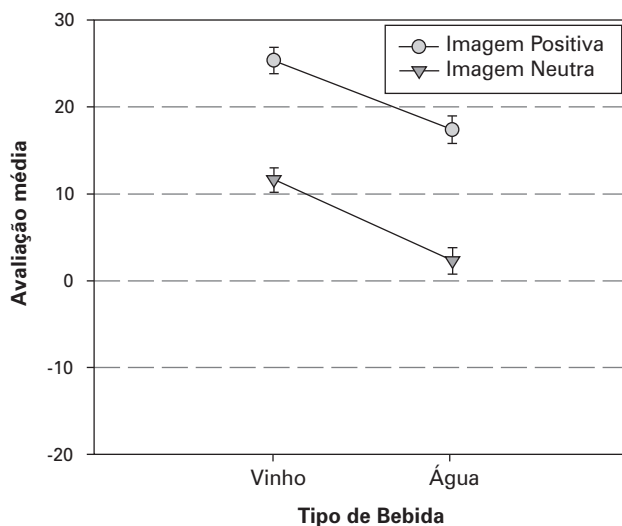
11.9.5.2 Cerveja *versus* Água, Percepção Negativa *versus* Neutra ②

O segundo termo de interação compara o nível 1 de bebida (cerveja) ao nível 3 (água) quando imagens negativas (nível 2) são utilizadas em comparação a neutras (nível 3). Esse contraste é significativo, $F(1, 19) = 6,75$, $p < 0,05$. Esse resultado nos diz que o decréscimo de simpatia encontrado quando imagens negativas são usadas (comparadas às imagens neutras) é diferente quando a cerveja é usada do que quando a água é utilizada. Em termos de gráfico de interação (Figura 11.14), significa que a distância entre o quadrado e o triângulo na condição cerveja (uma diferença pequena) é significativamente menor do que a distância entre o quadrado e o triângulo na condição da água (uma diferença maior). Podemos concluir que o decréscimo nas avaliações devido às imagens negativas (comparadas às neutras) encontrado quando água é usada no comercial é menor do que quando a cerveja é utilizada.

11.9.5.3 Vinho *versus* água, percepção positiva *versus* neutra ②

O terceiro termo de interação avalia o nível 2 da bebida (vinho) comparado ao nível 3 (água) quando imagens positivas (nível 1) são usadas em comparação a neutras (nível 3). Esse contraste não é significativo, indicando que o aumento da simpatia quando imagens positivas são usadas (comparadas às imagens neutras) é a mesma tanto para vinho quanto para água. Em termos de interação de gráfico (Figura 11.14), significa que a distância entre o círculo e o triângulo na condição do vinho é a mesma da distância entre o círculo e o triângulo na condição da água. Se apenas traçarmos essa seção do gráfico de interação, é fácil ver que as linhas são paralelas, indicando um efeito sem interação. Podemos concluir que o aumento das avaliações devido às imagens positivas comparadas às neutras não é afetado pelo fato de as pessoas estarem avaliando o vinho ou a água.

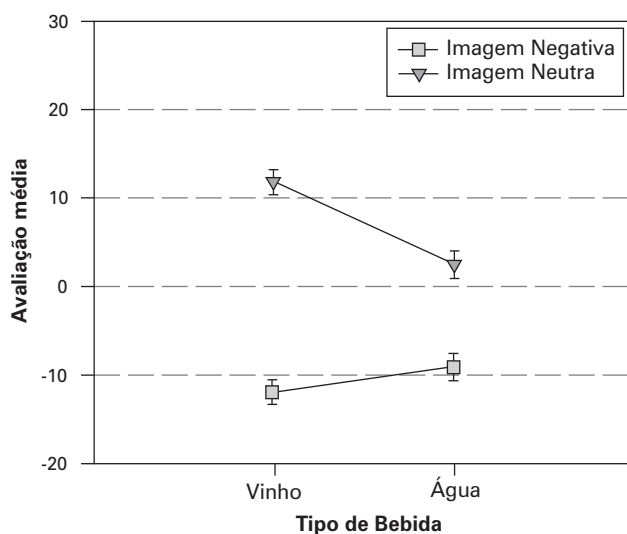




11.9.5.4 Vinho *versus* água, percepção negativa *versus* neutra ②

O termo final das interações observa o nível 2 da bebida (vinho) comparado ao nível 3 (água), quando imagens negativas (nível 2) são usadas em comparação a neutras (nível 3). Esse contraste é significativo, $F(1, 19) = 26,91$, $p < 0,001$. Esse resultado nos diz que o decréscimo de simpatia encontrado quando imagens negativas são usadas (comparadas às imagens neu-

tras) é diferente quando vinho é usado comparado com quando a água é usada. Em termos de interação no gráfico (Figura 11.14), significa que a distância entre o quadrado e o triângulo na condição do vinho (uma diferença grande) é significativamente maior do que a distância entre o quadrado e triângulo na condição da água (uma diferença menor). Podemos concluir que o decréscimo nas avaliações devido às imagens negativas (comparadas às neutras) é significati-



vamente maior quando o vinho é anunciado do que quando a água é anunciada.

11.9.5.5 Limitações desses contrastes ②

Esses contrastes, por sua natureza, não nos dizem nada sobre as diferenças entre as condições da cerveja e vinho (ou das condições positivas e negativas), e outros contrastes deverão ser feitos para descobrir mais. Entretanto, o que está claro, até agora, é que em relação à condição neutra imagens positivas aumentam para mais ou menos a simpatia pelos produtos sem levar em conta o produto; entretanto, imagens negativas tiveram um grande efeito no vinho e um efeito menor na cerveja. Essas diferenças não foram previstas. Embora possa parecer cansativo gastar tanto tempo interpretando uma análise, é recomendável adotar essa abordagem se você realmente quer entender os efeitos obtidos. Interpretar termos de interação é complexo, e eu conheço alguns pesquisadores

bem respeitados que ainda têm dificuldade com isso, assim, não fique desanimado se você achar difícil. Tente ser minucioso e analise cada efeito tanto quanto possível usando contrastes, e com sorte você encontrará algum esclarecimento.

11.10 TAMANHO DE EFEITO PARA A ANOVA FATORIAL DE MEDIDAS REPETIDAS ③



Calcular o ômega quadrado para uma ANOVA de um fator de medidas repetidas foi de arrancar os cabelos, e eu continuo dizendo, tamanhos de efeito são realmente mais úteis quando descrevem um efeito focado – assim, aconselho calcular os tamanhos de efeito para os contrastes quando você tem um delineamento fatorial (e qualquer efeito principal que compare somente dois grupos). A Saída 11.17 do SPSS

Dica da Samanta Ferrinho



- A ANOVA de dois fatores de medidas repetidas compara muitas médias quando existem duas variáveis independentes e os mesmos participantes foram usados em todas as condições experimentais.
- Teste a hipótese de esfericidade quando você tem três ou mais condições de medidas repetidas. Faça o teste de esfericidade usando o *teste de Mauchly*. Encontre a tabela com esse rótulo: se o valor na coluna denominada Sig. for menor do que 0,05, a hipótese foi violada. Se a significância do teste de Mauchly é maior do que 0,05, a hipótese de esfericidade foi encontrada. Você deve testar essa hipótese para todos os efeitos (numa ANOVA de dois fatores isso significa testar para os efeitos das variáveis e o termo de interação).
- A tabela denominada **Tests of Within-Subjects Effects** mostra o resultado principal para sua ANOVA. Numa ANOVA de dois fatores você tem três efeitos: um efeito principal para cada variável e a interação entre as duas. Para cada efeito, se a hipótese de esfericidade foi encontrada, olhe para a linha denominada *Sphericity Assumed*. Se a hipótese foi violada, leia a linha denominada *Greenhouse-Geisser* (você pode olhar Huynh-Feldt, mas você deve ler este capítulo para descobrir os méritos relativos dos dois procedimentos). Tendo selecionado a linha apropriada, olhe para a coluna denominada Sig.: se o valor for menor do que 0,05, as médias dos grupos são significativamente diferentes.
- Analise os efeitos principais e termos de interação usando contrastes. Esses contrastes aparecem na tabela denominada **Tests of Within-Subjects Contrasts**; mais uma vez, olhe a coluna denominada Sig. para descobrir se suas comparações são significativas (elas serão se o valor de significância for menor do que 0,05).

mostra os valores para vários contrastes, todos com graus de liberdade 1 para o modelo (isto é, eles representam uma comparação focada e interpretável) e tem 19 graus de liberdade residual. Podemos usar essas razões F e convertê-las para um tamanho de efeito r , usando o que já vimos anteriormente:

$$r = \sqrt{\frac{F(1, g l_R)}{F(1, g l_R) + g l_R}}$$

Para as duas comparações que fizemos para a variável bebida (Saída 11.17 do SPSS), vamos ter:

$$r_{\text{Cerveja vs. Água}} = \sqrt{\frac{6,22}{6,22 + 19}} = 0,50$$

$$r_{\text{Vinho vs. Água}} = \sqrt{\frac{18,61}{18,61 + 19}} = 0,70$$

Portanto, ambas as comparações produzem tamanhos de efeito bem grandes. Para as duas comparações que fizemos para a variável (Saída 11.17 do SPSS), temos:

$$r_{\text{Positiva vs. Neutra}} = \sqrt{\frac{142,19}{142,19 + 19}} = 0,94$$

$$r_{\text{Negativa vs. Neutra}} = \sqrt{\frac{47,07}{47,07 + 19}} = 0,84$$

Novamente, ambas as comparações produzem tamanhos de efeito muito grandes. Para o termo de interação temos quatro contrastes, mas podemos convertê-los em r porque todos têm 1 grau de liberdade para o modelo (Saída 11.17 do SPSS):

$$r_{\text{Cerveja vs. água, Positiva vs. Neutra}} = \sqrt{\frac{1,58}{1,58 + 19}} = 0,28$$

$$r_{\text{Cerveja vs. Água, Negativa vs. Neutra}} = \sqrt{\frac{6,75}{6,75 + 19}} = 0,51$$

$$r_{\text{Vinho vs. Água, Positiva vs. Neutra}} = \sqrt{\frac{0,24}{0,24 + 19}} = 0,11$$

$$r_{\text{Vinho vs. Água, Negativa vs. Neutra}} = \sqrt{\frac{26,91}{26,91 + 19}} = 0,77$$



Assim, os dois efeitos que eram significativos (cerveja *versus* água, negativa *versus* neutra e vinho *versus* água, negativa *versus* neutra) produzem tamanhos de efeito muito grandes. Os dois efeitos que não eram significativos produziram um tamanho de efeito médio (cerveja *versus* água, positiva *versus* neutra) e um tamanho de efeito pequeno (vinho *versus* água, positiva *versus* neutra).

11.11 RELATANDO OS RESULTADOS DA ANOVA FATORIAL DE MEDIDAS REPETIDAS ②

Podemos relatar uma ANOVA fatorial de medidas repetidas da mesma forma que relatamos qualquer outra ANOVA. Lembre que temos três efeitos para relatar e esses efeitos podem ter diferentes graus de liberdade. Para os efeitos principais de bebida e imagens, a hipótese de esfericidade foi violada e, assim, temos que relatar os graus de liberdade corrigidos de Greenhouse e Geisser. Podemos, portanto, começar relatando a violação da esfericidade:

- ✓ O teste de Mauchly indica que a hipótese de esfericidade foi violada para o efeito principal da bebida $\chi^2(2) = 23,75$, $p < 0,001$ e imagem $\chi^2(2) = 7,42$, $p < 0,05$. Portanto, os graus de liberdade foram corrigidos utilizando as estimativas de esfericidade de Greenhouse e Geisser ($\epsilon = 0,58$ para o efeito principal da bebida e 0,75 para o efeito principal da imagem).

Podemos relatar os três efeitos dessa análise como segue:

- ✓ Todos os efeitos são relatados como significativos em $p < 0,05$. Havia um efeito principal significativo do tipo de bebida nas avaliações da bebida, $F(1,15, 21,93) = 5,11$. Os contrastes revelaram que as avaliações para cerveja $F(1, 19) = 6,22$, $r = 0,50$ e vinho $F(1,19) = 18,61$, $r = 0,70$ eram significativamente mais altos do que para a água.

- ✓ Houve, também, um efeito significativo do tipo de imagem na avaliação das bebidas, $F(1,50, 28,40) = 122,57$. Os contrastes revelaram que as avaliações após a exibição de imagens positivas foram significativamente maiores do que após imagens neutras, $F(1,19) = 142,19$, $r = 0,94$. Já após imagens negativas elas foram significativamente mais baixas do que com as imagens neutras, $F(1,19) = 47,07$, $r = 0,84$.
- ✓ Houve uma interação significativa entre o tipo de bebida e o tipo de imagem usada, $F(4, 76) = 17,16$. Isso indica que as imagens tinham diferentes efeitos nas avaliações das pessoas dependendo do tipo de bebida que foi usado. Para analisar essa interação, contrastes foram realizados comparando todos os tipos de bebidas à sua base (água) e todos os tipos de imagens à sua base (imagens neutras). Esses contrastes revelaram interações significativas quando comparadas imagens negativas a imagens neutras tanto para a cerveja em relação à água, $F(1, 19) = 6,75$, $r = 0,51$, quanto para o vinho em relação à água, $F(1, 19) = 26,91$, $r = 0,77$. Observando o gráfico de interação, esses efeitos refletem que imagens negativas (comparadas às neutras) baixaram os escores significativamente mais para a água do que para cerveja e baixaram os escores significativamente mais para o vinho do que para a água. Os contrastes restantes não revelaram termos de interação significativos quando comparamos imagens positivas às imagens neutras para cerveja comparada à água, $F(1, 19) = 1,58$, $r = 0,28$ e para o vinho comparado à água, $F(1, 19) < 1$, $r = 0,11$. Entretanto, esses contrastes produziram tamanhos de efeito pequenos para médios.

11.12 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

Este capítulo nos ajudou a entender os delineamentos das medidas repetidas. A primeira coisa que aprendemos foi que com os

delineamentos de medidas repetidas existe, ainda, outra hipótese para nos preocuparmos: *esfericidade*. Tendo nos recuperado dessa revelação chocante, tivemos sorte de descobrir que essa hipótese, se violada, pode ser facilmente remediada. Em seguida, analisamos a teoria da ANOVA para medidas repetidas para uma variável de fator independente. Embora não tenha sido essencial, esse foi um exercício útil para demonstrar que aquela teoria é basicamente a mesma quando temos um delineamento independente (bem, existem algumas diferenças sutis, mas eu destaquei as semelhanças). Depois, vimos um exemplo no SPSS, antes de enfrentarmos o ômega ao quadrado. Lamento ter entrado nesse assunto, mas teimoso como sempre, perseverei. Isso nos levou a um delineamento fatorial de medidas repetidas e especificamente a situação onde temos 2 variáveis independentes. Aprendemos que, assim como com outros delineamentos fatoriais, temos que nos preocupar com os termos de interação. Mas também descobrimos algumas maneiras úteis de analisar esses termos usando contrastes. Se você pensou que tudo aquilo era ruim, irei acrescentar mais uma complicação nessa mistura. Podemos combinar medidas repetidas e delineamentos independentes, e o próximo capítulo analisa essa situação. Como se isso não fosse ruim o suficiente, também mostrarei um delineamento com três variáveis independentes (neste momento estou sentado na minha cadeira com os olhos vidrados babando e rindo como um maníaco).

11.13 TERMOS-CHAVE QUE DESCOBRIMOS

- Simetria composta
- Correção de Greenhouse e Geisser
- Correção de Huynh e Feldt
- Estimativas do limite inferior
- Teste de Mauchly
- ANOVA de medidas repetidas
- Esfericidade

11.14 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Imagine que desejo observar o efeito que o álcool tem sobre um “olhar de cobiça”. O efeito “olhar de cobiça” é a tendência das pessoas em relacionamentos de “conferir” o sexo oposto. Selecionei 20 homens e dei a eles óculos incrivelmente sofisticados que poderiam rastrear os movimentos dos seus olhos e gravar tanto o movimento dos seus olhos quanto o objeto sendo observado (esse é o ponto onde fica óbvio que estou inventando tudo isso). Durante quatro noites forneci a essas pobres almas 1, 2, 3 ou 4 canecas de uma cerveja forte numa boate. A cada noite, mesurei quantas mulheres diferentes eles conferiam (uma mulher era considerada conferida se o homem a olhou da cabeça aos pés à cabeça novamente). Para validar essa medida, também coletei a quantidade de baba no queixo do homem enquanto estava olhando para uma mulher. Os dados estão no arquivo **RovingEye.sav**; analise-os com uma ANOVA de um fator. ②
- **Tarefa 2:** No capítulo anterior, vimos o efeito álcool (*beer-goggles*): uma severa distorção da percepção após beber grandes quantidades de álcool. A distorção visual específica é que pessoas anteriormente consideradas não atraentes tornam-se bonitas. Resumindo, num minuto você está no zoológico admirando os orangotangos e no próximo você está pensando por que alguém colocaria Gail Porter (ou seja lá o nome que ela tem agora) numa jaula. De qualquer modo, naquele capítulo, um conjunto de dados descaradamente fabricado demonstrou que o efeito álcool era mais forte nos homens do que nas mulheres e fazia efeito somente após 2 canecas de cerveja. Imagine que queremos continuar com essas descobertas para observar em que fatores esse efeito interfere. Especificamente, achamos que o efeito álcool pode piorar pelo fato de que ele geralmente ocorre em boates com

pouca luz. Pegamos uma amostra de 26 homens (porque o efeito é mais forte nos homens) e demos a eles várias doses de álcool durante quatro semanas diferentes (nenhuma bebida, 2 canecas, 4 canecas e 6 canecas de uma cerveja forte). Essa é a nossa primeira variável independente, a qual chamaremos de consumo de álcool, e tem quatro níveis. Cada semana (e, portanto, em cada estágio de embriaguez), eles irão selecionar uma acompanhante numa boate normal (que tenha pouca iluminação) e, então, selecionar uma segunda acompanhante numa boate especialmente escolhida que tenha ótima iluminação. Desse modo, a segunda variável independente foi se o clube tinha pouca ou muita iluminação. A saída da mensuração foi a atratividade de cada acompanhante avaliada por um grupo de juízes independentes. Recapitulando, todos os participantes tomaram parte em todos os níveis na variável de consumo do álcool e selecionaram as acompanhantes em boates com muita ou pouca iluminação. Os dados estão no arquivo **BeerGogglesLighting.sav**; analise-os com uma ANOVA de dois fatores de medidas repetidas. ②

- **Tarefa 3:** Usando o Quadro 11.1, mude a sintaxe em **SimpleEffectsAttitude.sps** para observar o efeito de bebida em níveis diferentes de imagens. ③

11.15 LEITURAS COMPLEMENTARES

- FIELD, A. P. A bluffer's guide to sphericity. *Newsletter of the Mathematical, Statistical and Computing section of the British Psychological Society*. v. 6, n. 1, p. 13-22. 1988. (disponível no site www.artmed.com.br como **sphericity.pdf**).
- HOWELL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 2002. 5th ed.
- ROSENTHAL, R., ROSNOW, R. L., RUBIN, D. B. *Contrasts and effect sizes in behavioural research: a correctional approach*. Cambridge: Cambridge University Press. 2000. Este é bastante avançado, mas não poderia ser melhor em termos de contrastes e estimativa de tamanho de efeitos.

ANOVA COM DELINEAMENTOS MISTOS (MLG 5)

12.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②



O último delineamento que vou abordar é um em que temos uma mistura de variáveis entre grupos e de medidas repetidas: um *delineamento misto*. Para ter esse delineamento, obviamente é necessário pelo menos duas variáveis independentes, mas você pode ter cenários mais complexos também (por exemplo, duas entre grupos e uma de medidas repetidas, uma entre grupos e duas de medidas repetidas ou, até mesmo, duas de cada). O SPSS permite que você teste quase todos os delineamentos que quiser, praticamente com qualquer grau de complexidade. Entretanto, se os termos de interação são difíceis de interpretar com apenas duas variáveis, então imagine a dificuldade se incluirmos quatro! O melhor conselho que posso oferecer é permanecer com três ou menos variáveis independentes se você quiser conseguir interpretar os termos de interação¹ e, certa-

mente, não utilizar mais do que quatro, a não ser que queira uma bela dor de cabeça.

Este capítulo irá analisar um exemplo de uma *ANOVA mista*. Não haverá teoria alguma porque você definitivamente já teve teoria suficiente sobre a ANOVA (você pode interpretar isso como “É muito complexo para mim e eu irei disfarçar minha incompetência fazendo de conta que você não precisa saber sobre isso”). Vamos analisar um exemplo usando o SPSS e, depois, interpretar a saída. Com sorte, você aprenderá algo sobre interações e como analisá-las utilizando contrastes.

12.2 O QUE HOMENS E MULHERES PROCURAM EM UM COMPANHEIRO? ②

O exemplo que iremos usar neste capítulo continua com o tema do namoro. Muitas revistas falam sobre como homens e mulheres querem coisas diferentes de um relacionamento (ou talvez seja só a revista *Marie Claire* da minha namorada, que não leio – mesmo). A grande pergunta é: o que é mais importante, a aparência ou a personalidade? Imagine que você queira usar isso em um teste. Você idealiza um plano engenhoso onde organiza uma noite de encon-

¹ Apreciadores de ironia irão gostar da ANOVA de quatro fatores que realizei em Field e Davey (1999) e em muitas outras publicações!

tros rápidos (*speed dating*).² Mal sabem as pessoas que compareceram ao encontro que você chamou alguns dos seus amigos como pares. Especificamente, você colocou nove homens e nove mulheres para atuarem como pares. Em cada um desses grupos, três pessoas eram extremamente atraentes, mas diferiam em personalidade: uma tinha muito carisma³, outra tinha pouco carisma e a terceira era tão entediante quanto este livro. Outras três pessoas tinham uma atratividade mediana e diferiam, novamente, quanto a sua personalidade: uma era muito carismática, a outra tinha um pouco de carisma e a terceira pessoa era muito chata. As três últimas eram, não querendo ser indelicado, muito feias e, novamente, uma era carismática, outra tinha pouco carisma e a última pobre alma era um tédio. Os participantes foram as pessoas que se interessaram por essa noite de “encontros rápidos” e durante toda a noite eles passaram por todas as nove pessoas do sexo oposto que foram selecionadas para eles. Após uma conversa de 5 minutos, eles avaliaram o quanto gostariam de ter um encontro de fato com a pessoa em uma escala percentual (100% = “Eu pagaria pelo seu telefone”, 0% = “Eu pagaria por uma passagem de avião que me levasse para bem longe de você”). Cada participante avaliou nove pessoas diferentes que variaram quanto a atratividade e personalidade. Assim, existem duas variáveis de medidas repetidas: **looks** (aparência) (com três níveis, pois a pessoa pode ser atraente, mediana ou feia) e **personality** (personalidade) (novamente, com três níveis: muito carisma, pouco

carisma ou um tédio). Obviamente, as pessoas avaliando poderiam ser homens ou mulheres, assim, devemos também incluir o gênero (homem ou mulher); isso, é claro, será uma variável entre grupos. Os dados estão na Tabela 12.1.

12.3 ANOVA MISTA NO SPSS ②

12.3.1 A análise principal ②

Para entrar com esses dados no SPSS, usamos o mesmo procedimento realizado na ANOVA de dois fatores de medidas repetidas que vimos no capítulo anterior. Lembre que cada linha no editor de dados representa os dados de um único participante. Se a pessoa participar em todas as condições experimentais (nesse caso, elas passam por todas as pessoas que diferem em atratividade e todas as pessoas que diferem em carisma), cada condição experimental deve ser representada por uma coluna no editor de dados. Nesse experimento, existem nove condições experimentais, assim, os dados precisam ser colocados em nove colunas (o formato é idêntico ao da Tabela 12.1). Portanto, crie as seguintes nove variáveis no editor de dados com os nomes conforme determinado. Para cada uma, você também deve entrar um nome completo de variável (veja a Seção 2.4.2) para uma maior clareza na saída.

att_high	Attactive (Atraente)	+ High Charisma (Muito carisma)
av_high	Average Looks (Aparência mediana)	+ High Charisma
ug_high	Ugly (Feio)	+ High Charisma
att_some	Attractive	+ Some Charisma (Pouco carisma)
av_some	Average Looks	+ Some Charisma
ug_some	Ugly	+ Some Charisma
att_none	Attractive	+ Dullard (Chato)
av_none	Average Looks	+ Dullard
ug_none	Ugly	+ Dullard

Depois que as variáveis foram criadas, entre com os dados conforme apresentado na Tabela 12.1. Se você tiver problemas para digitar os da-

² No caso dos encontros rápidos (*speed dating*) terem saído de moda e ninguém saber do que eu estou falando, a ideia básica é: muitos homens e mulheres se encontrando em algum lugar (ou apenas homens ou apenas mulheres se for uma noite *gay*). Metade do grupo senta sozinho em pequenas mesas e o restante escolhe uma dessas mesas. Estes têm 3 minutos para impressionar a outra pessoa à mesa; quando o tempo acaba, uma campainha toca e eles se levantam e vão para outra mesa. Depois de passar por todas as mesas, os participantes terminam a noite ou perseguindo a pessoa que mais gostaram ou evitando algum excêntrico.

³ As pessoas muito atraentes e altamente carismáticas foram, é claro, jogadas de um penhasco porque a vida já é difícil o suficiente sem essas pessoas em volta fazendo você se sentir inadequado.

Tabela 12.1 Dados de **LooksOrPersonality.sav** (Att = Atraente, Av = Mediano, Ug = Feio)

Aparência	Muito Carisma			Pouco Carisma			Chato		
	Att	Av	Ug	Att	Av	Ug	Att	Av	Ug
Homem	86	84	67	88	69	50	97	48	47
	91	83	53	83	74	48	86	50	46
	89	88	48	99	70	48	90	45	48
	89	69	58	86	77	40	87	47	53
	80	81	57	88	71	50	82	50	45
	80	84	51	96	63	42	92	48	43
	89	85	61	87	79	44	86	50	45
	100	94	56	86	71	54	84	54	47
	90	74	54	92	71	58	78	38	45
Mulher	89	86	63	80	73	49	91	48	39
	89	91	93	88	65	54	55	48	52
	84	90	85	95	70	60	50	44	45
	99	100	89	80	79	53	51	48	44
	86	89	83	86	74	58	52	48	47
	89	87	80	83	74	43	58	50	48
	80	81	79	86	59	47	51	47	40
	82	92	85	81	66	47	50	45	47
	97	69	87	95	72	51	45	48	46
	95	92	90	98	64	53	54	53	45
	95	93	96	79	66	46	52	39	47

dos, use o arquivo **LooksOrPersonality.sav**. Primeiro temos que definir nossas variáveis de medidas repetidas, assim, acesse a caixa de diálogo **define factors** (definir fatores) usando o menu **Analyze⇒General Linear Model⇒Repeated Measures...** (Analisar⇒Modelo Linear Geral⇒Medidas Repetidas...). Da mesma forma que com a ANOVA de dois fatores de medidas repetidas (veja o capítulo anterior), precisamos nomear as nossas variáveis de medidas repetidas e especificar quantos níveis cada uma tem. Nesse caso, existem dois fatores entre sujeitos: **looks** (aparência) (atraente, mediano ou feio) e **charisma** (carisma) (muito carisma, pouco carisma e chato). Na caixa de diálogo **define factors**, troque a palavra **factor 1** pela palavra **looks**. Quando você tiver dado a esse fator de medidas repetidas um nome, diga ao computador que essa variável tem três níveis digitando o número 3 dentro da caixa **Number of Levels** (Número de níveis). Clique em **Add** (adicio-

nar) para incluir essa variável na lista de variáveis de medidas repetidas; ela irá aparecer no quadro em branco no final da caixa de diálogo como **looks** (3) (aparência). Agora, repita esse processo com a segunda variável independente. Insira a palavra **charisma** no espaço rotulado **Within-Subject Factor Name** e então, devido aos 3 níveis dessa variável, insira o número 3 no espaço rotulado **Number of Levels**. Clique em **Add** (incluir) para incluir essa variável na lista de fatores, que aparecerá como **charisma** (3). A caixa de diálogo completa é apresentada na Figura 12.1. Depois de entrar com ambos os fatores dentre sujeitos, clique em **Define** (Definir) para ir à caixa de diálogo principal.

A caixa de diálogo principal é a mesma de quando fizemos a ANOVA fatorial de medidas repetidas no capítulo anterior (veja a Figura 12.2). No alto da caixa **Within-Subjects Variables** (Variáveis entre participantes), o SPSS afirma que existem dois fatores: **looks**

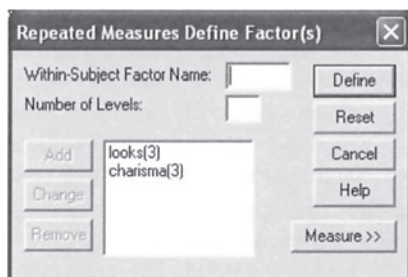


Figura 12.1 Caixa de diálogo *define factors* (definir fatores) para a ANOVA de medidas repetidas.

e **charisma**. Na caixa abaixo, existe uma série de pontos de interrogação seguidos por número entre parênteses. Os números entre parênteses representam os níveis dos fatores (variáveis independentes) – veja o capítulo anterior para uma explicação mais detalhada.

Nesse exemplo, existem duas variáveis independentes, portanto, existem dois números nos parênteses. O primeiro número se refere aos níveis do primeiro fator listado abaixo do quadro (nesse caso, **looks**). O segundo número se refere aos níveis do segundo fator acima do quadro (nesse caso, **charisma**). Assim como na outra ANOVA de medidas repetidas que vimos, precisamos substituir os pontos de interrogação pelas variáveis listadas no quadro esquerdo da caixa

de diálogo. No delineamento entre grupos, em que códigos variáveis são usados, os níveis de um fator particular são especificados pelos códigos designados a eles no editor de dados. Entretanto, em um delineamento de medidas repetidas, esse esquema de códigos não é utilizado, desse modo, determinamos que condição atribuir a um nível nesse estágio (novamente, retorne ao capítulo anterior para mais informações). Por isso, é imperativo que pensemos sobre o tipo de contrastes que queremos fazer *antes* de entrar com as variáveis na caixa de diálogo. Nesse experimento, se observarmos a primeira variável, **looks**, veremos que existem três condições: atraente, mediano e feio. De certo modo, faz sentido comparar as condições atraente e feio à mediano porque uma pessoa mediana representa a maioria (embora não fosse errado comparar, por exemplo, atraente e mediana à categoria feia). Essa comparação pode ser feita especificando um contraste simples (veja a Tabela 8.6), desde que tenhamos certeza de que a categoria mediana seja codificada como nossa primeira ou última categoria. Agora, vamos pensar sobre o nosso segundo fator. O fator **charisma** também tem uma categoria que representa a maioria, que é ter pouco carisma. Novamente, podemos usar isso como um controle para a comparação dos nossos dois extremos (muito carisma e chato). Portanto, podemos conduzir um contraste sim-

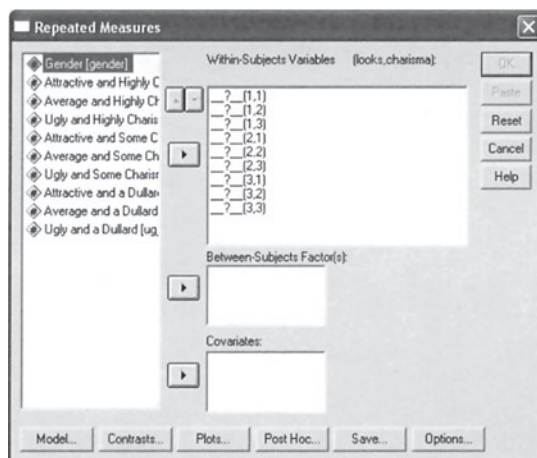












Figura 12.2 Definição dos níveis dos fatores para a ANOVA de medidas repetidas.

ples comparando todas as categorias com a de pouco carisma; isso precisa ser colocado como o primeiro ou como o último nível.

Com base no que foi discutido sobre o uso de contrastes, faz sentido ter a categoria mediana como o nível 3 do fator **looks** e pouco carisma como o terceiro nível do fator **charisma**. Os níveis restantes podem ser decididos arbitrariamente. Escolhi atraente como nível 1 e feio como nível 2 do fator **looks**. Para a variável **charisma**, escolhi muito carisma como nível 1 e chato como nível 2. Essas decisões significam que as variáveis devem entrar da seguinte forma:

att_high		_?(1,1)
att_none		_?(1,2)
att_some		_?(1,3)
ug_high		_?(2,1)
ug_none		_?(2,2)
ug_some		_?(2,3)
av_high		_?(3,1)
av_none		_?(3,2)
av_some		_?(3,3)

Diferentemente do capítulo anterior, troquei a ordem de entradas das variáveis no editor de dados. Fiz isso apenas para ilustrar que podemos entrar com as variáveis na ordem que quisermos. Até agora, o procedimento tem sido similar aos dos demais delineamentos de medidas repetidas. Entretanto, temos um delineamento misto aqui, desse modo, precisamos especificar ainda o fator entre grupos. Para fazer isso, selecione **gender** (gênero) na lista de variáveis e clique em  a fim de transferi-la para o quadro **Between-Subjects Factors** (Fatores entre participantes). A caixa de diálogo completa deve ser parecida com a da Figura 12.3. Já discuti as opções para os botões na parte de baixo da caixa de diálogo, assim, falarei somente daqueles importantes para esse exemplo.

12.3.2 Outras opções ②

Depois da análise principal, é interessante comparar os níveis das variáveis independentes para ver no que elas diferem. Como vimos, não existe um recurso para entrar com os códigos dos contrastes (a não ser que você use a sintaxe), portanto, temos que contar com os contrastes

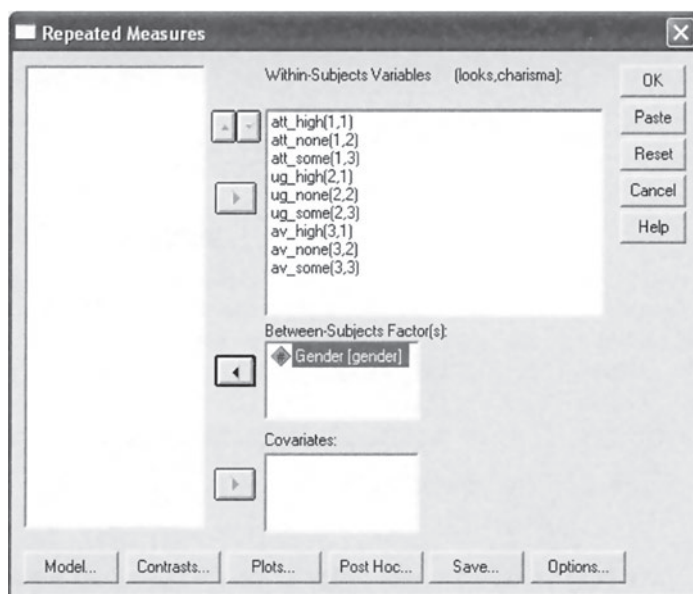


Figura 12.3 Caixa de diálogo completa para a ANOVA de delineamento misto.

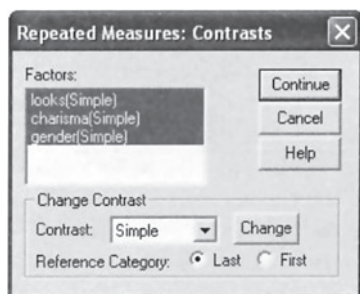


Figura 12.4 Contrastes para a ANOVA de delineamentos mistos.

disponíveis (veja a Tabela 8.6). A Figura 12.4 mostra a caixa de diálogo para conduzir contrastes e é obtida clicando em **Contrasts** (Contrastes) na caixa de diálogo principal. Na seção anterior, expliquei por que é interessante usar as condições atratividade mediana e pouco carisma como categorias base para os fatores **looks** e **charisma**, respectivamente. Usamos a caixa de diálogo **contrasts** (Contrastes) antes, nas Seções 9.3.3 e 11.4.2, assim, apenas direi que você deve selecionar um contraste simples para cada variável independente. Para ambas variáveis independentes, entramos com as variáveis de tal forma que a categoria de controle era a última, portanto, não precisamos trocar a categoria de referência para o contraste simples. Depois que os contrastes forem selecionados, clique em **Continue** (continuar) para retornar à caixa de diálogo principal.

Gender (gênero) tem somente dois níveis (masculino ou feminino), portanto, não precisamos especificar contrastes para essa variável. A adição do fator entre grupos também significa que podemos selecionar testes *post hoc* para essa variável clicando em **Post Hoc...**. Essa ação apresenta a caixa de diálogo testes *post hoc* (veja a Seção 8.3.3), que pode ser usada como explicado anteriormente. Entretanto, não precisamos especificar um teste *post hoc* aqui porque o fator entre grupos tem somente dois níveis.

Com a adição de uma variável extra, é necessário escolher um gráfico diferente daquele do exemplo capítulo anterior. Clique em **Plots...** (Diagramas) para acessar uma caixa de diálogo semelhante à da Figura 12.5. Coloque **looks** no espaço **Horizontal Axis** (Eixo horizontal) e

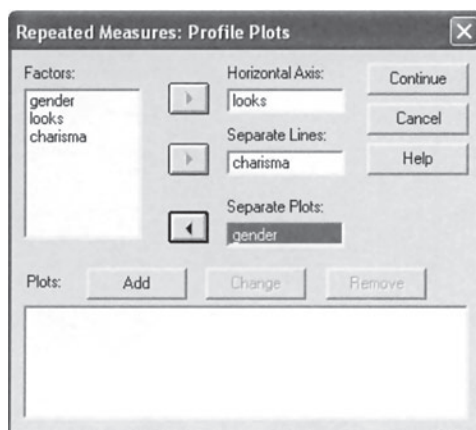


Figura 12.5 Caixa de diálogo diagramas (*plots*) para a ANOVA mista de três fatores.

charisma no espaço **Separate Line** (Linha separada); finalmente, coloque **gender** no espaço **Separate Plots** (Diagramas separados). Quando as três variáveis forem especificadas, não se esqueça de clicar em **Add** (adicionar) para acrescentar essa combinação à lista dos diagramas. Solicitando ao SPSS para traçar a interação $looks \times charisma \times gender$, teremos o gráfico de interação para **looks** e **charisma**, mas uma versão separada desse gráfico será produzida para participantes homens e mulheres.

No que se refere às outras opções, você deve selecionar as mesmas que foram escolhidas para o exemplo anterior (veja a Seção 11.8.4). Vale a pena selecionar médias marginais estimadas para todos os efeitos (porque esses valores irão ajudá-lo a entender qualquer efeito significativo), mas para economizar espaço, não solicitei intervalos de confiança para esses efeitos porque já vimos, com alguns detalhes, essa parte da saída. Quando todas as opções apropriadas forem selecionadas, execute a análise.

12.4 SAÍDAS PARA A ANOVA MISTA FATORIAL: ANÁLISE PRINCIPAL ③

A saída inicial é a mesma do exemplo da ANOVA de dois fatores: existe uma tabela listando as variáveis de medidas repetidas do editor de dados e o nível de cada variável in-

dependente que elas representam. A segunda tabela contém estatísticas descritivas (médias e desvios padrão) para cada uma das nove condições divididas entre homens ou mulheres (veja

a Saída 12.1 do SPSS). Os nomes nessa tabela são os que dei às variáveis no editor de dados (portanto, sua saída pode diferir da minha). Essas estatísticas descritivas são interessantes

Saída 12.1 do SPSS

Within-Subjects Factors (Fatores entre Participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

LOOKS (Aparência)	CHARISMA (Carisma)	Dependent Variable (Variável Dependente)
1	1	ATT_HIGH (Muito Atraente)
	2	ATT_NONE (Nada Atraente)
	3	ATT_SOME (Pouco Atraente)
2	1	UG_HIGH (Muito Feio)
	2	UG_NONE (Nada Feio)
	3	UG_SOME (Pouco Feio)
3	1	AV_HIGH (Muito Mediano)
	2	AV_NONE (Nada Mediano)
	3	AV_SOME (Pouco Mediano)

Descriptive Statistics (Estatísticas Descritivas)

	Gender (Gênero)	Mean (Média)	Std. Deviation (Desvio padrão)	N
Attractive and High Charismatic (Atraente e Muito Carismático)	Male (Masculino)	88.30	5.697	10
	Female (Feminino)	89.60	6.637	10
	Total (Total)	88.95	6.057	20
Attractive and Dullard (Atraente e Chato)	Male (Masculino)	87.30	5.438	10
	Female (Feminino)	51.80	3.458	10
	Total (Total)	69.55	18.743	20
Attractive and Some Charisma (Atraente e Pouco Carismático)	Male (Masculino)	88.50	5.740	10
	Female (Feminino)	87.10	6.806	10
	Total (Total)	87.80	6.170	20
Ugly and Highly Charismatic (Feio e Muito Carismático)	Male (Masculino)	56.80	5.731	10
	Female (Feminino)	86.70	5.438	10
	Total (Total)	71.75	16.274	20
Ugly and Dullard (Feio e Chato)	Male (Masculino)	45.80	3.584	10
	Female (Feminino)	46.10	3.071	10
	Total (Total)	45.95	3.252	20
Ugly and Some Charisma (Feio e Pouco Carismático)	Male (Masculino)	48.30	5.376	10
	Female (Feminino)	51.20	5.453	10
	Total (Total)	49.75	5.476	20
Average and Highly Charismatic (Mediano e Muito Carismático)	Male (Masculino)	82.80	7.005	10
	Female (Feminino)	88.40	8.329	10
	Total (Total)	85.60	8.022	20
Average and Dullard (Mediano e Chato)	Male (Masculino)	47.80	4.185	10
	Female (Feminino)	47.00	3.742	10
	Total (Total)	47.40	3.885	20
Average and Some Charisma (Mediano e Pouco Carismático)	Male (Masculino)	71.80	4.417	10
	Female (Feminino)	68.90	5.953	10
	Total (Total)	70.35	5.314	20

porque elas nos mostram o padrão das médias através das condições experimentais (assim, usamos essas médias para produzir os gráficos da interação de três fatores).

A Saída 12.2 do SPSS mostra os resultados do teste de esfericidade de Mauchly para cada um dos três efeitos de medidas repetidas no modelo. Nenhum dos efeitos viola a hipótese de esfericidade porque todos os valores na coluna denominada *Sig.* estão acima de 0,05; portanto, podemos assumir esfericidade quando olhamos as nossas estatísticas *F*.

A Saída 12.3 do SPSS mostra a tabela resumo dos efeitos das medidas repetidas na ANOVA com os valores *F* corrigidos. Assim como com a ANOVA fatorial de medidas repetidas, a saída é dividida em seções para cada um dos efeitos no modelo e seus termos erro associados. A única diferença é que as interações entre nossa variável de gênero entre grupos e os efeitos de medidas repetidas também estão incluídos.

Novamente, precisamos olhar a coluna *Sig.* e, se os valores nessa coluna são menores do que 0,05 para um efeito particular, então ele é estatisticamente significativo. Examinando a tabela a partir do topo, encontramos um efeito significativo de *looks*, informando que se ignorarmos que o acompanhante era carismático e se a avaliação foi de um homem

ou uma mulher, a atratividade de uma pessoa afetou significativamente as avaliações que eles receberam. A interação *looks* \times *gender* também é significativa, significando que a maneira como as avaliações foram afetadas pela atratividade foi diferente para avaliadores homens e mulheres.

A seguir, encontramos um efeito significativo de carisma, informando que se ignorarmos que o acompanhante era atraente e se a avaliação era de um homem ou uma mulher, o carisma de uma pessoa significativamente afetou as avaliações que ela recebeu. A interação *charisma* \times *gender* é, também, significativa, assim, a maneira como as avaliações foram afetadas pelo carisma foi diferente entre avaliadores homens e mulheres.

Em seguida, encontramos uma interação significativa entre *looks* e *charisma*, mostrando que se ignorarmos o gênero do avaliador, o perfil das avaliações através dos diferentes níveis de atratividade foi diferente para acompanhantes muito carismáticos, acompanhantes pouco carismáticos e chatos. (O contrário também é verdadeiro: o perfil das avaliações através de diferentes níveis de carisma foi diferente para acompanhantes atraentes, medianos e feios). Para nos confundirmos ainda mais, a interação *looks* \times *charisma* \times *gender* também é significativa, mostrando que *looks*

Saída do SPSS 12.2

(Mauchly's Test of Sphericity^a (Teste de Mauchly da Esfericidade)

Measure: Measure_1 (Medida: Medida_1)

Within-Subjects Effect (Efeito dentre participantes)	Mauchly's W (W de Mauchly)	Approx. Chi-Square (Qui-Quadrado Aproximado)	df (gl)	Sig. (Sig.)	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound (Limite inferior)
LOOKS (Aparência)	0.960	0.690	2	0.708	0.962	1.000	0.500
CHARISMA (Carisma)	0.929	1.246	2	0.536	0.934	1.000	0.500
LOOKS*CHARISMA (Aparência*Carisma)	0.613	8.025	9	0.534	0.799	1.000	0.250

Testa a hipótese nula de que a matriz de covariância dos erros das variáveis dependentes transformadas e ortonormalizadas é proporcional a uma matriz identidade.

a Pode ser utilizado para ajustar os graus de liberdade dos testes de significância para médias. Os testes corrigidos são mostrados em camadas (por padrão) na tabela dos testes de efeitos dentre participantes.

b Delineamento: Intercepto + GÊNERO.

Delineamento dentre participantes: APARÊNCIA + CARISMA + APARÊNCIA*CARISMA.

Saída 12.3 do SPSS

Tests of Within-Subjects Effects (Testes de Efeitos entre Participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)		<i>Type III Sum of Squares</i> (Soma dos Quadrados do Tipo III)	<i>df</i> (gl)	<i>Mean Square</i> (Média dos Quadrados)	<i>F</i>	<i>Sig.</i>
LOOKS (Aparência)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	20779.633	2	10389.817	423.733	0.000
	Greenhouse-Geisser	20779.633	1.923	10803.275	423.733	0.000
	Huynh-Feldt	20779.633	2.000	10389.817	423.733	0.000
	<i>Lower-bound</i> (Limite Inferior)	20779.633	1.000	20779.633	423.733	0.000
LOOKS * GENDER (Aparência * Gênero)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	3944.100	2	1972.050	80.427	0.000
	Greenhouse-Geisser	3944.100	1.923	2050.527	80.472	0.000
	Huynh-Feldt	3944.100	2.000	1972.050	80.427	0.000
	<i>Lower-bound</i> (Limite Inferior)	3944.100	1.000	3944.100	80.427	0.000
Error(LOOKS) (Erro(Aparência))	<i>Sphericity Assumed</i> (Esfericidade Assumida)	882.711	36	24.520		
	Greenhouse-Geisser	882.711	34.622	25.496		
	Huynh-Feldt	882.711	36.000	24.520		
	<i>Lower-bound</i> (Limite Inferior)	882.711	18.000	49.040		
CHARISMA (Carisma)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	23233.600	2	11616.800	328.250	0.000
	Greenhouse-Geisser	23233.600	1.868	12437.761	328.250	0.000
	Huynh-Feldt	23233.600	2.000	11616.800	328.250	0.000
	<i>Lower-bound</i> (Limite Inferior)	23233.600	1.000	23233.600	328.250	0.000
CARISMA * GENDER (Carisma * Gênero)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	4420.133	2	2210.067	62.449	0.000
	Greenhouse-Geisser	4420.133	1.868	2366.252	62.449	0.000
	Huynh-Feldt	4420.133	2.000	2210.067	62.449	0.000
	<i>Lower-bound</i> (Limite Inferior)	4420.133	1.000	4420.133	62.449	0.000
Error(CHARISMA) (Erro(CARISMA))	<i>Sphericity Assumed</i> (Esfericidade Assumida)	1274.044	36	35.390		
	Greenhouse-Geisser	1274.044	33.624	37.891		
	Huynh-Feldt	1274.044	36.000	35.390		
	<i>Lower-bound</i> (Limite Inferior)	1274.044	18.000	70.780		
LOOKS * CHARISMA (Aparência * Carisma)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	4055.267	4	1013.817	36.633	0.000
	Greenhouse-Geisser	4055.267	3.197	1268.295	36.633	0.000
	Huynh-Feldt	4055.267	4.000	1013.817	36.633	0.000
	<i>Lower-bound</i> (Limite Inferior)	4055.267	1.000	4055.267	36.633	0.000
LOOKS * CHARISMA * GENDER (Aparência * Carisma * Gênero)	<i>Sphericity Assumed</i> (Esfericidade Assumida)	2669.667	4	667.417	24.116	0.000
	Greenhouse-Geisser	2669.667	3.197	834.945	24.116	0.000
	Huynh-Feldt	2669.667	4.000	667.417	24.116	0.000
	<i>Lower-bound</i> (Limite Inferior)	2669.667	1.000	2669.667	24.116	0.000
Error(LOOKS * CHARISMA) (Erro(Aparência * Carisma))	<i>Sphericity Assumed</i> (Esfericidade Assumida)	1992.622	72	27.675		
	Greenhouse-Geisser	1992.622	57.554	34.622		
	Huynh-Feldt	1992.622	72.000	27.675		
	<i>Lower-bound</i> (Limite Inferior)	1992.622	18.000	110.701		

× *charisma* foi significativamente diferente em participantes homens e mulheres!

Isso é muito para se assimilar, assim, iremos analisar esses efeitos em partes nas próximas seções. Primeiro, precisamos entender o que aconteceu com nosso efeito principal de gênero.

12.4.1 O efeito de gênero ②

O efeito principal do gênero está listado separadamente dos efeitos de medidas repetidas em uma tabela denominada **Tests of Between-Subjects Effects** (Testes dos efeitos entre sujeitos). Antes de olharmos para essa tabela, é importante checar as hipóteses de homogeneidade da variância usando o teste de Levene (veja a Seção 3.6).

O SPSS produz uma tabela listando o teste de Levene para cada uma das variáveis de medidas repetidas no editor de dados e deve-

mos procurar por qualquer variável que tenha um valor significativo. A Saída 12.4 do SPSS mostra ambas as tabelas. A tabela mostrando o teste de Levene indica que as variâncias são homogêneas para todos os níveis das variáveis de medidas repetidas (porque todos os valores de significância são maiores do que 0,05). Se qualquer valor fosse significativo, isso poderia comprometer a precisão do teste *F* para gênero e teríamos que considerar a transformação de todos os nossos dados para a estabilização das variâncias entre grupos (veja o Capítulo 3). Felizmente, nesse exemplo a transformação não é necessária. A segunda tabela mostra o resumo da ANOVA para o efeito principal do gênero e isso revela um efeito não-significativo (porque a significância de 0,946 é maior do que o ponto de corte padrão de 0,05).

Podemos relatar que houve um efeito principal não-significativo (*ns*) de gênero,

Saída 12.4 do SPSS

Levene's Test of Equality of Error Variances^a
(Teste de Levene da Igualdade das Variâncias do Erro)

	<i>F</i>	<i>df1</i> (gl1)	<i>df2</i> (gls2)	<i>Sig.</i>
<i>Attractive and High Charismatic</i> (Atraente e Muito Carismático)	1.131	1	18	0.302
<i>Attractive and Dullard</i> (Atraente e Chato)	1.949	1	18	0.180
<i>Attractive and Some Charisma</i> (Atraente e pouco Carismático)	0.599	1	18	0.449
<i>Ugly and Highly Charismatic</i> (Feio e Muito Carismático)	0.005	1	18	0.945
<i>Ugly and Dullard</i> (Feio e Chato)	0.082	1	18	0.778
<i>Ugly and Some Charisma</i> (Feio e pouco Carismático)	0.124	1	18	0.729
<i>Average and Highly Charismatic</i> (Mediano e Muito Carismático)	0.102	1	18	0.753
<i>Average and Dullard</i> (Mediano e Chato)	0.004	1	18	0.950
<i>Average and Some Charisma</i> (Mediano e com Pouco Carismático)	1.763	1	18	0.201

Testa a hipótese nula de que a variância do erro da variável dependente é igual ao longo dos grupos.

a Delineamento: Intercepto + GÊNERO.

Delineamento dentre participantes: APARÊNCIA + CARISMA + APARÊNCIA*CARISMA.

Tests of Within-Subjects Effects (Testes de Efeitos dentre Participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Transformed Variable: Average (Variável transformada: Mediano)

Source (Fonte)	<i>Type III Sum of Squares</i> (Soma dos Quadrados do Tipo III)	<i>df</i> (gl)	<i>Mean Square</i> (Média dos Quadrados)	<i>F</i>	<i>Sig.</i>
<i>Intercept</i> (Intercepto)	94027.756	1	94027.756	20036.900	0.000
<i>GENDER</i> (Gênero)	0.022	1	0.022	0.005	0.946
<i>Error</i> (Erro)	84.469	18	4.693		

$F(1, 18) < 1$, *ns*. Esse efeito nos diz que se ignorarmos todas as outras variáveis, as avaliações dos participantes masculinos foram basicamente iguais às dos participantes femininos. Se você solicitou que o SPSS mostrasse as médias para o efeito do gênero, deve examinar a saída para encontrar a tabela na seção chamada de **Estimated Marginal Means** (Médias marginais estimadas). A Saída 12.5 do SPSS é uma tabela de médias para o efeito principal do gênero com os erros padrão associados. Essa informação é mostrada na Figura

12.6. A partir desse gráfico, fica claro que as avaliações de homens e mulheres foram em geral as mesmas.

12.4.2 O efeito da aparência ②

Vimos o efeito principal da aparência (*looks*) na Saída 12.3 do SPSS. Agora iremos ver o que esse efeito significa. Podemos relatar que havia um efeito principal significativo da aparência (*looks*), $F(2, 36) = 423,73$, $p < 0,001$. Esse efeito nos diz que se ignorarmos todas as

Saída 12.5 do SPSS

1 Gender (Gênero)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Gender (Gênero)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
Male (Masculino)	68,600	0,685	67,161	70,039
Female (Feminino)	68,533	0,685	67,094	69,973

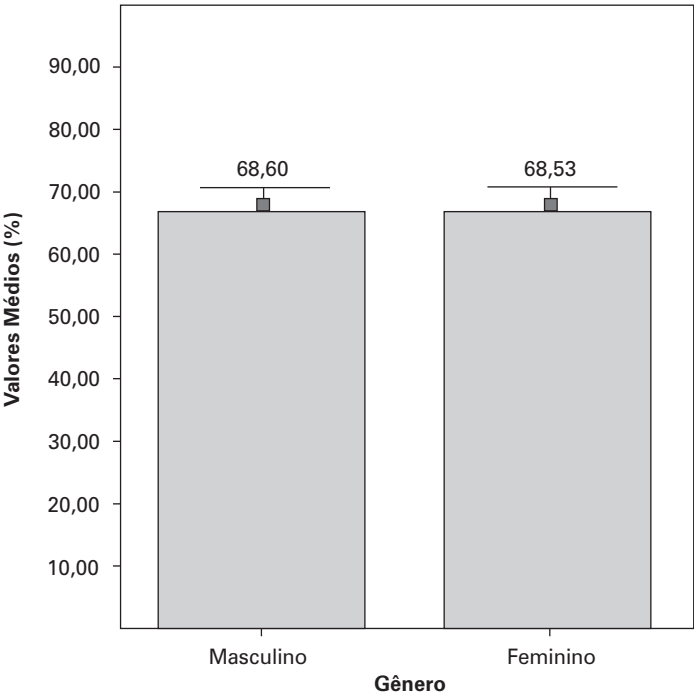


Figura 12.6 Diagrama de Barras de Erro para Gênero mostrando um IC de 95% para a média.

outras variáveis, as avaliações foram diferentes para os acompanhantes atraentes, medianos e feios. Se solicitarmos ao SPSS a exibição das médias para o efeito aparência (de agora em diante, suporei que você fez isso), você encontrará a tabela na seção denominada **Estimated Marginal Means** (Médias marginais estimadas). A Saída 12.6 do SPSS é uma tabela de médias para o efeito principal da aparência com os erros padrão associados. Os níveis da aparência estão rotulados de 1, 2 e 3, e você deve lembrar como entrou com a variável (ou você pode olhar a tabela resumo que o SPSS produz no início da saída – veja a Saída 12.1 do SPSS). Se você seguiu o que eu fiz, o nível 1 é atraente, o nível 2 é feio e o nível 3 é mediano. Para facilitar, essa informação está apresentada na Figura 12.7. Você pode ver que à medida que a atratividade diminui, a média das avaliações também cai. Assim, esse efeito principal parece refletir que os avaliadores tinham maior probabilidade de

Saída do SPSS 12.6

2 Looks (Aparência)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Looks (Aparência)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	82.100	0.652	80.729	83.471
2	55.817	0.651	54.449	57.184
3	67.783	0.820	66.061	69.505

expressar um grande interesse em sair com pessoas atraentes do que medianas ou feias. Entretanto, precisamos olhar alguns contrastes para descobrir exatamente o que está acontecendo.

A Saída 12.7 do SPSS mostra os contrastes que solicitamos. Por enquanto, apenas olhe

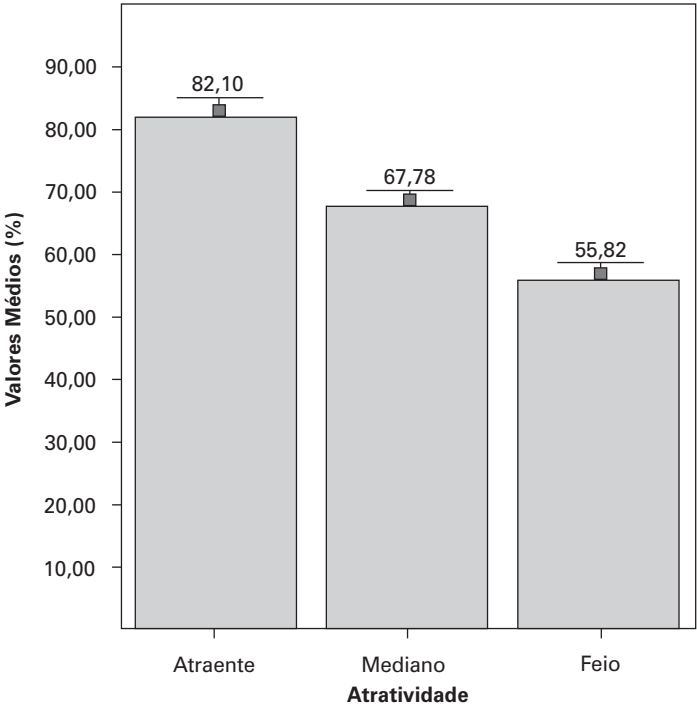


Figura 12.7 Diagrama de Barras de Erro para a Atratividade mostrando um IC de 95% para a média.

Saída 12.7 do SPSS

Tests of Within-Subjects Contrasts (Testes de Contrastes entre Participantes)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Source (Fonte)	LOOKS (Aparência)	CHARISMA (Carisma)	Type III Sum of Squares (Soma dos Quadrados do Tipo III)	df (gl)	Mean Square (Média dos Quadrados)	F	Sig.
LOOKS (Aparência)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		4099.339	1	4099.339	226.986	0.000
	Level 2 vs. Level 3 (Nível 1 vs. Nível 3)		2864.022	1	2864.022	160.067	0.000
LOOKS * GENDER (Aparência * Gênero)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		781.250	1	781.250	43.259	0.000
	Level 2 vs. Level 3 (Nível 1 vs. Nível 3)		540.800	1	540.800	30.225	0.000
Error(LOOKS) (ErroAparência)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)		325.078	18	18.060		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)		322.067	18	17.893		
CHARISMA (Carisma)		Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	3276.800	1	3276.800	109.937	0.000
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	4500.000	1	4500.000	227.941	0.000
CHARISMA * GENDER (Carisma * Gênero)		Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	810.689	1	810.689	27.199	0.000
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	665.089	1	665.089	33.689	0.000
Error(CHARISMA) (ErroCHARISMA)		Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	536.511	18	29.806		
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	355.356	18	19.742		
LOOKS * CHARISMA (Aparência * Carisma)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	3976.200	1	3976.200	21.944	0.000
		Level 2 vs. Level 3 (Nível 1 vs. Nível 3)	441.800	1	441.800	4.091	0.058
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	911.250	1	911.250	6.231	0.022
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	7334.450	1	7334.450	88.598	0.000
LOOKS * CHARISMA * GENDER (Aparência * Carisma * Gênero)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	168.200	1	168.200	0.928	0.348
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	6552.200	1	6552.200	60.669	0.000
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	1711.250	1	1711.250	11.701	0.003
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	110.450	1	110.450	1.3334	0.263
Error(LOOKS * CHARISMA) (ErroAparência * Carisma)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	3261.600	18	181.200		
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	1944.000	18	108.000		
	Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	2632.500	18	146.250		
		Level 2 vs. Level 3 (Nível 2 vs. Nível 3)	1490.100	18	82.783		

para a linha LOOKS (Aparência). Lembre que fizemos um contraste simples, assim, temos um contraste comparando o nível 1 com o nível 3 e, então, comparando o nível 2 com o nível 3. Por causa da ordem em que entramos as variáveis, esses contrastes representam atraente comparado com mediano (nível 1 vs. nível 3) e feio comparado com mediano (nível 2 vs. nível 3). Observando os valores de F para cada contraste e seus valores de significância relacionados, isso nos informa que o efeito da atratividade representa o fato de que acompanhantes atraentes foram avaliados significativamente mais alto do que acompanhantes medianos, $F(1, 18) = 226,99, p < 0,001$, e os acompanhantes medianos foram avaliados significativamente mais alto do que os feios, $F(1, 18) = 160,07, p < 0,001$.

12.4.3 O efeito do carisma ②

O efeito principal do carisma está na Saída 12.3 do SPSS. Podemos relatar que houve um efeito principal significativo do carisma, $F(2, 36) = 328,25, p < 0,001$. Esse efeito nos diz que se ignorarmos todas as outras variáveis, as avaliações diferem em muito carismáticos, pouco carismáticos e chatos. A tabela CHARISMA (Carisma) na seção chamada de **Estimated Marginal Means** (Médias Marginais Estimadas) nos informa o que esse efeito significa (Saída 12.8 do SPSS). Novamente, os níveis do carisma estão classificados como 1, 2 e 3. Se você seguiu o que eu

fiz, o nível 1 é muito carisma, o nível 2 é sem carisma e nível 3 é algum carisma. Essa informação está traçada na Figura 12.8: à medida que o carisma declina, a média das avaliações também diminui. Assim, esse efeito principal parece refletir que os avaliadores tinham maior probabilidade de expressar um interesse maior em sair com pessoas carismáticas do que com pessoas medianas ou chatas.

A Saída 12.7 do SPSS mostra os contrastes que solicitamos. Observando a tabela na linha CHARISMA (Carisma) e lembrando nós solicitamos contrastes simples, conseguimos um contraste comparando o nível 1 com o nível 3 e, então, comparando o nível 2 com o nível 3. Como interpretamos esses contrastes depende da ordem que entramos com as variáveis de medidas repetidas: nesse caso, esses contrastes representam muito carisma comparado com pouco carisma (nível 1 vs. nível 3) e chato comparado com pouco carisma (nível 2 vs. nível 3). Os contrastes nos dizem que o efeito do carisma representou o fato de que acompanhantes com muito carisma foram avaliados significativamente mais alto do que com pouco carisma, $F(1, 18) = 109,94, p < 0,001$ e acompanhantes medianos foram avaliados significativamente mais alto do que os feios, $F(1, 18) = 227,94, p < 0,001$.

12.4.4 A interação entre gênero e aparência ②

A Saída 12.3 do SPSS indica que o gênero interagiu de alguma maneira com a atratividade do acompanhante. Da tabela resumo devemos relatar que houve uma interação significativa entre a atratividade do acompanhante e o gênero do participante, $F(2, 36) = 80,43, p < 0,001$. Esse efeito nos diz que o perfil das avaliações entre os acompanhantes de atratividade diferente foi distinto para homens e mulheres. Podemos usar as médias marginais estimadas para determinar a natureza dessa interação (ou poderíamos ter solicitado ao SPSS um gráfico de *gender* \times *looks* usando uma caixa de diálogo semelhante à da Figura 12.5). As médias e o gráfico de interação (Saída 12.9 do SPSS e Figura 12.9) mostram o significado desse resultado. O gráfico mostra a

Saída 12.8 do SPSS

3 CHARISMA (Carisma)

Measure: MEASURE_1 (Medida: MEDIDA_1)

CHARISMA (Carisma)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
			Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	82.100	1.010	79.978	84.222
2	54.300	0.573	53.096	55.504
3	69.300	0.732	67.763	70.837

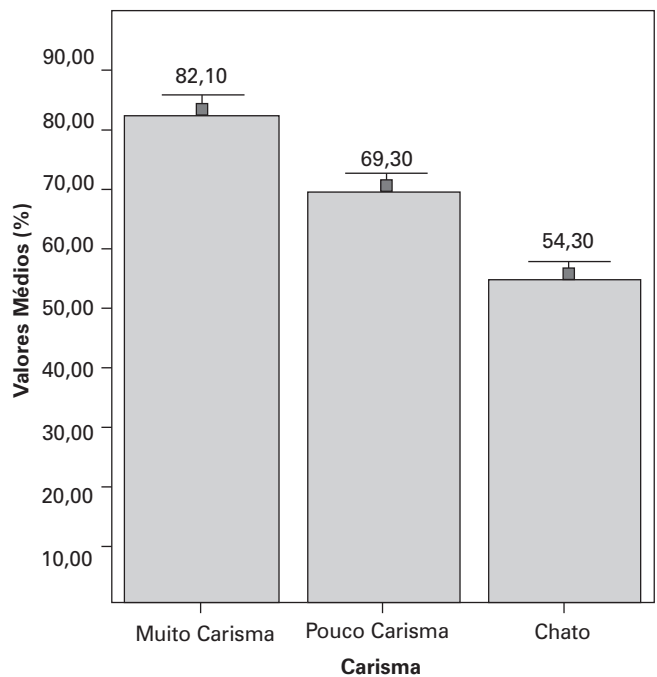


Figura 12.8 Diagrama de Barras de Erro para o Carisma mostrando um IC de 95% para a média.

Saída 12.9 do SPSS

4. Gender * LOOKS (Gênero * Aparência)

Measure: MEASURE_1 (Medida: MEDIDA_1)

GENDER (Gênero)	LOOKS (Aparência)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
				Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
Male (Masculino)	1	88.033	0.923	86.095	89.972
	2	50.300	0.921	48.366	52.234
	3	67.467	1.159	65.031	69.902
Female (Feminino)	1	76.167	0.923	74.228	78.105
	2	61.333	0.921	59.399	63.267
	3	68.100	1.159	65.665	70.535

média das avaliações dos acompanhantes dos homens de diferentes atratividades ignorando quão atrativo o acompanhante era (círculos). Os escores das mulheres são mostrados como quadrados. O gráfico, claramente, mostra que as avaliações de homens e mulheres são muito semelhantes para acompanhantes com aparência mediana, mas os homens dão avaliações mais al-

tas (isto é, eles preferem sair com essas pessoas) que as mulheres para acompanhantes atraentes, enquanto as mulheres expressam maior interesse em sair com pessoas feias do que homens. Em geral, essa interação parece sugerir que o interesse dos homens em sair com uma pessoa é mais influenciado por sua aparência do que para as mulheres. Embora o interesse de homens e

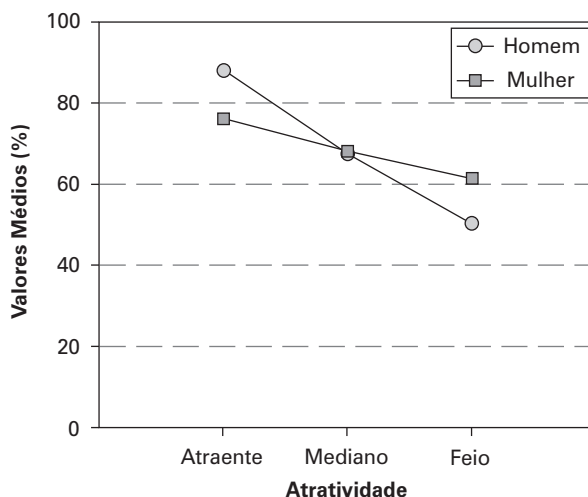


Figura 12.9 Diagrama de Interação entre Gênero e Atratividade.

mulheres diminua à medida que a atratividade diminui, esse decréscimo é mais pronunciado nos homens. Essa interação pode ser esclarecida usando os contrastes especificados antes da análise.

12.4.4.1 Interação 1 aparência × gênero: atrativo × mediano, homem × mulher ②

O primeiro termo de interação analisa o nível 1 da aparência (atraente) comparado ao nível 3 (mediano), comparando os escores de homens e mulheres. O contraste é altamente significativo, $F(1, 18) = 43,26, p < 0,001$. Esse resultado nos mostra que o aumento de interesse em acompanhantes atraentes comparado aos acompanhantes de aparência mediana nos homens é mais significativo do que nas mulheres. Assim, na Figura 12.9 a inclinação da linha entre os círculos representando as avaliações masculinas de acompanhantes atraentes e acompanhantes medianos é mais íngreme do que a linha ligando os quadrados representando as avaliações das mulheres para os acompanhantes atraentes e medianos. Podemos concluir que a preferência por acompanhantes atraentes comparado aos de aparência mediana é maior para os homens do que para as mulheres.

12.4.4.2 Interação 2 aparência × gênero: feio × mediano, homem × mulher ②

O segundo termo de interação compara o nível 2 de aparência (feio) com o nível 3 (mediano), comparando os escores de homens e mulheres. Esse contraste é altamente significativo, $F(1, 18) = 30,23, p < 0,001$. Isso nos diz que o decréscimo do interesse em acompanhantes feios comparado aos acompanhantes medianos para homens é significativamente maior do que para as mulheres. Assim, na Figura 12.9 a inclinação da linha entre os círculos representando as avaliações dos homens para acompanhantes feios e medianos é mais íngreme do que a linha unindo os quadrados representando as avaliações femininas para acompanhantes feios e medianos. Podemos concluir que a preferência para acompanhantes de aparência mediana, comparado com acompanhantes feios, é maior para os homens do que para as mulheres.

12.4.5 A interação entre gênero e carisma ②

A Saída 12.3 do SPSS indicou que o gênero interagiu de alguma maneira com o quanto carismático o acompanhante era. A partir da

tabela resumo podemos relatar que havia uma interação significativa entre a atratividade do acompanhante e o gênero do participante, $F(2, 36) = 62,45, p < 0,001$. Esse efeito nos diz que o perfil das avaliações através dos acompanhantes dos diferentes níveis do carisma foi diferente para homens e mulheres. A média marginal estimada (ou um gráfico do gênero \times carisma usando a caixa de diálogo semelhante à da Figura 12.5) nos revela o significado dessa interação (veja a Saída 12.10 do SPSS e a figura 12.10 que mostra o significado desse resultado). O gráfico mostra a média das avaliações dos acompanhantes feitas pelos homens nos diferentes níveis do carisma ignorando quão atraentes eles eram (círculos). Os escores das mulheres são apresentados como quadrados. O gráfico mostra um padrão quase contrário para os dados de atratividade; novamente, as avaliações dos homens e mulheres são muito similares com uma quantidade normal de carisma, mas dessa vez, os homens mostram mais interesse em acompanhantes entediantes do que as mulheres e as mulheres mostram um interesse um pouco maior em acompanhantes muito carismáticos do que os homens. Em geral, essa interação parece sugerir que o interesse das mulheres em sair com uma pessoa é mais influenciado pelo seu carisma do que os homens. Embora tanto o interesse dos homens e mulheres diminua à medida que o carisma diminui, esse decréscimo é mais pronunciado nas mulheres. Essa interação pode ser esclarecida usando os contrastes especificados antes da análise.

12.4.5.1 Interação 1 carisma \times gênero:
muito \times pouco carisma, homem
 \times mulher ②

O primeiro termo de interação analisa o nível 1 do carisma (muito carisma) em relação ao nível 3 (pouco carisma), comparando os escores dos homens e mulheres. Esse contraste é altamente significativo, $F(1, 18) = 27,20, p < 0,001$. Esse resultado nos diz que o aumento do interesse em acompanhantes com muito carisma comparado com acompanhantes com pouco carisma encontrado em mulheres é significativamente maior do que para os homens. Assim, na Figura 12.10 a inclinação da linha entre os quadrados representando as avaliações femininas de acompanhantes muito carismáticos e com pouco carisma é mais íngreme do que a linha ligando os círculos representando as avaliações dos homens para os acompanhantes com muito carisma e com pouco carisma. Podemos concluir que a preferência para os acompanhantes carismáticos, comparados com os acompanhantes pouco carismáticos, é maior para as mulheres do que para os homens.

12.4.5.2 Interação 2 carisma \times gênero:
chato \times pouco carisma, homem
 \times mulher ②

O segundo termo de interação analisa o nível 2 de carisma (chato) em relação ao nível 3 (pouco carisma), comparando os escores de homens e mulheres. Esse contraste é altamen-

Saída 12.10 do SPSS

5. Gender * CHARISMA (Gênero * Carisma)

Measure: MEASURE_1 (Medida: MEDIDA_1)

GENDER (Gênero)	CHARISMA (Carisma)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
				Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
Male (Masculino)	1	75.967	1.428	72.966	78.967
	2	60.300	0.810	58.598	62.002
	3	69.533	1.035	67.360	71.707
Female (Feminino)	1	88.233	1.428	85.233	91.234
	2	48.300	0.810	46.598	50.002
	3	69.067	1.035	66.893	71.240

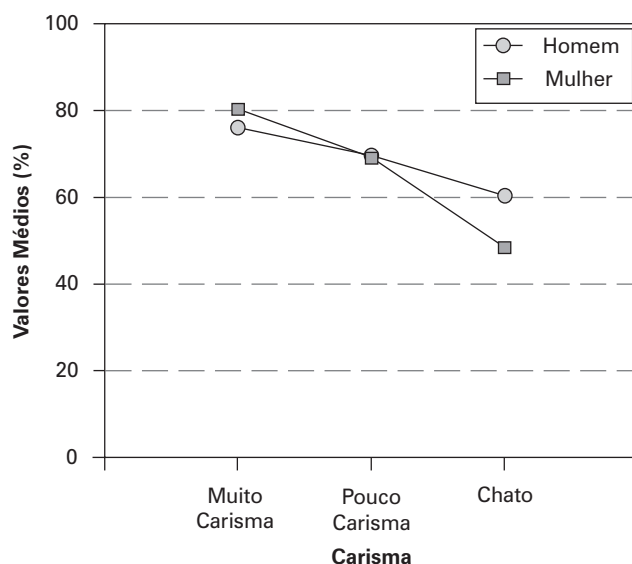


Figura 12.10 Diagrama de Interação entre Gênero e Carisma.

te significativo, $F(1, 18) = 33,69, p < 0,001$. Esse resultado nos diz que o decréscimo do interesse em acompanhantes chatos comparado aos com carisma mediano para as mulheres é significativamente maior do que para homens. Assim, na Figura 12.10, a inclinação da linha entre os quadrados representando as avaliações dos acompanhantes femininos com pouco carisma ou acompanhantes chatos é mais íngreme do que a linha unindo os círculos representando as avaliações dos homens para seus acompanhantes com pouco carisma e acompanhantes chatos. Podemos concluir que a preferência para acompanhantes com pouco carisma sobre os chatos foi maior nas mulheres do que nos homens.

12.4.6 A interação entre atratividade e carisma ②

A Saída 12.3 do SPSS indicou que a atratividade do acompanhante interagia de alguma maneira com quão carismático ele era. Da tabela resumo devemos relatar que havia uma interação significativa entre a atratividade do acompanhante e o seu carisma, $F(4, 72)$

$= 36,63, p < 0,001$. Esse efeito nos informa que o perfil das avaliações ao longo dos acompanhantes nos diversos níveis de carisma foi diferente para acompanhantes atraentes, medianos e feios. As médias marginais estimadas (ou um gráfico da aparência \times carisma usando a caixa de diálogo da Figura 12.5) nos informam o significado dessa interação (veja a Saída 12.11 do SPSS e a Figura 12.11, que mostram o significado desse resultado).

O gráfico mostra a média das avaliações dos acompanhantes de diferentes níveis de atratividade quando o acompanhante também tinha muito carisma (círculos), pouco carisma (quadrados) e era chato (triângulos). Olhe primeiro para as diferenças entre acompanhantes atraentes e medianos. O interesse em acompanhantes muito carismáticos não muda (a linha é mais ou menos reta entre esses dois pontos), mas para acompanhantes com pouco carisma ou chatos os níveis de interesse declinam. Assim, se você tem muito carisma, mas uma aparência mediana, as pessoas ainda vão querer sair com você. Agora, se você olhar para a diferença entre os acompanhantes de aparência mediana e os feios, um padrão diferente é ob-

Saída 12.11 do SPSS

6. LOOKS * CHARISMA (Aparência * Carisma)

Measure: MEASURE_1 (Medida: MEDIDA_1)

LOOKS (Aparência)	CHARISMA (Carisma)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
				Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
1	1	89.950	1.383	86.045	91.855
	2	69.550	1.019	67.409	71.691
	3	87.800	1.408	84.842	90.758
2	1	71.750	1.249	69.126	74.374
	2	45.950	0.746	44.382	47.518
	3	49.750	1.211	47.206	52.294
3	1	85.600	1.721	81.985	89.215
	2	47.400	0.888	45.535	49.265
	3	70.350	1.172	67.888	72.812

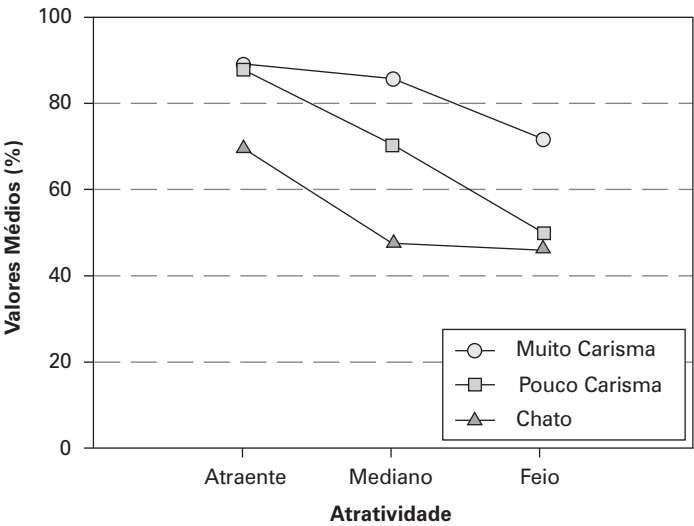
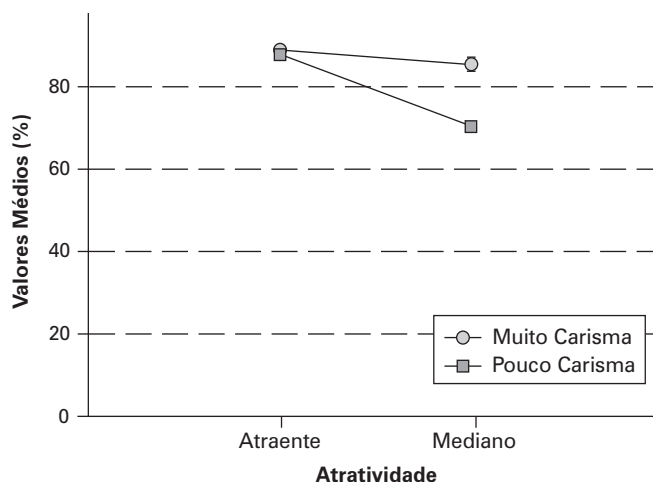


Figura 12.11 Diagrama de Interação entre Atratividade e Carisma.

servado. Para os acompanhantes chatos (triângulos), não existe diferença entre pessoas feias e medianas (assim, se você for chato, precisa ser muito atraente para alguém convidá-lo para sair). Entretanto, para aqueles com carisma, existe um declínio no interesse se for feio (ou seja, se você for feio, ter carisma não irá ajudá-lo muito). Essa interação é muito complexa, mas podemos analisá-la usando os contrastes especificados antes da análise.

12.4.6.1 Interação 1 aparência × carisma:
atraente × mediano, muito
carisma × pouco carisma ②

O primeiro termo de interação investiga o nível 1 da aparência (atraente) em relação ao nível 3 (aparência mediana), comparando o nível 1 do carisma (muito carisma) ao nível 3 do carisma (pouco carisma). É como se perguntássemos: “a diferença entre muito carisma e pouco

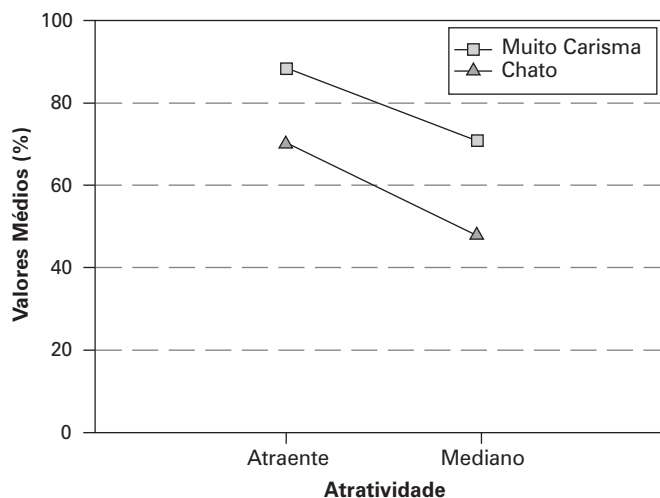


carisma é a mesma para pessoas atraentes e pessoas de aparência mediana?”. A melhor maneira de entender o que esse contraste está testando é extrair a parte relevante do gráfico de interação na Figura 12.11. Se você observar isso, pode ver que o interesse (como indicado pelas altas avaliações) por acompanhantes atraentes foi o mesmo apesar de eles terem muito ou pouco carisma. Entretanto, para acompanhantes de aparência mediana, havia um interesse maior quando a pessoa tinha muito carisma em vez de pouco. O contraste é altamente significativo, $F(1, 18) = 21,94$, $p < 0,001$ e nos informa que

à medida que os acompanhantes se tornam menos atraentes, há um declínio maior no interesse quando o carisma é pouco comparado com quando o carisma é muito.

12.4.6.2 Interação 2 aparência \times carisma: atraente \times mediano, chato \times pouco carisma ②

O segundo termo de interação investiga o nível 1 da aparência (atraente) em relação ao nível 3 (aparência mediana), quando comparado ao nível 2 do carisma (chato) ao



nível 3 do carisma (pouco carisma). É como se perguntássemos: “a diferença entre chato e pouco carisma é a mesma para pessoas atraentes e pessoas de aparência mediana?”. Novamente, a melhor maneira de entender o que o contraste está testando é extrair a parte relevante do gráfico de interação da Figura 12.11. Se você observar isso, pode ver que o interesse (como indicado pelas altas avaliações) diminui dos acompanhantes de aparência mediana para os feios tanto para os acompanhantes com muito carisma quanto nos de pouco carisma; entretanto, essa queda é levemente maior em acompanhantes com pouco carisma (a linha conectando os quadrados é levemente mais inclinada). O contraste é significativo, $F(1,18) = 6,23$, $p < 0,05$ e nos informa que à medida que os acompanhantes se tornam menos atraentes, existe um grande declínio no interesse quando o carisma é pouco ou quando não existe carisma.

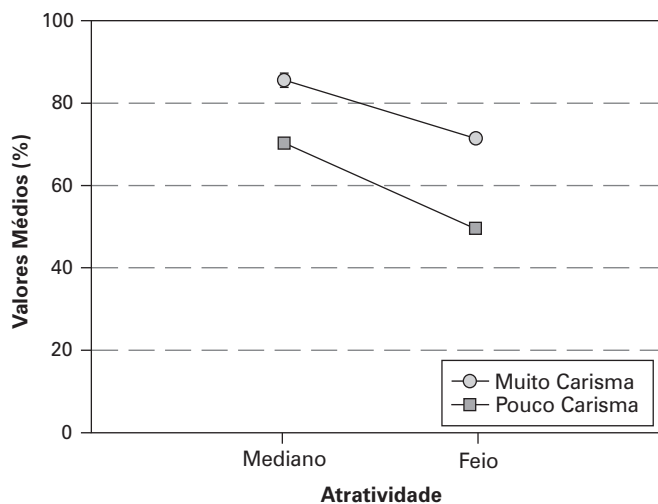
12.4.6.3 Interação 3 aparência × carisma: feio × mediano, muito carisma × pouco carisma ②

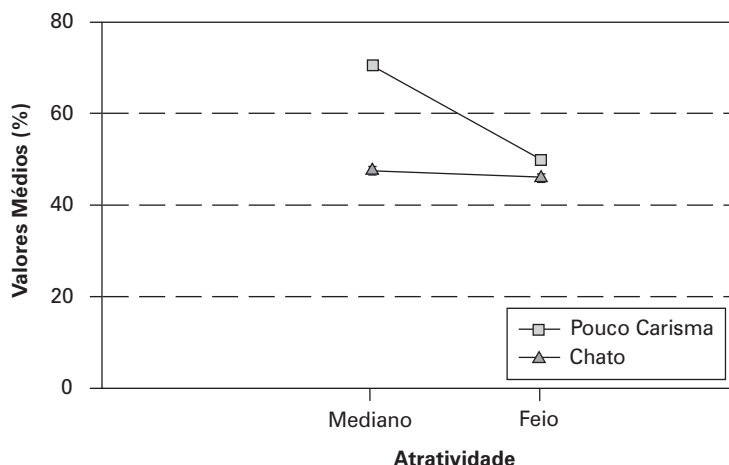
O terceiro termo de interação investiga o nível 2 da aparência (feio) em relação ao nível 3 (aparência mediana), comparando o nível 1 do carisma (muito carisma) ao nível 3 do carisma (pouco carisma). É como se perguntássemos: “a diferença entre muito carisma e pouco carisma é a mesma para pessoas feias e pessoas com aparência mediana?”. Se extrairmos a parte relevante do gráfico de interação na Figura 12.11, podemos ver que o interesse (como

de aparência mediana?”. Se, novamente, extrairmos a parte relevante do gráfico de interação na Figura 12.11, podemos ver que o interesse (como indicado pelas altas avaliações) diminui dos acompanhantes de aparência mediana para os feios tanto para os acompanhantes com muito carisma quanto nos de pouco carisma; entretanto, essa queda é levemente maior em acompanhantes com pouco carisma (a linha conectando os quadrados é levemente mais inclinada). O contraste é significativo, $F(1,18) = 6,23$, $p < 0,05$ e nos informa que à medida que os acompanhantes se tornam menos atraentes, existe um grande declínio no interesse quando o carisma é pouco comparado com quando o carisma é muito.

12.4.6.4 Interação 4 aparência × carisma: feio × mediano, chato × pouco carisma ②

O último termo de interação analisa o nível 2 da aparência (feio) em relação ao nível 3 (aparência mediana), quando comparando ao nível 2 do carisma (chato) ao nível 3 do carisma (pouco carisma). É o mesmo que perguntar: “a diferença entre chato e pouco carisma é a mesma para pessoas feias e pessoas com aparência mediana?”. Se extrairmos a parte relevante do gráfico de interação na Figura 12.11, podemos ver que o interesse (como





indicado pelas altas avaliações) em acompanhantes de aparência mediana foi maior quando eles tinham pouco carisma do que quando eles eram chatos, mas para acompanhantes feios as avaliações foram aproximadamente as mesmas do nível do carisma. Esse contraste é altamente significativo, $F(1, 18) = 88.60$, $p < 0,001$ e nos informa que à medida que os acompanhantes se tornam menos atraentes, o declínio do interesse em acompanhantes com um pouco de carisma é significativamente maior do que para os chatos.

12.4.7 A interação entre aparência, carisma e gênero ③



A interação de três fatores nos informa se a interação aparência \times carisma, descrita acima, é a mesma tanto para os homens quanto para as mulheres (isto é, se o efeito combinado da atratividade do acompanhante e do seu nível de carisma é o mesmo tanto para os participantes homens quanto para mulheres). A Saída 12.3 do SPSS nos informa que existe uma interação de três fatores significativa da aparência \times carisma \times gênero, $F(4, 72) = 24,12$, $p < 0,001$. A natureza dessa interação

está revelada na Figura 12.12, que mostra a interação da aparência \times carisma para homens e mulheres separadamente (as médias em que o gráfico é baseado aparecem na Saída 12.12 do SPSS). O gráfico dos homens mostra que quando os acompanhantes são atraentes, os homens expressarão alto interesse apesar dos níveis de carisma (círculo, quadrado e triângulo se sobrepõem). No lado oposto da escala de atratividade, quando o acompanhante é feio, apesar do carisma, os homens expressarão pouco interesse (as avaliações são bem baixas). O único momento em que o carisma faz diferença para um homem é se a acompanhante tem uma aparência mediana; nesse caso, ter muito carisma aumenta o interesse, ser chato reduz o interesse e ter pouco carisma não prejudica nem melhora. A conclusão é que os homens são uns cretinos superficiais mais interessados em atributos físicos. O quadro para as mulheres é bem diferente. Se o acompanhante tem muito de carisma, não importa a sua aparência, as mulheres irão expressar interesse neles (a linha dos círculos é relativamente reta). No outro extremo, se o acompanhante for chato, elas não irão demonstrar interesse apesar da sua atratividade (a linha dos triângulos é relativamente reta). A única ocasião em que a atratividade faz uma diferença é quando o acompanhante tem pouco carisma; nesse caso, ser atraente aumenta o interesse e ser feio o reduz. Ou seja, as mulheres

Saída 12.12 do SPSS

7. Gender * LOOKS * CHARISMA (Gênero* Aparência* Carisma)

Measure: MEASURE_1 (Medida: MEDIDA_1)

Gender (Gênero)	LOOKS (Aparência)	CHARISMA (Carisma)	Mean (Média)	Std. Error (Erro Padrão)	95% Confidence Interval (Intervalo de 95% de confiança)	
					Lower Bound (Limite Inferior)	Upper Bound (Limite Superior)
Male (Masculino)	1	1	88.300	1.956	84.191	92.409
		2	87.300	1.441	84.273	90.327
		3	88.500	1.991	84.317	92.683
	2	1	56.800	1.767	53.089	60.511
		2	45.800	1.055	43.583	48.017
		3	48.300	1.712	44.703	51.897
	3	1	82.800	2.434	77.687	87.913
		2	47.800	1.255	45.163	50.437
		3	71.800	1.657	68.318	75.282
Male (Feminino)	1	1	89.600	1.956	85.491	93.709
		2	51.800	1.441	48.773	54.827
		3	87.100	1.991	82.917	91.283
	2	1	86.700	1.767	82.989	90.411
		2	46.100	1.055	43.883	48.317
		3	51.200	1.712	47.603	54.797
	3	1	88.400	2.434	83.287	93.513
		2	47.000	1.255	44.363	49.637
		3	68.900	1.657	65.418	72.382

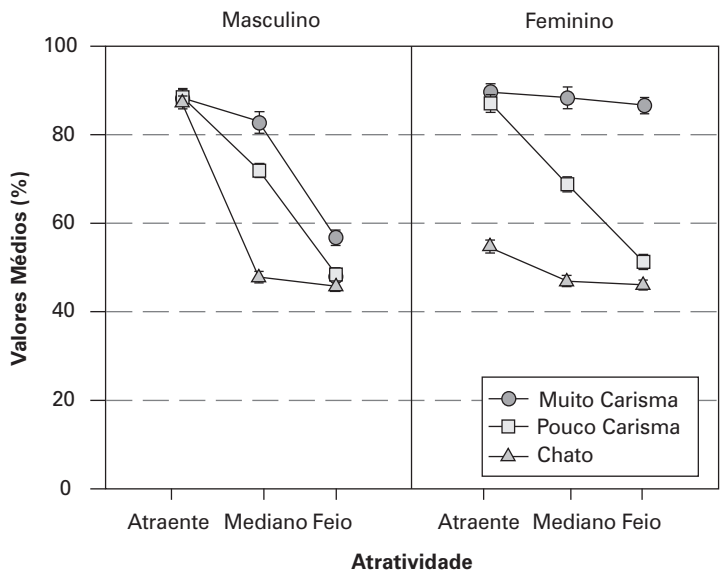
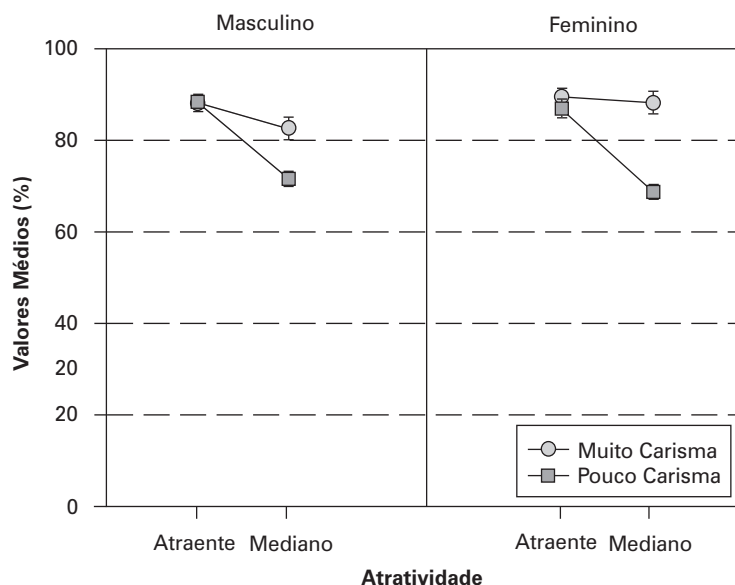


Figura 12.12 Gráficos mostrando a interação aparência × carisma para homens e mulheres. As linhas representam alto carisma (círculos), pouco carisma (quadrados) e chato (triângulos).



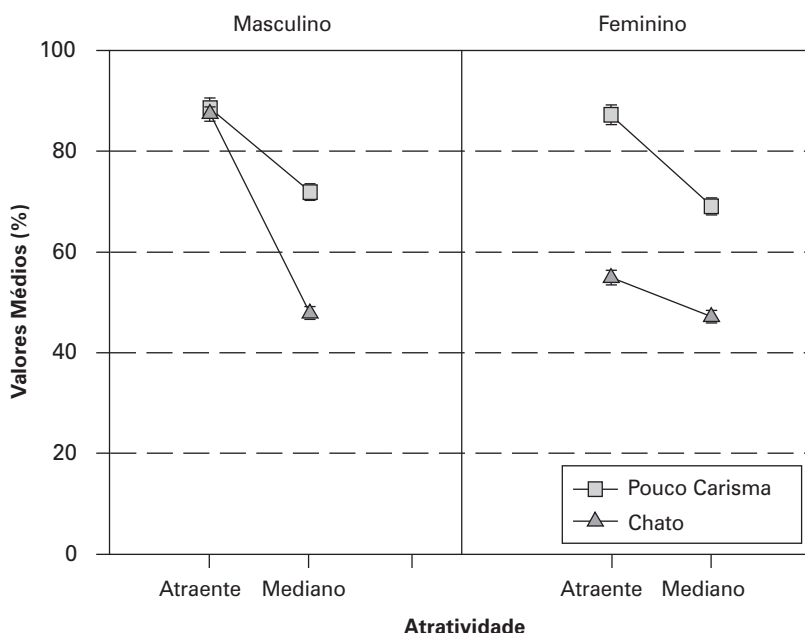
priorizam o carisma à aparência física. Novamente, podemos analisar alguns contrastes para entender essa interação. Esses contrastes são semelhantes àqueles da interação aparência \times carisma, mas agora eles também levam em conta o gênero!

12.4.7.1 Interação 1 aparência \times carisma \times gênero: atraente \times mediano, muito carisma \times pouco carisma, homem \times mulher ③

O primeiro termo de interação relaciona o nível 1 da aparência (atraente) ao nível 3 (aparência mediana), quando o nível 1 do carisma (muito carisma) é comparado ao nível 3 do carisma (pouco carisma) em homens comparados às mulheres, $F(1, 18) < 1, ns$. Se extrairmos as partes relevantes do gráfico de interação na Figura 12.12, podemos ver que o interesse (como indicado pelas altas avaliações) em acompanhantes atraentes foi o mesmo apesar de muito ou pouco carisma. Entretanto, para acompanhantes de aparência mediana, havia maior interesse quando a pessoa tinha mais carisma do que pouco. Mais importante, esse padrão de resultados é o mesmo em homens e mulheres e isso está refletido na não-significância desse contraste.

12.4.7.2 Interação 2 aparência \times carisma \times gênero: atraente \times mediano, chato \times pouco carisma, homem \times mulher ③

O segundo termo de interação compara o nível 1 da aparência (atraente) ao nível 3 (aparência mediana) quando o nível 2 do carisma (chato) é relacionado ao nível 3 do carisma (pouco carisma), em homens comparados às mulheres. Novamente, extraímos a parte relevante do gráfico de interação da Figura 12.12 e você pode ver que os padrões são diferentes para homens e mulheres. Isso é refletido pelo fato de que o contraste é significativo, $F(1, 18) = 60,67, p < 0,001$. Para entender isso, precisamos olhar o gráfico. Primeiro, se analisarmos os acompanhantes de aparência mediana, tanto para os homens quanto para as mulheres o maior interesse é expresso quando os acompanhantes têm pouco carisma do que quando são chatos (e a distância entre o quadrado e o triângulo é quase a mesma). Assim, a diferença não parece estar aqui. Se agora olharmos os acompanhantes atraentes, vemos que os homens estão igualmente interessados em seus acompanhantes independentemente do seu carisma, mas para as mulheres, elas

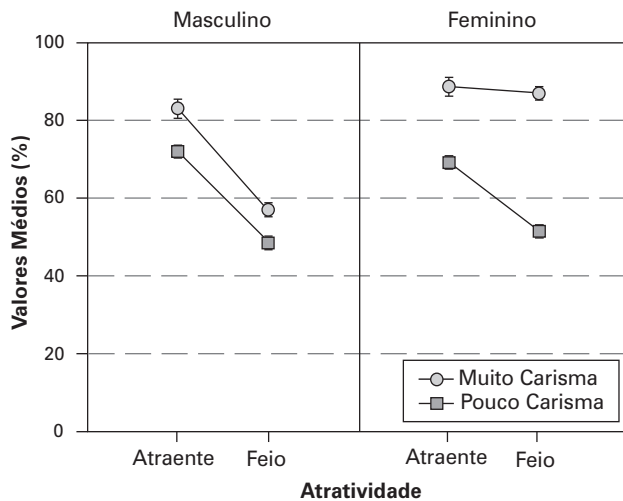


são menos atraídas por uma pessoa atraente se ela for chata. Ou seja, para acompanhantes atraentes, a distância entre o quadrado e o triângulo é muito menor para homens do que para mulheres. Outra maneira de analisar isso é que para acompanhantes com pouco carisma, a redução no interesse quando a atratividade diminui é quase a mesma em homens e mulheres (as linhas com os quadrados têm a mesma inclinação). Entretanto, para acompanhantes chatos, o decréscimo no interesse se esses acompanhantes são de aparência mediana em vez de atraentes, é muito maior em homens do que em mulheres (a linha com triângulos é mais inclinada para homens do que para as mulheres). Isso é o que o contraste significativo nos informa.

12.4.7.3 Interação 3 aparência × carisma × gênero: feio × mediano, muito carisma × pouco carisma, homem × mulher ③

O terceiro termo de interação compara o nível 2 da aparência (feio) com o nível 3 (aparência mediana), quando o nível 1 do ca-

risma (muito carisma) é relacionado ao nível 3 do carisma (pouco carisma), em homens em comparação às mulheres. Novamente, extraímos a parte relevante do gráfico de interação da Figura 12.12 e você pode ver que os padrões são diferentes para homens e mulheres. Isso é refletido pelo fato de que o contraste é significativo, $F(1, 18) = 11,70, p < 0,01$. Para entender isso, precisamos olhar o gráfico. Primeiro, vamos analisar os homens. Para os homens, à medida que a atratividade diminui, o interesse também diminui quando a acompanhante tem muito ou pouco carisma. De fato, as linhas são paralelas. Assim, apesar do carisma, existe uma redução similar no interesse à medida que a atratividade diminui. Para as mulheres, o quadro é bem diferente: quando tem muito carisma, não há declínio no interesse à medida que a atratividade diminui (a linha conectando os círculos é reta), mas quando tem pouco carisma, a atratividade do acompanhante não interessa e o interesse é menor num acompanhante feio do que num de aparência mediana. Outra maneira de analisar isso é que para acompanhantes com pouco carisma, a



redução no interesse, à medida que a atratividade diminui, é quase a mesma em homens e mulheres (as linhas com quadrados têm a mesma inclinação). Entretanto, para acompanhantes que tem muito carisma, o decréscimo no interesse se os acompanhantes são feios em vez de aparência mediana, é muito maior em homens do que em mulheres (a linha com círculos é muito mais inclinada para homens do que para mulheres). Isso é o que o contraste significativo nos informa.

12.4.7.4 Interação 4 aparência × carisma × gênero: atraente × mediano, chato × pouco carisma, homem × mulher ③

O último termo de interação compara o nível 2 da aparência (feio) com o nível 3 (aparência mediana), quando relacionado o nível 2 do carisma (chato) ao nível 3 do carisma (pouco carisma), em homens comparados às mulheres. Se extrairmos as partes relevantes do

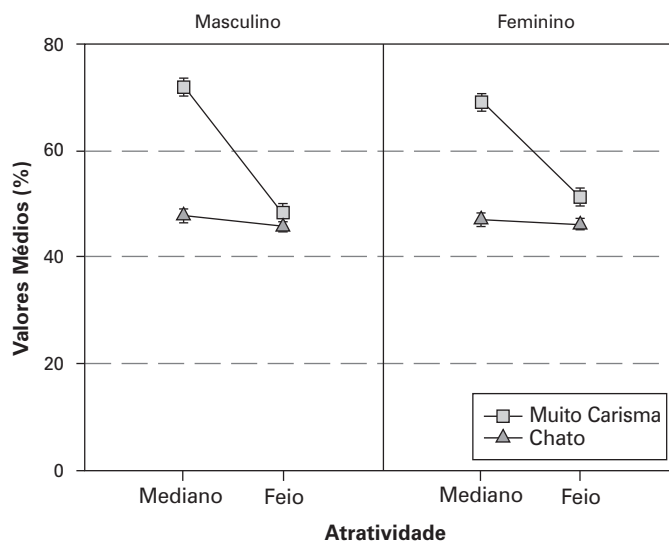


gráfico de interação da Figura 12.12, podemos ver que o interesse (como indicado pelas altas avaliações) em acompanhantes feios foi o mesmo independentemente de eles terem muito ou pouco carisma. Entretanto, para acompanhantes de aparência mediana, havia mais interesse quando a pessoa tinha pouco carisma do que se ela fosse chata. Mais importante, esse padrão de resultados é o mesmo para homens e mulheres e isso está refletido na não-significância desse contraste, $F(1, 18) = 1,33, ns$.

12.4.8 Conclusões ③

Esses contrastes nada nos informam sobre as diferenças entre as condições atraente e feio ou as condições de muito ou pouco carisma

porque elas nunca foram comparadas. Podemos executar a análise novamente e especificar nossos contrastes de forma diferente para conseguir esses efeitos. Entretanto, o que está claro dos nossos dados é que existem diferenças entre homens e mulheres em termos de como eles são afetados pela aparência e personalidade dos acompanhantes. Os homens parecem querer sair com qualquer mulher que seja atraente independentemente de quão horrível é sua personalidade. As mulheres são quase o oposto: querem sair com qualquer homem que tenha muito carisma independentemente da sua aparência (e rejeitam sair com pessoas chatas não importa quão atraentes sejam). A única semelhança entre homens e mulheres é quando existe algum carisma (mas não muito): nesse caso,

Dica da Samanta Ferrinho



- A ANOVA mista compara várias médias quando existem duas ou mais variáveis independentes e, pelo menos, uma das variáveis foi mensurada usando os mesmos participantes e, no mínimo, outra variável foi mensurada usando participantes diferentes.
- Teste as hipóteses da *esfericidade* para a(s) variável(is) de medidas repetidas quando elas tem três ou mais condições usando o teste de Mauchly. Se o valor na coluna Sig. é menor do que 0,05, a hipótese foi violada. Se a significância do teste de Mauchly é maior do que 0,05, a hipótese de esfericidade está mantida. Você deve testar essa hipótese para todos os efeitos (se há duas ou mais variáveis de medidas repetidas, isso significa que você deve testar a hipótese para todas as variáveis e para os termos de interação correspondentes).
- A tabela *Efeitos entre participantes* mostra os resultados para a ANOVA de medidas repetidas e todos os efeitos de interação. Para cada efeito, se a hipótese de esfericidade foi satisfeita, olhe para linha Esfericidade Assumida (*Sphericity Assumed*). Se a hipótese foi violada, leia a linha *Greenhouse e Geiser* (você também pode olhar *Huynh-Feldt*, mas deve ler o capítulo anterior para descobrir os prós e contras dos dois procedimentos). Tendo selecionado a linha apropriada, olhe para a coluna Sig.: se o valor for menor do que 0,05, as médias dos grupos são significativamente diferentes.
- A tabela “Testes dos Efeitos entre Participantes” (*Tests of Between-Subjects Effects*) mostra os resultados da ANOVA para as variáveis entre grupos. Observe a tabela Sig.: se o valor for menor do que 0,05, as médias dos grupos são significativamente diferentes.
- Analise os efeitos principais e os termos de interação usando contrastes. Esses contrastes aparecem na tabela “Testes dos Contrastes entre Participantes” (*Tests of Within-Subjects Contrasts*); novamente, observe a coluna Sig. para descobrir se as suas comparações são significativas (elas serão se o valor de significância for menor do que 0,05).
- Observe as médias, ou melhor, faça gráficos para ajudá-lo a interpretar esses contrastes.

para ambos os gêneros a atratividade influencia o desejo de sair com essa pessoa.

Neste capítulo deve ficar claro que quando mais de duas variáveis independentes são usadas numa ANOVA, ela gera efeitos de interação complexos que exigem muita concentração para interpretar (imaginem interpretar uma interação de quatro fatores!). Portanto, é essencial ter uma abordagem sistemática para a interpretação, e traçar gráficos é um método inteligente. Também é aconselhável pensar cuidadosamente sobre os contrastes apropriados para usar na resposta das questões que você tem sobre seus dados. São esses contrastes que irão ajudá-lo a interpretar as interações, portanto, verifique se você selecionou os mais adequados!

12.5 CALCULANDO O TAMANHO DE EFEITO ③



Os tamanhos dos efeitos são mais úteis quando resumem um efeito focado. Isso me dá boa desculpa para burlar as complexidades do ômega quadrado em delineamentos mistos (é o caminho para a loucura, eu asseguro). Portanto, apenas calcule o tamanho dos efeitos para os seus contrastes quando você tem um delineamento fatorial (e qualquer outro efeito que compare somente dois grupos). A Saída 12.7 do SPSS mostra os valores para muitos contrastes, todos tendo 1 grau de liberdade para o modelo (isto é, eles representam uma comparação interpretável e focada) e 18 graus de liberdade residuais. Podemos usar as razões F e convertê-las em um tamanho de efeito r usando uma fórmula que já vimos anteriormente:

$$r = \sqrt{\frac{F(1, g l_R)}{F(1, g l_R) + g l_R}}$$

Primeiro, podemos lidar com o efeito principal do gênero porque ele compara somente dois grupos:

$$r_{\text{Gênero}} = \sqrt{\frac{0,005}{0,005 + 18}} = 0,02$$

Para as duas comparações que fizemos para a variável aparência (Saída 12.7 do SPSS), obtemos:

$$r_{\text{Atraente vs. Mediano}} = \sqrt{\frac{226,99}{226,99 + 18}} = 0,96$$

$$r_{\text{Feio vs. Mediano}} = \sqrt{\frac{160,07}{160,07 + 18}} = 0,95$$

Portanto, ambas as comparações geram tamanhos massivos do efeito. Para as duas comparações que fizemos para a variável do carisma (Saída 12.7 do SPSS), temos:

$$r_{\text{Muito vs. Pouco}} = \sqrt{\frac{109,94}{109,94 + 18}} = 0,93$$

$$r_{\text{Chato vs. Pouco}} = \sqrt{\frac{227,94}{227,94 + 18}} = 0,96$$

Novamente, ambas as comparações geram massivos tamanhos do efeito. Para a interação aparência \times gênero, temos, novamente, dois contrastes:

$$r_{\text{Atraente vs. Mediano, Homens vs. Mulheres}} = \sqrt{\frac{43,26}{43,26 + 18}} = 0,84$$

$$r_{\text{Feio vs. Mediano, Homens vs. Mulheres}} = \sqrt{\frac{30,23}{30,23 + 18}} = 0,79$$

Novamente, esses são efeitos massivos. Para a interação carisma \times gênero, os dois contrastes nos dão:

$$r_{\text{Muito vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{27,20}{27,20 + 18}} = 0,78$$

$$r_{\text{Chato vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{33,69}{33,69 + 18}} = 0,81$$

Mais uma vez efeitos massivos (nem dá para notar que os dados são inventados!). Na interação aparência \times carisma, temos quatro contrastes:

$$r_{\text{Atraente vs. Mediano, Muito vs. Pouco}} = \sqrt{\frac{21,94}{21,94 + 18}} = 0,74$$

$$r_{\text{Atraente vs. Mediano, Chato vs. Pouco}} = \sqrt{\frac{4,09}{4,09 + 18}} = 0,43$$

$$r_{\text{Feio vs. Mediano, Muito vs. Pouco}} = \sqrt{\frac{6,23}{6,23 + 18}} = 0,51$$

$$r_{\text{Alto vs. Mediano, Chato vs. Pouco}} = \sqrt{\frac{88,60}{88,60 + 18}} = 0,91$$

Todos esses efeitos estão num intervalo de médio para massivo. Finalmente, para a interação aparência \times carisma \times gênero, temos quatro contrastes:

$$r_{\text{Atraente vs. Mediano, Muito vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{0,93}{0,93 + 18}} = 0,22$$

$$r_{\text{Atraente vs. Mediano, Chato vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{60,67}{60,67 + 18}} = 0,88$$

$$r_{\text{Feio vs. Mediano, Muito vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{11,70}{11,70 + 18}} = 0,63$$

$$r_{\text{Feio vs. Mediano, Chato vs. Pouco, Homens vs. Mulheres}} = \sqrt{\frac{1,33}{1,33 + 18}} = 0,26$$



Desse modo, os dois efeitos que são significativos (atraente vs. mediano, chato vs. algum, homens vs. mulheres, e feio vs. mediano, alto vs. algum, homens vs. mulheres) geram tamanhos dos efeitos grandes. Os dois efeitos que não são significativos geram tamanhos de efeito próximos a médio.

12.6 RELATANDO OS RESULTADOS DA ANOVA MISTA ②

Quando você tem mais do que duas variáveis independentes, há muita informação para relatar. Você deve relatar todos os efeitos principais, todas as interações e qualquer contraste que tenha feito. Isso pode tomar muito espaço, e uma boa dica é reservar os detalhes para os efeitos que realmente importam (isto é, os efeitos principais em geral não são muito interessantes se você tem uma interação significativa que inclui aquela variável). Sou um grande fã de dar breves explicações dos resultados para

que você entenda o que o efeito está dizendo, assim, tenho a tendência de relatar não somente os resultados, mas também alguma interpretação. Alguns editores adoram me dizer que as minhas seções de resultados são muito grandes. Portanto, provavelmente você deve ignorar tudo o que eu disse. Presumindo que você queira relatar todos os seus efeitos, pode fazer algo assim (embora não como uma lista!):

- ✓ Todos os efeitos são relatados como significativos a $p < 0,05$. Havia um efeito significativo da atratividade do acompanhante no interesse expressado pelo participante, $F(2, 36) = 423,73$. Os contrastes revelaram que acompanhantes atraentes eram mais desejados do que os de aparência mediana, $F(1, 18) = 226,99$, $r = 0,96$ e os acompanhantes feios eram menos desejados do que os de aparência mediana, $F(1, 18) = 160,07$, $r = 0,95$.
- ✓ Houve, também, um efeito significativo do carisma do acompanhante no interesse em sair com eles, $F(2, 36) = 328,25$. Os contrastes revelaram que acompanhantes com muito carisma eram mais desejados do que os acompanhantes com pouco carisma $F(1, 18) = 109,94$, $r = 0,93$, e chatos eram menos desejados do que acompanhantes com pouco carisma, $F(1, 18) = 227,94$, $r = 0,96$.
- ✓ Não houve um efeito significativo do gênero indicando que as avaliações dos participantes homens e mulheres eram geralmente as mesmas, $F(1, 18) < 1$, $r = 0,02$.
- ✓ Houve um efeito de interação significativo entre a atratividade do acompanhante e o gênero do participante, $F(2, 36) = 80,43$. Isso indica que a vontade de sair com acompanhantes dos diferentes níveis de atratividade diferia em homens e mulheres. Para analisar essa interação, contrastes foram feitos comparando cada nível de atratividade à aparência mediana em participantes homens e mulheres. Isso revelou interações significativas quando comparando os escores dos homens e mulheres para acompanhantes atraentes em relação aos de aparência mediana, F

(1, 18) = 43,26, $r = 0,84$ e para acompanhantes feios comparados aos de aparência mediana, $F(1, 18) = 30,23$, $r = 0,79$. Observando o gráfico de interação, as avaliações dos homens e mulheres são muito similares para acompanhantes de aparência mediana, mas os homens dão mais valor a acompanhantes atraentes do que as mulheres, enquanto as mulheres dão mais valor aos acompanhantes feios do que os homens. Embora tanto o interesse dos homens quanto o das mulheres diminua à medida que a atratividade diminui, esse decréscimo foi maior nos homens, sugerindo que quando o carisma é ignorado, o interesse dos homens em sair com uma pessoa é mais influenciado pela sua aparência do que para as mulheres.

- ✓ Houve um efeito de interação significativo entre o nível de carisma do acompanhante e o gênero do participante, $F(2, 36) = 62,45$. Isso indica que a vontade de sair com acompanhantes dos diferentes níveis do carisma foi diferente em homens e mulheres. Para analisar essa interação, contrastes foram feitos comparando cada nível do carisma à categoria “pouco carisma” entre participantes homens e mulheres. Isso revelou interações significativas quando comparando os escores dos homens e mulheres para acompanhantes muito carismáticos e acompanhantes com pouco carisma, $F(1, 18) = 27,20$, $r = 0,78$, e para chatos comparados a acompanhantes com pouco carisma, $F(1, 18) = 33,69$, $r = 0,81$. O gráfico de interação mostra que os homens demonstram mais interesse em acompanhantes que são chatos do que as mulheres e as mulheres demonstram um interesse levemente maior em acompanhantes muito carismáticos do que os homens. O interesse tanto dos homens quanto das mulheres diminui à medida que o carisma diminui, mas esse decréscimo é maior em mulheres, sugerindo que o interesse das mulheres em sair com uma pessoa é mais influenciado pelo seu carisma do que para os homens.

- ✓ Houve um efeito de interação significativo entre o nível do carisma do acompanhante e a atratividade do acompanhante $F(4, 72) = 36,63$. Isso indica que a vontade de sair com os acompanhantes dos diferentes níveis de carisma diferiu de acordo com sua atratividade. Para analisar essa interação, foram feitos contrastes comparando cada nível do carisma com a categoria do meio (pouco carisma) ao longo de cada nível da atratividade comparado à categoria de atratividade mediana. O primeiro contraste revelou uma interação significativa na comparação entre acompanhantes atraentes e os de aparência mediana quando o acompanhante tinha muito carisma em relação a pouco carisma, $F(1, 18) = 21,94$, $r = 0,74$, e nos informou que à medida que os acompanhantes se tornam menos atraentes, existe um grande declínio no interesse quando o carisma é pouco comparado com quando o carisma é muito. O segundo contraste comparou acompanhantes atraentes aos acompanhantes de aparência mediana quando o acompanhante era chato em relação a quando ele tinha pouco carisma. Isso não foi significativo, $F(1, 18) = 4,09$, $r = 0,43$, e indica que à medida que os acompanhantes se tornam menos atraentes, existe um declínio no interesse quando o carisma é baixo ou nulo. O terceiro contraste comparou acompanhantes feios aos acompanhantes de aparência mediana quando os acompanhantes tinham muito carisma em relação a pouco carisma. Isso foi significativo, $F(1, 18) = 6,23$, $r = 0,51$, e indica que à medida que os acompanhantes se tornam menos atraentes, existe um grande declínio no interesse quando o carisma é pouco comparado com quando o carisma é muito. O contraste final comparou acompanhantes feios aos de aparência mediana quando o acompanhante era chato em relação a quando ele tinha pouco carisma. Esse contraste foi altamente significativo, $F(1, 18) = 88,60$, $r = 0,91$, e informa que à medida que os acompanhantes se tornam menos atraentes, o declínio no interesse em acompanhantes com pouco

carisma é significativamente maior do que para os chatos.

- ✓ Finalmente, a interação aparência \times carisma \times gênero foi significativa, $F(4, 72) = 24,12$. Isso indica que a interação aparência \times carisma descrita previamente foi diferente em participantes homens e mulheres. Novamente, contrastes foram usados para analisar essa interação; esses contrastes compararam os escores dos homens e mulheres em cada nível do carisma em relação à categoria do meio de “pouco carisma” através de cada nível de atratividade comparado com a categoria da atratividade mediana. O primeiro contraste revelou uma diferença não significativa entre as respostas dos homens e mulheres quando relacionando os acompanhantes atraentes aos de aparência mediana quando o acompanhante tinha alto carisma comparado a pouco carisma, $F(1, 18) < 1$, $r = 0,22$, e indica que para os homens e as mulheres, à medida que os acompanhantes se tornam menos atraentes, existe um grande declínio no interesse quando o carisma é baixo comparado ao muito carisma. O segundo contraste investigou diferenças entre homens e mulheres relacionando acompanhantes atraentes aos de aparência mediana quando os acompanhantes eram chatos comparado aos com pouco carisma. Isso foi significativo, $F(1, 18) = 60,67$, $r = 0,88$, e informa que para acompanhantes com pouco carisma, a redução no interesse à medida que a atratividade diminui é quase a mesma para homens e mulheres, mas para acompanhantes chatos, o decréscimo no interesse se esses acompanhantes são de aparência mediana em vez de atraentes é muito maior em homens do que em mulheres. O terceiro contraste buscou diferenças entre homens e mulheres quando comparando acompanhantes feios aos de aparência mediana quando o acompanhante tinha muito carisma em relação a pouco carisma. Esse foi significativo, $F(1, 18) = 11,70$, $r = 0,63$, e indica que para os acompanhantes com pouco carisma, a redução no interesse à

medida que a atratividade diminui é quase a mesma em homens e mulheres, mas para acompanhantes que tem pouco carisma, o decréscimo no interesse se os acompanhantes são feios em vez de com aparência mediana é muito maior em homens do que em mulheres. O último contraste procurou diferenças entre homens e mulheres quando comparando acompanhantes feios aos de aparência mediana quando o acompanhante era chato comparado a um que tinha pouco carisma. Esse contraste não foi significativo, $F(1, 18) = 1,33$, $r = 0,26$, e indica que, tanto para homens quanto para mulheres, à medida que os acompanhantes se tornam menos atraentes, o declínio no interesse em acompanhantes com um pouco de carisma é significativamente maior do que para os chatos.

12.7 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

A ANOVA de três fatores é muito difícil de decifrar. Já fiz centenas de ANOVAs de três fatores na minha vida e mesmo assim foi complicado escrever este capítulo (portanto, se você estiver confuso após ler este capítulo, a culpa não é sua, é minha). Com sorte, você deve ter descoberto que esse tipo de ANOVA é flexível o suficiente para que você possa misturar e relacionar variáveis independentes que são mensuradas usando os mesmos ou diferentes participantes. Além disso, observamos como a ANOVA é também flexível o bastante para ir além de simplesmente incluir duas variáveis independentes. Espero que você também tenha entendido por que existem boas razões para limitar o número de variáveis independentes que você inclui (para o bem da interpretação). E, é claro, você aprendeu que os homens são criaturas superficiais que preferem aparência ao carisma e que as mulheres saíam até com o corcunda de Notre Dame se ele tivesse carisma o suficiente. Agora vamos diminuir um pouco o ritmo e observar testes que podemos usar quando as hipóteses necessárias para os testes paramétricos são violadas e não podem ser corrigidas.

12.8 TERMOS-CHAVE QUE DESCOBRIMOS

- ANOVA mista
- Delineamento misto

12.9 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Estenderei o exemplo do capítulo anterior (propaganda e imagens diferentes) acrescentando uma variável entre grupos ao delineamento.⁴ Re-

capitulando, no caso de você não ter lido o capítulo anterior, os participantes assistiram a um total de nove comerciais falsos durante três sessões. Nesses comerciais, havia três produtos (uma marca de cerveja, Brain Death, uma marca de vinho, Dangleberry, e uma marca de água, Puritan). Elas poderiam ser apresentadas com imagens positivas, negativas e neutras. Durante as três sessões e nove comerciais, cada tipo de produto foi apresentado com um tipo de imagem (leia o último capítulo se você precisar de mais detalhes). Após cada comercial, os participantes avaliaram as bebidas numa escala variando de -100 (detesta) passando por 0 (neutro) a 100 (gosta muito). O delineamento, até agora, tem duas variáveis independentes: o tipo de bebida (cerveja, vinho e água) e o tipo de imagem utilizada (positiva, negativa ou neutra). Essas duas variáveis se sobrepõem completamente produzindo nove condições experimentais. Agora imagine que eu também anotei o gênero de cada pessoa. Depois da análise do capítulo anterior, me ocorreu que homens e mulheres podem ter uma resposta diferente aos produtos (porque, conforme o estereótipo, os homens, na sua maioria, gostam mais de cerveja e as mulheres, de vinho). Portanto, gostaria de analisar novamente os dados levando

em consideração essa variável adicional. O gênero é uma variável entre grupos porque um participante pode ser somente homem ou mulher: eles não podem participar como homem e depois trocar para mulher e participar novamente! Os dados são os mesmos do capítulo anterior (Tabela 11.3) e estão no arquivo **MixedAttitude.sav**. Execute uma ANOVA mista com esses dados. ③

- **Tarefa 2:** Mensagens de texto são muito populares entre donos de telefones celulares, a ponto de livros sobre como escrever em linguagem de texto terem sido publicados (espero q vc saiba o q eu kero dizer com linguagem de txtu). Há uma preocupação de que as crianças usem muito essa forma de comunicação e isso atrapalhe a sua habilidade de aprender o idioma corretamente. Um pesquisador conduziu um experimento no qual um grupo de crianças foi incentivado a mandar mensagens de texto dos seus celulares por um período de seis meses. O segundo grupo foi proibido de mandar mensagens de texto pelo mesmo período. Para assegurar que as crianças desse último grupo não usassem seus celulares, foram colocadas braçadeiras que administravam choques doloridos na presença de micro-ondas (como as emitidas dos celulares)⁵. Havia 50 participantes diferentes: 25 foram incentivados a enviar mensagens de texto e 25 foram proibidos. A saída foi um escore num teste de gramática (como uma porcentagem), mensurada antes e depois do experimento. A primeira variável independente foi, portanto, o uso da mensagem de texto (*text messagers versus controls*)(mensagens de texto *versus* controle)) e a segunda variável independente foi o período em que a habilidade gramatical foi avaliada (antes e depois do experimento). Os dados estão no arquivo **TextMessages.sav**. ③

⁴ Anteriormente, o exemplo continha duas variáveis de medidas repetidas (tipo de bebida e tipo de imagem), mas agora ele irá incluir três variáveis (duas medidas repetidas e uma entre grupos).

⁵ Embora isso as punisse por qualquer tentativa de uso do celular, porque os celulares de outras pessoas também emitem microondas; um efeito colateral infeliz foi que essas crianças adquiriram um medo patológico de qualquer pessoa usando um celular!

As respostas estão no arquivo **Answers (Chapter12).pdf**, disponível no *site* www.artmed.com.br. Alguns comentários mais detalhados sobre a tarefa 2 podem ser encontrados em Field e Hole (2003).

12.10 LEITURAS COMPLEMENTARES

HOWELL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury. 2002. 5ª edição.

13

TESTES NÃO PARAMÉTRICOS

13.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①



Vimos nos últimos capítulos como utilizar várias técnicas para determinar diferenças entre médias. Contudo, todos os testes apresentados têm por base hipóteses paramétricas (principalmente dados normalmente distribuídos). No Capítulo 3, vimos que os dados nem sempre são tão amigáveis nem aparecem em pacotes normalmente distribuídos. Ainda, nem sempre é possível corrigir os problemas da distribuição dos dados. O que fazer nesses casos? A resposta é que devemos utilizar procedimentos estatísticos especiais chamados de testes não-paramétricos. Esses testes são também conhecidos como testes de distribuição livre porque fazem poucas – ou nenhuma – suposições sobre o tipo de dados que pode ser

utilizado.¹ Muitos desses testes trabalham com a ideia de categorizar os dados: isto é, encontrar o menor valor e atribuir a ele um posto (posição) 1, achar o próximo maior valor e atribuir a ele o posto 2 e assim por diante. Esse processo resulta em escores altos sendo representados por altos postos e escores baixos sendo representados por postos baixos. A análise é então executada sobre os postos (*ranks*) em vez de sobre os dados observados. O processo é uma forma engenhosa de utilizar dados que não se encaixam nas hipóteses paramétricas. Algumas pessoas acreditam que os testes paramétricos têm menos poder do que seus contrapartes paramétrico, mas como veremos no Quadro 13.2, isso nem sempre é verdadeiro. Neste capítulo, veremos quatro dos procedimentos não-paramétricos mais comuns: os testes de Mann-Whitney, por postos de Wilcoxon, de Friedman e de Kruskal-Wallis. Para cada um deles, iremos descobrir como realizar a análise com o SPSS e como interpretar e relatar os resultados.

¹ Testes não-paramétricos são normalmente denominados testes de distribuição livre, com a explicação de que não é necessário suposições sobre a distribuição dos dados. Tecnicamente isso não é verdadeiro: eles de

fato exigem hipóteses sobre distribuições (por exemplo, todos os deste capítulo pressupõem que a distribuição é contínua), mas eles são menos restritivos que os paramétricos.

13.2 COMPARANDO DUAS CONDIÇÕES INDEPENDENTES: O TESTE DA SOMA DOS POSTOS DE WILCOXON E O TESTE DE MANN-WHITNEY ①

Quando queremos testar diferenças entre duas condições e diferentes participantes foram selecionados em cada condição, temos duas escolhas: o teste de Mann-Whitney (Mann e Whitney, 1947) e o teste da soma dos postos de Wilcoxon (Wilcoxon, 1945). Esses testes são as versões não-paramétricas equivalentes ao teste paramétrico *t* (Student). De fato, os dois testes são equivalentes e ainda existe outro, mais famoso, o teste de Wilcoxon – é difícil não nos sentirmos confusos.

Por exemplo, uma neurologista quer realizar um experimento para investigar o efeito depressivo de determinada substância usada para o divertimento. Ela testou, ao todo, 20 frequentadores de boates: a 10 deles foi dado um tablete de *ecstasy* para tomar no sábado a noite e aos outros 10 foi permitido ingerir somente álcool. Os níveis de depressão foram mensurados utilizando o BDI (Inventário de Depressão de Beck, *BDI – Beck Depression Inventory*) no dia posterior e no meio da semana. Os dados estão na Tabela 13.1.

13.2.1 Teoria ②



A lógica por trás dos testes da soma de postos de Wilcoxon e de Mann-Whitney são incrivelmente elegantes. Primeiro, vamos imaginar um cenário onde não existem diferenças nos níveis de depressão entre os usuários de *ecstasy* e álcool. Se fôssemos organizar (ordenar) os dados ignorando a qual grupo a pessoa pertence a partir do menor para o maior (isto é, atribuindo ao menor valor o posto 1 e ao próximo o posto 2 e assim por diante), o que deveríamos encontrar? Bem, se não existe diferença entre os dois grupos, esperaríamos encontrar um número semelhante de postos altos e baixos em cada grupo; especificamente, se somarmos os postos, esperaríamos que a soma total dos postos em cada grupo fosse aproximadamente a mesma.



Figura 13.1 Frank Wilcoxon.

Tabela 13.1 Dados para o experimento com substâncias

Participante	Droga	BDI (Domingo)	BDI (Quarta)
1	<i>Ecstasy</i>	15	28
2	<i>Ecstasy</i>	35	35
3	<i>Ecstasy</i>	16	35
4	<i>Ecstasy</i>	18	24
5	<i>Ecstasy</i>	19	39
6	<i>Ecstasy</i>	17	32
7	<i>Ecstasy</i>	27	27
8	<i>Ecstasy</i>	16	29
9	<i>Ecstasy</i>	13	36
10	<i>Ecstasy</i>	20	35
11	Álcool	16	5
12	Álcool	15	6
13	Álcool	20	30
14	Álcool	15	8
15	Álcool	16	9
16	Álcool	13	7
17	Álcool	14	6
18	Álcool	19	17
19	Álcool	18	3
20	Álcool	18	10

Agora pense sobre o que aconteceria se existisse uma diferença entre os dois grupos. Vamos imaginar que o grupo do *ecstasy* tenha

mais depressivos do que o do álcool. Se repetirmos o procedimento anterior, esperaríamos que os postos maiores estivessem no grupo do *ecstasy* e os postos mais baixos no grupo do álcool. Novamente, se somarmos os postos de cada grupo, esperaríamos que a soma dos postos fosse maior no grupo do *ecstasy* do que no do álcool.

Os testes de Mann-Whitney e da soma de postos de Wilcoxon trabalham com essa ideia. De fato, quando os grupos apresentam um número diferente de participantes, a estatística teste (W_s) para o teste da soma dos postos de Wilcoxon é simplesmente a soma dos postos no grupo que contém o menor número de participantes; quando os tamanhos dos grupos são iguais, a estatística é a menor das duas somas. Vamos dar uma olhada em como isso funciona na prática.



A Figura 13.2 mostra o processo de obtenção dos postos tanto para os dados do domingo quanto para os da quarta-feira. Para começar, vamos utilizar os dados da quarta-feira. Organize os

dados em ordem crescente, colocando um rótulo em cada valor para saber a que grupo eles pertencem (eu uso A para o grupo do álcool e E para o do *ecstasy*). Depois, partindo do menor valor, atribua o postos potenciais começando com o valor 1 e aumentando até o total existente de dados. A razão para acrescentar o termo potencial aos postos é que algumas vezes o mesmo valor ocorre mais de uma vez no conjunto global de dados (por exemplo, o valor 6 ocorre duas vezes e o valor 35, três). Nesse caso, temos um empate de valores e precisamos atribuir a eles o mesmo posto, assim, o que fazemos é atribuir um posto que é a média dos postos potenciais para esses valores. Desse modo, com os dois valores 6, em virtude de eles terem postos potenciais de 3 e 4, iremos atribuir a eles a média desses dois postos (3,5) e utilizar esse mesmo valor para as duas ocorrências desse escore. Faremos o mesmo com o

valor 35 que apresenta postos potenciais de 16, 17 e 18; o posto utilizado será o valor médio desses três postos potenciais, isto é, $(16 + 17 + 18)/3 = 17$. Quando tivermos terminado de ordenar os dados, somamos os postos de cada um dos dois grupos. Assim, some os postos dos escores do grupo do álcool (você deve encontrar o valor 59) e faça o mesmo para o grupo do *ecstasy* (esse valor deverá ser 151). Tomamos a menor dessas duas somas como nossa estatística teste; dessa forma, a estatística de Wilcoxon para os dados de Quarta-Feira é $W_s = 59$.

Agora, como faremos o mesmo com os dados de domingo, que apresentam muitos empates e é geralmente horrível! A resposta está na Figura 13.2; você deve encontrar que quando ordenar os dados e somar os postos para os dois grupos, a soma dos postos para o grupo do Álcool é 90,5 e para o grupo de *Ecstasy*, 119,5. Tomamos o menor dessas duas somas como nossa estatística; portanto, a estatística teste para os dados de domingo é $W_s = 90,5$.

A próxima questão é: como determinar se esse valor é significativo? Sabemos que a média (\bar{W}_s) e o erro padrão dessa estatística (EP_{W_s}) podem ser facilmente calculados a partir dos tamanhos dos dois grupos (n_1 sendo o tamanho do grupo 1 e n_2 o tamanho do grupo 2):

$$\bar{W}_s = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$EP_{W_s} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Para os nossos dados que, de fato, apresentam grupos de mesmo tamanho com 10 pessoas em cada um, então $n_1 = n_2 = 10$. Portanto, a média e o erro padrão são:

$$\bar{W}_s = \frac{10(10 + 10 + 1)}{2} = 105$$

$$EP_{W_s} = \sqrt{\frac{(10 \times 10)(10 + 10 + 1)}{12}} = 13,23$$

Se soubermos a estatística teste, a média e o erro padrão da estatística teste, podemos facilmente converter a estatística teste em um

	Dados de quarta-feira																			
Escore	3	5	6	6	7	8	9	10	17	24	27	28	29	30	32	35	35	35	36	39
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Posto Real	1	2	3.5	3.5	5	6	7	8	9	10	11	12	13	14	15	17	17	17	19	20
Grupo	A	A	A	A	A	A	A	A	A	E	E	E	E	A	E	E	E	E	E	E
	Soma dos postos para o Álcool (A) = 59										Soma dos postos para o Ecstasy (E) = 151									
	Dados de domingo																			
Escore	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Posto	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Posto Real	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Grupo	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E
	Soma dos postos para o Álcool (A) = 90,5										Soma dos postos para o Ecstasy (E) = 119,5									

Figura 13.2 Postos dos escores de depressão de quarta-feira.

escore- z , utilizando a equação que já foi vista no Capítulo 1:

$$z = \frac{X - \bar{X}}{s} = \frac{W_s - \bar{W}_s}{EP_{W_s}}$$

Se calcularmos esses valores para os escores de depressão de domingo e quarta-feira, obteremos:

$$z_{\text{Domingo}} = \frac{W_s - \bar{W}_s}{EP_{W_s}} = \frac{90,5 - 105}{13,23} = 1,10$$

$$z_{\text{Quarta}} = \frac{W_s - \bar{W}_s}{EP_{W_s}} = \frac{59 - 105}{13,23} = -3,48$$

Se esses valores são maiores do que 1,96 (ignorando o sinal de menos), o teste é significativo no nível de $p < 0,05$. Assim, parece que existe uma diferença significativa entre os grupos na quarta-feira, mas não no domingo.

O procedimento que descrevemos é o teste da soma dos postos de Wilcoxon. O teste de

Mann-Whitney, com o qual muitos de vocês já estão familiarizados, é basicamente o mesmo. Ele é baseado na estatística U , derivada de modo semelhante ao procedimento de Wilcoxon (de fato, existe um relacionamento direto entre os dois). Se você está interessado, U é calculado utilizando uma equação na qual n_1 e n_2 são os tamanhos dos grupos 1 e 2, respectivamente, e R_1 é a soma dos postos para o grupo 1:

$$U = N_1 N_2 + \frac{N_1(N_1 + 1)}{2} - R_1$$

Assim, para nossos dados obteremos o seguinte (lembre que temos 10 pessoas em cada grupo e que a soma dos postos para o grupo 1, o de *Ecstasy*, foi 119,5 para os dados de domingo e 151 para os dados de quarta-feira):

$$U_{\text{Domingo}} = (10 \times 10) + \frac{10(11)}{2} - 119,50 = 35,50$$

$$U_{\text{Quarta}} = (10 \times 10) + \frac{10(11)}{2} - 151,00 = 4,00$$



O SPSS produz as duas estatísticas e existe um relacionamento direto entre as duas, portanto, não importa qual delas for utilizada!

13.2.2 Atribuindo dados e análise provisional ①

Quando os dados são coletados com um número diferente de participantes em cada grupo, precisamos entrar com os dados utilizando uma variável codificadora. Assim, o editor de dados terá três colunas de dados. A primeira coluna será a variável código (que pode ser denominada **drug** (droga)) que, nesse caso, terá somente dois valores (por conveniência, sugiro 1 = grupo do *Ecstasy* e 2 = grupo do Álcool). A segunda coluna apresentará os valores da variável dependente (BDI) medida no dia seguinte (chame essa variável de **sunbdi**) e a terceira terá os escores do meio da semana do mesmo questionário (chame essa variável de **wedbdi**).

Quando você inserir os dados no SPSS, informe ao computador que o código 1 representa o grupo do *Ecstasy* e que o código 2 representa o grupo do Álcool (veja a Seção 2.4.4). Não existe uma previsão específica sobre qual droga terá o maior efeito e, assim, a análise deve ser bilateral. Primeiro, podemos rodar uma análise exploratória nos dados e, em virtude de estarmos procurando diferenças entre os grupos, precisamos fazer essa análise exploratória para cada um dos grupos (veja as Seções 3.5 e 3.6). Se você fizer essas análises deverá encontrar tabelas idênticas às apresentadas na Saída 13.1 do SPSS. Essas tabelas mostram primeiro que para os dados de domingo a distribuição para o *Ecstasy*, $D(10) = 0,28$, $p < 0,05$, parece não ser normal enquanto a dos dados do Álcool, $D(10) = 0,17$, ns, é normal: podemos dizer isso se a significância dos testes K-S e Shapiro-Wilk são menores do que 0,05 (e, dessa forma, significativas) ou maior do que 0,05 (e, portanto, não-significativas (ns)). Para os dados de quarta-feira, embora os dados para o *Ecstasy* sejam normais, $D(10) = 0,24$, ns, os dados para o Álcool parecem ser significativamente não-normais $D(10) = 0,31$, $p <$

0,01. Esses achados nos alertam para o fato de que um teste não-paramétrico deve ser utilizado nos dados de domingo e quarta-feira porque uma das variáveis de cada não é normalmente distribuída. Você deve notar que a estatística de Shapiro-Wilk fornece valores com significância exata enquanto o teste K-S, às vezes, fornece uma aproximação de 0,2 para a significância (veja os dados de domingo para o grupo do Álcool), porque o SPSS não pode calcular significâncias exatas. Esse fato ressalta uma diferença importante entre os testes K-S e Shapiro-Wilk: em geral, o teste de Shapiro-Wilk é mais preciso. A segunda tabela na Saída 13.1 do SPSS mostra os resultados do teste de Levene. Para os dados de domingo $F(1, 18) = 3,64$, ns, e para os de quarta-feira, $F(1, 18) 0,51$, ns, as variâncias não são significativamente diferentes, indicando que a hipótese de homogeneidade foi satisfeita. Não obstante, precisamos utilizar procedimentos estatísticos não-paramétricos nos dados de domingo e quarta-feira em virtude da não normalidade dos dados.

13.2.3 Executando a análise ①

Primeiro, acesse a caixa de diálogo principal utilizando o seguinte caminho (menu) **Analyze⇒Nonparametric Tests⇒2 Independent Samples...** (Analisar⇒Testes não-paramétricos⇒Duas Amostras Independentes...) (veja a Figura 13.3). Com essa caixa de diálogo ativada, selecione as duas variáveis dependentes da lista (clique no Inventário de Depressão de Beck domingo [**sunbdi**], depois, pressionando o botão do *mouse*, arraste Inventário de Depressão de Beck quarta-feira [**wedbdi**] e transfira as duas para o quadro denominado **Test Variable List** (Lista de Variáveis de Teste) clicando em . A seguir, selecione a variável independente (a variável de grupo) – nesse caso, **drug** (tipo de droga) – e transfira-a para o quadro denominado **Grouping Variable** (Variável de Agrupamento). Quando tiver feito isso, o botão **Define Groups...** ficará ativo e você deve clicar nele para ativar a caixa de diálogo **define groups** (definir grupos). O SPSS precisa saber quais códigos você atribuiu para cada grupo e existe um espaço para

Saída 13.1 do SPSS


Tests of Normality (Testes de Normalidade)


	Type of Drug (Tipo de Droga)	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic (Estatística)	df (gl)	Sig. (Sig.)	Statistic (Estatística)	df (gl)	Sig. (Sig.)
Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))	Ecstasy (Ecstasy)	0.276	10	0.030	0.811	10	0.020
	Alcohol (Álcool)	0.170	10	0.200*	0.959	10	0.780
Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta-feira))	Ecstasy (Ecstasy)	0.235	10	0.126	0.941	10	0.566
	Alcohol (Álcool)	0.305	10	0.009	0.753	10	0.004

* This is a lower bound of the true significance (Esse é um limite inferior da verdadeira significância)
a Lilliefors Significance Correction (Correção da significância de Lilliefors)

Test of Homogeneity of Variance (Teste de Homogeneidade da Variância)

		Levene Statistic (Estatística de Levene)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))	Based on Mean (Com base na Média)	3.644	1	18	0.072
	Based on Median (Com base na mediana)	1.880	1	18	0.187
	Based on Median and with adjusted df (Com base na mediana e com gl ajustado)	1.880	1	10.076	0.200
	Based on on trimmed mean (Com base na média interna)	2.845	1	18	0.109
Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta-feira))	Based on Mean (Com base na Média)	0.508	1	18	0.485
	Based on Median (Com base na mediana)	0.091	1	18	0.766
	Based on Median and with adjusted df (Com base na mediana e com gl ajustado)	0.091	1	11.888	0.768
	Based on trimmed mean (Com base na média interna)	0.275	1	18	0.606

digitar esses valores. Nesse exemplo, o grupo do *Ecstasy* foi codificado como 1 e o grupo do *Álcool*, como 2, e você deve digitar esses dois valores nos espaços apropriados. Quando tiver definido os grupos, clique em  para retornar à caixa de diálogo principal. Essa caixa fornece também opções para realizar outros testes além do de Mann-Whitney e essas alternativas são apresentadas no Quadro 13.1.

Se você clicar em , outra caixa de diálogo aparecerá.² Por padrão, o SPSS cal-

cula a significância do teste de Mann-Whitney utilizando um método que é preciso com amostras grandes (denominado **Asymptotic Method** (Método Assintótico); contudo, quando as amostras são pequenas, ou os dados são pobremente distribuídos, métodos mais precisos estão disponíveis. A melhor opção é solicitar um teste exato (*Exact*), que calcula a significância do teste de Mann-Whitney de forma exata. Contudo, existe um preço para obter tal precisão e, em virtude da complexidade dos cálculos, o SPSS pode demorar para encontrar uma solução, especialmente com grandes amostras. Um método menos trabalhoso é estimar a significância utilizando o

² Esse botão aparecerá somente se você tiver o módulo *Exact tests* (Testes exatos) do SPSS instalado. Lembre disso também nas próximas seções.

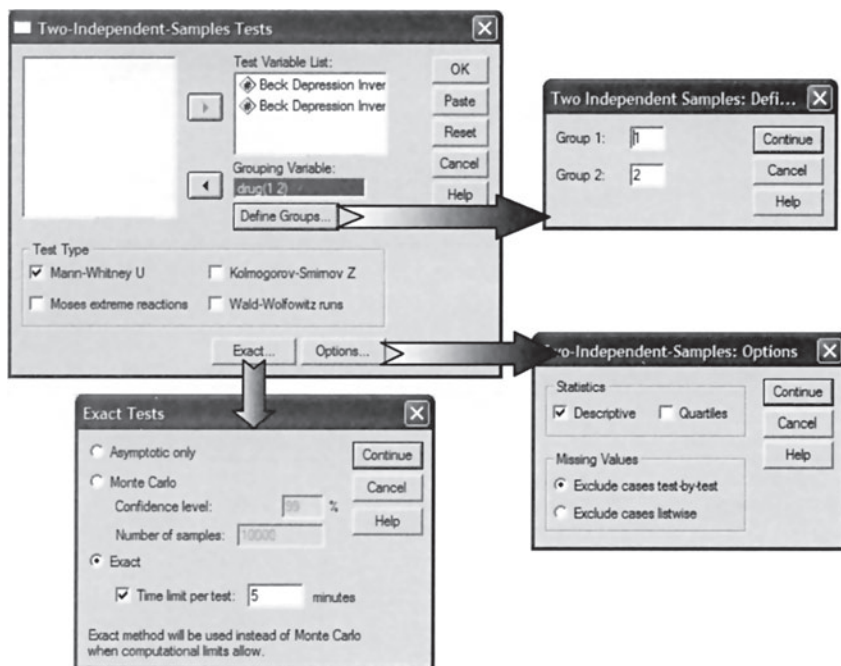


Figura 13.3 Caixas de diálogo para o teste de Mann-Whitney.

Quadro 13.1

Outras opções para o teste de Mann-Whitney ②

Na caixa de diálogo principal existem algumas outras opções que podem ser selecionadas:

- O teste de **Kolmogorov-Smirnov** (*Kolmogorov-Smirnov Z*): No Capítulo 3, aprendemos sobre o teste de Kolmogorov-Smirnov, utilizado para testar se uma população era normalmente distribuída. Esse é um teste diferente! De fato, ele testa se dois grupos foram retirados de uma mesma população (não importando o tipo de população). Isso significa que ele é muito semelhante ao teste de Mann-Whitney! Contudo, o teste K-S tende a ter mais poder do que o de Mann-Whitney quando os tamanhos amostrais são inferiores a 25 por grupo. Assim, vale a pena selecioná-lo se esse for o caso.
- **Reações extremas de Moses** (*Moses Extreme Reactions*): Grande nome – me faz pensar em um homem barbudo no alto do Monte Sinai lendo umas tábuas de pedra e que, subitamente irado, quebra as tábuas e grita: “o que você quer dizer com não louvar outro Deus?”. Infelizmente, esse teste não é tão excitante quanto a minha imaginação. Ele basicamente compara a variabilidade dos escores nos dois grupos, assim, ele é um pouco semelhante ao teste não-paramétrico de Levene.
- **Corridas de Wald-Wolfowitz** (*Wald-Wolfowitz runs*): Apesar do nome estranho, é outra variante do teste de Mann-Whitney. Nesse teste, os escores são transformados em postos e ordenados da mesma forma que no teste de Mann-Whitney, mas em vez de analisar os postos, esse teste procura “corridas” ou sequências de um mesmo grupo dentro dos postos ordenados. Assim, se não existirem diferenças entre os dois grupos, os postos dos dois grupos devem estar aleatoriamente interespçados. Contudo, se os grupos diferem, você deve ver mais postos de um grupo no início e mais postos do outro grupo no final. Procurando por aglomerados de postos, o teste pode determinar se os dois grupos diferem.

método de Monte Carlo. Isso basicamente envolve criar uma distribuição semelhante àque-la encontrada na amostra e tomar várias amos-tras (o padrão é 10.000) dessa distribuição; a partir dessas amostras, o valor médio da signi-ficância e um intervalo de confiança para esse valor podem ser determinados. Se isso não faz sentido algum para você, não tenha medo: a regra prática é que quando tiver grandes ta-manhos amostrais em geral você deve optar pelo método de Monte Carlo; quando tiver tamanhos amostrais pequenos (como nesse caso), é melhor utilizar o método exato (como foi feito neste exemplo). Finalmente, clican-do em **Options...**, uma nova caixa de diálogo que fornece novas opções para a análise é aberta. Essas opções não são muito úteis porque, por exemplo, a opção que fornece as estatísticas descritivas faz isso para todos os dados (não considerando os grupos). Por esse motivo, re-comendo obter as estatísticas descritivas uti-lizando os métodos que foram apresentados nas Seções 3.4 e 3.5. Para executar a análise, retorne à caixa de diálogo principal e clique em **OK**.

13.2.4 Saídas para o teste de Mann-Whitney ①

Expliquei na Seção 13.2.1 que o teste de Mann-Whitney funciona procurando por dife-renças nas posições ordenadas dos escores nos diferentes grupos. Portanto, a primeira parte da saída resume os dados após eles terem sido ordenados (*ranqueados*). Especificamente, o SPSS informa o total e a média dos postos em cada condição (veja a Saída 13.2 do SPSS).

Lembre que o teste de Mann-Whitney se baseia nos escores transformados em postos: portanto, o grupo com a menor média é o grupo com o maior número de postos baixos. De forma se-melhante, o grupo com a maior média de postos deve ter o maior número de postos altos. Por-tanto, essa tabela inicial pode ser utilizada para assegurar que grupo tem os maiores postos, o que é útil no caso de precisarmos interpretar um resultado significativo. Você deve notar que a soma dos postos são as mesmas calculadas na Seção 13.2.1 (o que é um alívio para mim!).

A segunda tabela (Saída 13.2 do SPSS) fornece as estatísticas reais para o teste de Mann-Whitney, o procedimento de Wilco-xon e o correspondente escore-z (veja a Seção 13.2.1). A Saída 13.3 do SPSS tem uma colu-na para cada variável (uma para **sunbdi** e uma para **wedbdi**) e em cada coluna existe um va-lor da estatística U de Mann-Whitney, o valor da estatística de Wilcoxon e uma aproximação do valor associado z. Note que os valores de U, W_s e do escore-z associado são os mesmos que foram calculados na Seção 13.2.1!

A parte importante da tabela é o valor da significância do teste, que fornece a probabili-dade bilateral de que a magnitude da estatística teste seja um resultado casual. Esse valor da sig-nificância pode ser utilizado quando nenhuma previsão tenha sido feita sobre qual grupo irá diferir. Contudo, se uma previsão tiver sido feita (por exemplo, se dissermos que os usuários de *Ecstasy* terão mais depressão do que os usuários de Álcool no dia seguinte), precisamos calcular a probabilidade unilateral dividindo por dois o valor da significância bilateral. Para esses da-

Saída 13.2 do SPSS

Ranks (Postos)				
	Type of Drug (Tipo de Droga)	N	Mean Rank (Média dos Postos)	Sum of Ranks (Soma dos Postos)
Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))	Ecstasy (Ecstasy)	10	11.95	119.50
	Alcohol (Álcool)	10	9.05	90.50
	Total (Total)	20		
Beck Depression Inventory (Wed- nesday) (Inventário de Depressão de Beck (Quarta-feira))	Ecstasy (Ecstasy)	10	15.10	151.00
	Alcohol (Álcool)	10	5.90	59.00
	Total (Total)	20		

dos, o teste de Mann-Whitney não é significativo (bilateral) para os escores de depressão coletados no Domingo. Esse resultado indica que o *Ecstasy* não é mais depressivo, no dia seguinte, do que o *Álcool*: os dois grupos têm o mesmo nível de depressão. Contudo, para as medidas do meio da semana, os resultados são altamente significativos ($p < 0,001$). O valor médio dos postos indica que o grupo do *Ecstasy* apresenta níveis significativamente mais altos de depressão no meio da semana do que o grupo do *Álcool*. Essa conclusão é obtida observando que os escores da quarta-feira apresentam um posto médio bem mais alto dos usuários de *Ecstasy* (15,10) do que dos usuários de *Álcool* (5,90).

A Saída 13.4 do SPSS mostra a saída para o teste de Mann-Whitney quando a opção pela significância exata é selecionada. Inclui isso para lhe mostrar que aparecem algumas linhas extras que fornecem os valores da significância exata (tanto unilateral quanto bilateral). Esses valores não alteram nossas conclusões, mas saiba que você deve provavelmente consultar esses valores em vez do valor assintótico, especialmente quando os tamanhos amostrais são pequenos.

Saída 13.4 do SPSS (Com a significância exata por Monte Carlo)

Test Statistics^a (Estatísticas Teste)		
	<i>Beck Depression Inventory</i> (Sunday) (Inventário de Depressão de Beck (domingo))	<i>Beck Depression Inventory</i> (Wednesday) (Inventário de Depressão de Beck (quarta-feira))
<i>Mann-Whitney U</i> (U de Mann-Whitney)	35.500	4.000
<i>Wilcoxon W</i> (W de Wilcoxon)	90.500	59.000
<i>Z</i>	-1.105	-3.484
<i>Asymp. Sig. (2-tailed)</i> (Sig. Assint.) (Bilateral)	0.269	0.000
<i>Exact Sig. [2*(1-tailed Sig.)]</i> (Sig. Exata) [2*(Sig. Unilateral)	0.280 ^a	0.000 ^a
<i>Exact Sig. (2-tailed.)</i> (Sig. Exata) (Bilateral)	0.288	0.000
<i>Exact Sig. (1-tailed)</i> (Sig. Exata) (Unilateral)	0.144	0.000
<i>Point Probability</i> (Probabilidade Pontual)	0.013	0.000

a Não corrigido para empates.
b Variável de Agrupamento: Tipo de Droga.

Saída 13.3 do SPSS (Sem a significância exata por Monte Carlo)

Test Statistics^b (Estatísticas Teste)		
	<i>Beck Depression Inventory</i> (Sunday) (Inventário de Depressão de Beck (Domingo))	<i>Beck Depression Inventory</i> (Wednesday) (Inventário de Depressão de Beck (Quarta-Feira))
<i>Mann-Whitney U</i> (U de Mann-Whitney)	35.500	4.000
<i>Wilcoxon W</i> (W de Wilcoxon)	90.500	59.000
<i>Z</i>	-1.105	-3.484
<i>Asymp. Sig. (2-tailed)</i> (Sig. Assint.) (Bilateral)	0.269	0.000
<i>Exact Sig. [2*(1-tailed Sig.)]</i> (Sig. Exata) [2*(Sig. Unilateral)	0.280 ^a	0.000 ^a

a Não corrigido para empates.
b Variável de Agrupamento: Tipo de Droga.

13.2.5 Calculando o tamanho de efeito ②

Já ressaltamos ao longo do texto a importância de relatar o tamanho de efeito de forma que as pessoas tenham uma medida padronizada do valor que você observou e, assim, eles possam comparar com os seus ou outros estudos. O SPSS não calcula o tamanho de efeito, mas podemos calcular tamanhos de efeito aproximados facilmente porque o SPSS converte a estatística teste em um escore- z . A equação que converte um escore- z em uma estimativa do tamanho de efeito, r , é a seguinte (Rosenthal, 1991, p. 19):

$$r = \frac{Z}{\sqrt{N}}$$

Onde z é o escore- z calculado pelo SPSS e N é o tamanho dos valores do estudo (isto é, o total de observações realizadas) na qual

z foi baseado. Nesse caso, a Saída 13.3 do SPSS informa que o valor de z é $-1,11$ para os dados de domingo e $-3,48$ para os dados de quarta-feira. Nos dois casos, tivemos 10 usuários de *Ecstasy* e 10 usuários de *Álcool* e, assim, o número total de observações é 20. Os tamanhos de efeito são, portanto:

$$r_{\text{Domingo}} = \frac{-1,11}{\sqrt{20}} = -0,25$$

$$r_{\text{Quarta}} = \frac{-3,48}{\sqrt{20}} = -0,78$$

Isso representa um pequeno para médio efeito para os dados de domingo (isto é, um valor abaixo do critério 0,3 para ser considerado um efeito médio) e um grande efeito para a quarta-feira (o valor está bem acima do limite de 0,5 considerado um grande efeito. Muito importante: os dados de domingo mostram que um tamanho de efeito razoavelmente grande pode não ser significativo quando temos uma amostra pequena!

13.2.6 Escrevendo os resultados ①

Para o teste de Mann-Whitney, precisamos apenas relatar a estatística teste (que é indica-

da por U) e sua significância. É claro, também podemos incluir o tamanho de efeito. Assim, o relatório pode ser semelhante ao seguinte:

- ☑ Usuários de *Ecstasy* (Mediana = 17,50) parecem não diferir quanto ao nível de depressão dos usuários de *álcool* (Mediana = 16,00) no dia seguinte ao uso da droga, $U = 35,50$, ns, $5 = -0,25$. Contudo, na quarta-feira, os usuários de *Ecstasy* (Mediana = 33,50) estavam significativamente mais deprimidos que os usuários de *álcool* (Mediana = 7,50), $U = 4,00$, $p < 0,001$, $r = -0,78$.

Note que relatei a mediana para cada condição – essa estatística é mais apropriada que a média para testes não-paramétricos. Podemos, ainda, escolher relatar a estatística teste U de Wilcoxon em vez da de Mann-Whitney e isso poderia ser feito da seguinte forma:

- ☑ Usuários do *Ecstasy* (Mediana = 17,50) não parecem diferir nos níveis de depressão dos usuários de *Álcool* (Mediana = 16,00) no dia seguinte ao uso da droga, $W_s = 90,50$, ns, $r = -0,25$. Contudo, na quarta-feira, os usuários de *Ecstasy* (Mediana = 33,50) estavam significativamente

Dica da Samanta Ferrinho



- O teste de Mann-Whitney e a soma dos postos de Wilcoxon **comparam duas condições quando diferentes pessoas participam em cada condição e os dados resultantes não são normalmente distribuídos ou violam a hipótese do teste t independente**.
- Olhe para a linha denominada Sig. Assint. (bilateral) (*Asymp. Sig. (2-tailed)*): se o valor for menor do que 0,05, as médias dos dois grupos são significativamente diferentes. (Se você optou por um teste exato, olhe a coluna denominada Sig. Exata (bilateral) (*Exact. Sig. (2-tailed)*):
- Olhe para o valor dos postos para saber como os grupos diferem (o grupo com os escores mais altos apresentará também os postos mais altos).
- O SPSS fornece apenas uma significância bilateral; se você quiser a significância unilateral, basta dividir o valor por 2 (se você optou pelo teste exato, olhe para a linha denominada Sig. Exata (unilateral) (*Exact. Sig. (1-tailed)*):
- Relate a estatística U (ou a W_s se preferir), o valor z correspondente e o valor da significância. Relate, ainda, as medianas e suas amplitudes correspondentes (ou trace um diagrama de caixa e bigodes).
- Você deve calcular o tamanho de efeito e relatá-lo também!

Quadro 13.2

Testes não-paramétricos e poder estatístico ②



Atribuir postos aos dados é uma maneira útil de contornar as hipóteses paramétricas sobre a forma das distribuições, mas há um preço a pagar: transformando os dados em postos, perdemos informações sobre a magnitude da diferença entre os valores. O resultado é que os testes não-paramétricos podem ser menos poderosos do que os correspondentes paramétricos. Apresentei a noção de poder estatístico na Seção 1.8.5: ela se refere à habilidade de um teste encontrar um efeito que, de fato, exista. Assim, afirmar que os testes não-paramétricos são menos poderosos significa que se existe um efeito genuíno nos nossos dados, um teste paramétrico terá maior probabilidade de detectá-lo do que o correspondente não-paramétrico. Contudo, essa afirmação é verdadeira somente se as hipóteses do teste paramétrico forem satisfeitas. Desse modo, se utilizarmos um teste paramétrico e um não-paramétrico nos mesmos dados e esses dados são normalmente distribuídos, o teste paramétrico terá um poder maior de detectar o efeito do que o teste não-paramétrico.

O problema é que para definir o poder de um teste, precisamos estar certos de que ele controla a taxa de erro do Tipo I (o número de vezes que um teste encontrará um efeito significativo quando, de fato, não existir tal efeito – veja a Seção 1.8.2). Vimos no Capítulo I que Fisher disse que esse erro deveria ser de 5%. Sabemos que quando os dados são normalmente distribuídos, a taxa de erro do Tipo I de um teste baseado nessa distribuição é 5% e, assim, podemos determinar o poder. Contudo, quando os dados não são normais, a taxa de erro do Tipo I de um teste baseado na distribuição não será de 5% (de fato, não sabemos qual será essa taxa, pois ela depende da forma da distribuição); desse modo, não temos uma maneira de calcular o poder (porque o poder está relacionado à taxa de erro do Tipo I – veja a Seção 1.8.5). Assim, embora você frequentemente leia (na primeira edição deste livro, por exemplo) que os testes não-paramétricos aumentam a probabilidade de erro do Tipo II (isto é, mais chance de aceitar o fato de que não existe uma diferença entre os grupos, quando ela, de fato, existe), isso é verdadeiro apenas se os dados são normalmente distribuídos.



te mais deprimidos que os usuários de Álcool (Mediana = 7,50), $W_s = 59,00$, $p < 0,001$, $r = -0,78$.

13.3 COMPARANDO DUAS CONDIÇÕES RELACIONADAS: O TESTE DOS POSTOS COM SINAIS DE WILCOXON ①

O teste dos postos com sinais de Wilcoxon (Wilcoxon, 1945) – não confunda com o teste da soma dos postos da seção anterior – é utilizado em situações em que existem dois conjuntos de escores a serem comparados, mas esses escores são provenientes dos mesmos participantes. Pense nele como um equivalente não-paramétrico ao teste t dependente (ou ao teste de Mann-Whitney para dados com medidas repetidas). Imagine que o pesquisador na se-

ção anterior estava interessado na *mudança* nos níveis de depressão, entre as pessoas, para cada uma das duas drogas. Agora queremos comparar os escores do BDI (Inventário de Depressão de Beck) do domingo com os da quarta-feira. Ainda temos que utilizar um teste não-paramétrico porque as distribuições dos escores para as duas drogas não são normais em um dos dois dias (veja a Saída 13.1 do SPSS).

13.3.1 Teoria do teste dos postos com sinais de Wilcoxon ②



O teste dos postos com sinais de Wilcoxon funciona de forma semelhante ao teste t para variáveis dependentes (Capítulo 7), uma vez que ele tem por base as diferenças entre os escores nas duas con-

dições que estamos comparando. Depois de calculadas, essas diferenças são transformadas em postos (da mesma forma que em 13.2.1), mas o sinal da diferença (positiva ou negativa) é atribuído a cada posto. Se utilizarmos os mesmos dados de antes, podemos comparar os escores de depressão do domingo com os de quarta-feira para as duas drogas separadamente.

A Tabela 13.2 mostra os postos para esses dados. Lembre que estamos determinando os postos das duas drogas em separado. Primeiro, calcule as diferenças entre domingo e quarta-feira (isto é, apenas subtraia os valores de domingo dos de quarta-feira). Se a diferença é zero (isto é, os valores dos dois dias são iguais), exclua esses dados da análise. Anota-

mos o sinal das diferenças (isto é, quais escores são positivos ou negativos) e ordenamos as diferenças (começando com a menor) ignorando se ela é positiva ou negativa. É a mesma maneira da Seção 13.2.1 e lidamos com escores empatados do mesmo modo. Finalmente, juntamos todas as diferenças que apresentaram o mesmo sinal e as somamos. A soma das diferenças positivas é representada por T_+ e as negativas, por T_- . Assim, para o *Ecstasy*, $T_+ = 36$ e $T_- = 0$ (não tivemos escores negativos), e para o *Álcool*, $T_+ = 8$ e $T_- = 47$. A estatística teste, T , é a menor dos dois valores, e é 0 para a *Ecstasy* e 8 para o *Álcool*.

Para calcular a significância da estatística teste (T), novamente olhamos para a média

Tabela 13.2 Postos no testes dos postos com sinais de Wilcoxon

BDI (Domingo)	BDI (Quarta-feira)	Diferença	Sinal	Posto	Posto Positivo	Posto Negativo
Ecstasy						
15	28	13	+	2,5	2,5	
35	35	0	Excluir			
16	35	19	+	6	6	
18	24	6	+	1	1	
19	39	20	+	7	7	
17	32	15	+	4,5	4,5	
27	27	0	Excluir			
16	29	13	+	2,5	2,5	
13	36	23	+	8	8	
20	35	15	+	4,5	4,5	
Total =					36	0
Álcool						
16	5	-11	-	9		9
15	6	-9	-	7		7
20	30	10	+	8	8	
15	8	-7	-	3,5		3,5
16	9	-7	-	3,5		3,5
13	7	-6	-	2		2
14	6	-8	-	5,5		5,5
19	17	-2	-	1		1
18	3	-15	-	10		10
18	10	-8	-	5,5		5,5
Total =					8	47

(\bar{T}) e para o erro padrão (EP_T), que, da mesma forma que no teste da soma dos postos e o de Mann-Whitney na seção anterior, são funções do tamanho da amostra, n (porque utilizamos os mesmos participantes, existe apenas um tamanho amostral):

$$\bar{T} = \frac{n(n+1)}{4}$$

$$EP_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Nos dois grupos, n é simplesmente 10 (porque esse é o número de participantes que foi utilizado). Contudo, lembre que para o grupo do *Ecstasy*, duas pessoas foram excluídas por apresentarem uma diferença igual a zero; portanto, o tamanho da amostra a ser utilizado é 8 e não 10. Assim, temos:

$$\bar{T}_{Ecstasy} = \frac{8(8+1)}{4} = 18$$

$$EP_{T_{Ecstasy}} = \sqrt{\frac{8(8+1)(16+1)}{24}} = 7,14$$

Não houve exclusões para o grupo Álcool, assim, temos:

$$\bar{T}_{Álcool} = \frac{10(10+1)}{4} = 27,50$$

$$EP_{T_{Álcool}} = \sqrt{\frac{10(10+1)(20+1)}{24}} = 9,81$$

Como antes, se soubermos o valor da estatística teste, a sua média e o erro padrão, podemos facilmente convertê-la em um escore- z utilizando a equação que foi vista no Capítulo 1 e na seção anterior:

$$z = \frac{X - \bar{X}}{s} = \frac{T - \bar{T}}{EP_T}$$

Se calcularmos esses valores para os escores do *Ecstasy* e do Álcool, obteremos:

$$z_{Ecstasy} = \frac{T - \bar{T}}{EP_T} = \frac{0 - 18}{7,14} = -2,52$$

$$z_{Quarta} = \frac{T - \bar{T}}{EP_T} = \frac{8 - 27,5}{9,81} = -1,99$$



Se esses valores são maiores do 1,96 (ignorando o sinal), eles são significativos no nível $p < 0,05$. Assim, o teste mostra que existem diferenças significativas entre os escores de depressão das Quartas-Feiras e dos Domingos para as duas drogas (*Ecstasy* e Álcool).

13.3.2 Executando a análise ①

Para realizar a mesma análise com o SPSS, podemos utilizar os mesmos dados que foram utilizados antes, mas como queremos verificar as alterações de cada droga separadamente, precisamos utilizar o comando *split file* (dividir arquivo) e solicitar que o SPSS divida o arquivo pelo tipo de droga [**drug**]. Esse processo assegura que qualquer análise posterior será feita nos grupos do *Ecstasy* e do Álcool separadamente. Uma vez que o arquivo tenha sido dividido, selecione a caixa de diálogo para o teste de Wilcoxon utilizando o menu **Analyze**⇒**Nonparametric Tests**⇒**2 Related Samples...** (Analisar⇒Testes Não-paramétricos⇒Duas Amostras Relacionadas...) (veja a Figura 13.4). Essa caixa de diálogo permite que você selecione, também, outros testes (veja o Quadro 13.3).

Com essa caixa de diálogo ativada, selecione as duas variáveis da lista (clique na primeira variável com o *mouse* e depois na segunda). A primeira variável que você selecionar (Inventário de Depressão de Beck Domingo [**sunbdi**]) será denominada **Variable 1** (Variável 1) no quadro **Current Selections** (Seleções atuais) e a segunda variável que você selecionar (Inventário de Depressão de Beck Quarta-Feira [**wedbdi**]) aparecerá como **Variable 2** (Variável 2). Quando você tiver selecionado as variáveis, transfira as duas para o quadro **Test Pair(s) List** (Lista de Pares do Teste) clicando em . Se você quiser executar vários testes de Wilcoxon, pode selecionar outro par de variáveis e transferi-lo para a lis-

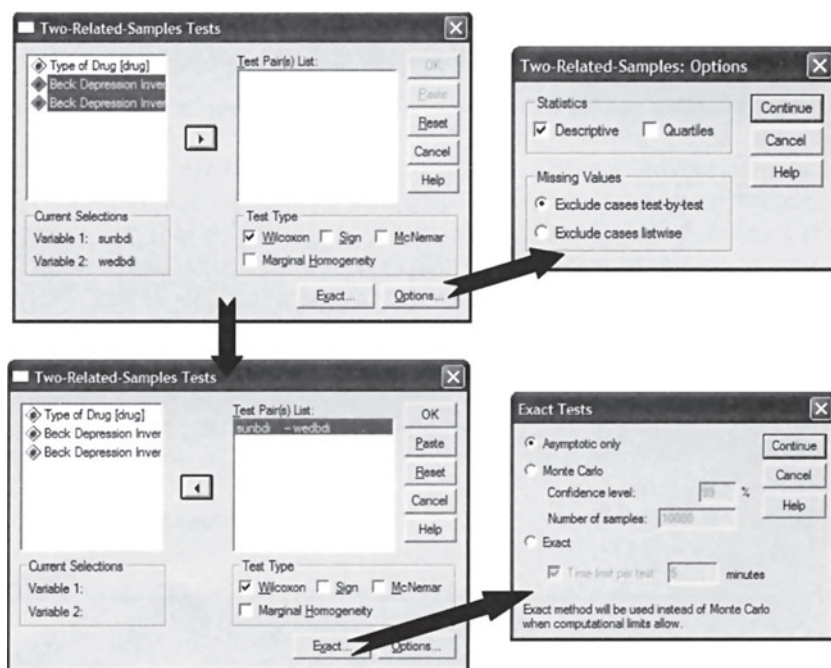


Figura 13.4 Caixas de diálogo para o teste dos postos com sinais de Wilcoxon.

Quadro 13.3

Outras opções para o teste dos postos com sinais de Wilcoxon ②

Na caixa de diálogo principal existem outros testes que podem ser selecionados:

- **Sign** (Sinal). Esse teste basicamente faz o mesmo que o teste dos postos com sinal de Wilcoxon, exceto que ele é baseado na direção das diferenças (positiva ou negativa). A magnitude da mudança é completamente ignorada (diferente do teste de Wilcoxon onde os postos nos dão informações sobre a magnitude das mudanças). Por esses motivos, o teste dos sinais tem baixo poder (ele não é bom para detectar efeitos), a menos que os tamanhos amostrais sejam pequenos (seis ou menos). Assim, francamente, não vejo por que utilizá-lo!
- **McNemar**: Esse teste é útil quando tivermos dados nominais em vez de ordinais. Ele é tipicamente utilizado quando estamos procurando por mudanças nos escores de pessoas e ele compara a quantidade de pessoas que alteraram suas respostas em uma direção (isto é, os valores aumentaram) com aqueles que mudaram na direção oposta (os valores diminuíram). Assim, esse teste precisa ser utilizado quando você tiver duas variáveis dicotômicas relacionadas.
- **Marginal Homogeneity** (Homogeneidade Marginal): Esse procedimento é uma extensão do teste de McNemar, mas para variáveis ordinais. Ele faz praticamente o mesmo que o teste de Wilcoxon.

ta de variáveis e depois selecionar outro par e assim por diante. Nesse caso, queremos testar apenas um par de variáveis. Se você clicar em **Exact...** (Exato), outra caixa de diálogo aparece e permite que a significância exata seja calculada (veja a Seção 13.2.3). Não vou retomar esse assunto novamente, mas devo dizer que quando os tamanhos amostrais são grandes deve-se optar pelo método de Monte Carlo, e quando tivermos pequenos tamanhos amostrais, é melhor optar pelo cálculo exato. Não optei por nenhum deles nesse exemplo. Se você clicar em **Options...** (Opções), uma caixa de diálogo aparece tornando possível selecionar estatísticas descritivas. Diferentemente do teste de Mann-Whitney, as estatísticas descritivas valem a pena, porque elas são uma mudança relevante entre variáveis (colunas do editor de dados). Para realizar a análise, retorne à caixa de diálogo principal e clique em **OK**.

13.3.3 Saídas para o Grupo do Ecstasy ①



Se você dividiu o arquivo, o primeiro conjunto de resultados obtido será para o grupo do *Ecstasy* (Saída 13.5 do SPSS). A primeira tabela fornece informações sobre os postos dos escores.

Ela nos informa o número de postos negativos (aquelas pessoas para as quais o escore de domingo foi maior do que o de quarta-feira) e o número de escores positivos (pessoas para as quais os escores da quarta-feira foram maiores do que os de domingo). A Tabela mostra que 8 dos 10 participantes apresentaram escores na quarta-feira maiores do que os de domingo, indicando uma grande depressão no meio da semana quando comparada com a manhã seguinte. Foram encontrados dois postos empatados (isto é, participantes com os mesmos valores nos dois dias). A tabela mostra também o número médio de postos positivos e negativos, bem como a soma desses postos. Abaixo da tabela aparecem notas informando a que os postos positivos e negativos estão se

Saída 13.5 do SPSS

Ranks^d (Postos)

		<i>N</i>	<i>Mean Rank</i> (Média dos postos)	<i>Sum of Ranks</i> (Soma dos postos)
<i>Beck Depression Inventory (Wednesday)</i> (Inventário de Depressão de Beck (Quarta-feira)) –	<i>Negative Ranks</i> (Postos Negativos)	0 ^a	0.00	0.00
<i>Beck Depression Inventory (Sunday)</i> (Inventário de Depressão de Beck (Domingo))	<i>Positive Ranks</i> (Postos Positivos)	8 ^b	4.50	36.00
	<i>Ties</i> (Empates)	2 ^c		
	<i>Total</i> (total)	10		

- a Inventário de Depressão de Beck (Quarta) Inventário de Depressão de Beck (Domingo).
b Inventário de Depressão de Beck (Quarta) Inventário de Depressão de Beck (Domingo).
c Inventário de Depressão de Beck (Quarta) Inventário de Depressão de Beck (Domingo).
d Tipo de Droga = *Ecstasy*.

Test Statistics^{b,c} (Estatísticas Teste)

	<i>Beck Depression Inventory (Wednesday)</i> (Inventário de Depressão de Beck (Domingo)) – <i>Beck Depression Inventory (Sunday)</i> (Inventário de Depressão de Beck (Quarta-feira))
<i>Z</i> <i>Asymp. Sig. (2-tailed)</i> (Sig. Assint.) (Bilateral)	–2.527 ^a 0.012

- a Com base em postos negativos.
b Teste dos Postos com Sinal de Wilcoxon.
c Tipo de Droga = *Ecstasy*.

referindo (fornecendo assim o mesmo tipo de explicação que acabei de fornecer! – viu, eu não sou esperto, apenas li as notas da tabela!). Na Seção 13.3.1, expliquei que a estatística teste, *T*, é o menor valor dos dois tipos de postos, assim, o nosso valor de teste aqui é a soma dos postos negativos (isto é, zero). Também mostrei como esse valor pode ser convertido em um escore-*z* e isso é o que o SPSS faz. A

vantagem dessa abordagem é que ela permite que o valor exato da significância seja calculado com base na distribuição normal. A segunda Tabela da Saída 13.5 do SPSS mostra que a estatística teste está baseada nos postos negativos, que o *escore-z* é $-2,53$ (o valor calculado na Seção 13.3.1) e que esse valor é significativo a $p = 0,012$. Assim, em virtude desse valor ser baseado nos postos *negativos* (e porque a estatística teste é o menor valor entre os postos positivos e negativos, a maioria dos postos deve ser positiva), podemos concluir que tomar *Ecstasy* aumenta significativamente a depressão (medida pelo Inventário de Depressão de Beck (BDI)) da manhã seguinte ao meio da semana. Se a estatística teste tivesse sido baseada nos postos positivos, isso nos informaria que os resultados seriam opostos (isto é, os escores do BDI seriam maiores na manhã seguinte ao uso do que no meio da semana). Assim, podemos concluir que para os usuários de *Ecstasy* existe um aumento significativo da depressão a partir do dia seguinte para o meio da semana ($z = -2,53$, $p < 0,05$).

13.3.4 Saídas para o Grupo do Álcool ①

O restante da saída deve conter as mesmas duas tabelas anteriores, mas agora para o grupo do Álcool (se isso não aconteceu, você provavelmente esqueceu de dividir o arquivo). Como antes, a primeira tabela na Saída 13.6 do SPSS fornece informações sobre os postos dos escores. Ela nos informa sobre o número de postos negativos (aqueles em que as pessoas tiveram mais depressão no domingo do que na quarta-feira) e o número de postos positivos (pessoas que tiveram mais depressão na quarta-feira). A tabela mostra que para 9 dos 10 participantes, seus escores no domingo foram maiores do que na quarta-feira, indicando um grau de depressão maior na manhã seguinte comparado com o meio da semana. Diferentemente dos usuários do *Ecstasy*, não houve postos empatados nesse caso. A tabela mostra, ainda, o número médio de postos positivos e negativos, bem como a soma deles. Abaixo da tabela existem notas informando com o que os postos positivos e negativos

Saída 13.6 do SPSS

Ranks^d (Postos)

		<i>N</i>	Mean Rank (Média dos Postos)	Sum of Ranks (Soma dos Postos)
Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta-Feira)) – Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))	Negative Ranks (Postos Negativos)	9 ^a	5.22	47.00
	Positive Ranks (Postos Positivos)	1 ^b	8.00	8.00
	Ties (Empates)	0 ^c		
	Total (total)	10		

a Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta)) < Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))
b Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta)) > Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))
c a Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta)) = Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))
d Type of Drug = Alcohol (Tipo de Droga = Álcool)

Test Statistics^{b,c} (Estatísticas Teste)

	Beck Depression Inventory (Wednesday) (Inventário de Depressão de Beck (Quarta-Feira)) – Beck Depression Inventory (Sunday) (Inventário de Depressão de Beck (Domingo))
Z	-1.990 ^a
Asymp. Sig. (2-tailed) (Sig. Assint.) (Bilateral)	0.047

a Based on negative ranks (Com base em postos negativos)
b Wilcoxon Signed Rank Test (Teste dos Postos com Sinal de Wilcoxon)
c Type of Drug = Alcohol (Tipo de Droga = Álcool)

estão relacionados. Como antes, o menor valor dos dois grupos de postos é convertido no *escore-z* (nesse caso, o valor 8). A segunda tabela informa que a estatística teste tem como base os postos positivos e que o *escore-z* é $-1,99$ (esse é o valor calculado em 13.3.1 – viu, meus cálculos manuais funcionam!) e que

esse valor é significativo no nível $p < 0,047$. Portanto, concluímos (baseados no fato que postos positivos foram utilizados) que quando alguém toma Álcool, existe um declínio significativo do nível de depressão (mensurada pelo BDI) da manhã seguinte para o meio da semana ($z = -1,99$, $p < 0,05$).

A partir dos resultados dos dois grupos podemos ver que existe um efeito oposto quando o Álcool é ingerido do que quando o *Ecstasy* é tomado. O Álcool torna você levemente deprimido na manhã seguinte, mas essa depressão some no meio da semana. O *Ecstasy* também causa depressão na manhã seguinte ao consumo, mas essa depressão aumenta no meio da semana. É claro, para se ver o verdadeiro efeito na manhã seguinte teríamos que ter medido a depressão antes de a droga ser ingerida! Esse efeito oposto entre grupos de pessoas é conhecido como uma interação (isto é, você obtém um efeito sob certas circunstâncias e um efeito diferente sob outras circunstâncias) – vimos isso nos Capítulos 10, 11 e 12.

13.3.5 Calculando o tamanho de efeito ②

O tamanho de efeito pode ser calculado da mesma forma que para o teste de Mann-Whitney (veja a Equação na Seção 13.2.5). Nesse caso, a Saída 13.6 do SPSS nos informa que para o grupo do *Ecstasy*, z é $-2,53$, e que para o grupo do Álcool, z é $-1,99$. Nos dois casos temos 20 observações (mas utilizamos apenas 10 pessoas e as testamos duas vezes; o número de observações, e não o número de pessoas, é importante aqui). O tamanho de efeito é, desta forma;

$$r_{\text{Ecstasy}} = \frac{-2,53}{\sqrt{20}} = -0,57$$

$$r_{\text{Álcool}} = \frac{-1,99}{\sqrt{20}} = -0,44$$

Isso representa uma grande mudança nos níveis de depressão quando o *Ecstasy* é utilizado (o valor está acima do limite de 0,50 proposto por Cohen) e um efeito médio para

grande na depressão quando Álcool é ingerido (o valor obtido está entre os níveis de 0,3 e 0,5 para um efeito médio para grande proposto por Cohen).

13.3.6 Escrevendo e interpretando os resultados ①

Para o teste de Wilcoxon precisamos relatar apenas a estatística teste (que é representada pela letra T e é a menor das duas somas dos postos), a sua significância e de preferência o tamanho de efeito. Assim, o relatório pode ser assim:

- ✓ Para os usuários do *Ecstasy*, os níveis de depressão foram significativamente mais altos na quarta-feira (Mediana = 33,50) do que no domingo (Mediana = 17,50), $T = 0$, $p < 0,05$, $r = -0,57$. Contudo, para os usuários de Álcool, o oposto é verdadeiro: os níveis de depressão foram significativamente menores na quarta-feira (Mediana = 7,50) do que no domingo (Mediana = 16,0), $T = 8$, $p < 0,05$, $r = -0,44$.

Podemos também apresentar os valores de z :

- ✓ Para os usuários do *Ecstasy*, os níveis de depressão foram significativamente mais altos na quarta-feira (Mediana = 33,50) do que no domingo (Mediana = 17,50), $z = -2,53$, $p < 0,05$, $r = -0,57$. Contudo, para os usuários de Álcool o oposto é verdadeiro: os níveis de depressão foram significativamente menores na quarta-feira (Mediana = 7,50) do que no domingo (Mediana = 16,0), $z = -1,99$, $p < 0,05$, $r = -0,44$.

13.4 DIFERENÇAS ENTRE VÁRIOS GRUPOS INDEPENDENTES: O TESTE DE KRUSKAL-WALLIS ①

No Capítulo 8, descobrimos uma técnica denominada ANOVA independente de um fator que pode ser utilizada para testar diferenças entre vários grupos independentes. Mencionei

Dica da Samanta Ferrinho



- O teste dos postos com sinais de Wilcoxon **compara duas condições quando as mesmas pessoas participam em cada condição e os dados resultantes não são normalmente distribuídos ou violam a hipótese do teste t dependente.**
- Observe a linha denominada Sig. Assint. (bilateral) (*Asymp. Sig. (2-tailed)*): se o valor for menor do que 0,05, os dois grupos são significativamente diferentes.
- Observe os postos positivos e negativos (e as notas da tabela explicando o que eles significam) para saber como os grupos diferem (um grande número de postos em uma direção em particular nos informa qual a direção do resultado).
- Como já aconteceu com outros testes, o SPSS fornece apenas o valor da significância bilateral; se você quer a significância unilateral, basta dividir o valor da bilateral por 2.
- Relate a estatística T , o valor z correspondente e o valor da significância e o tamanho de efeito, se possível. Relate, ainda, as medianas e suas amplitudes correspondentes (ou trace um diagrama de caixa e bigodes).

várias vezes naquele capítulo que a ANOVA é uma técnica robusta à violação de suas hipóteses. Vimos também que podem ser tomadas algumas medidas quando existir heterogeneidade das variâncias (Quadro 8.3). Entretanto, existe alternativa: a ANOVA independente de um fator tem uma contraparte não-paramétrica chamada teste de **Kruskal-Wallis** (Kruskal e Wallis, 1952; Figura 13.5). Se você tem dados não normalmente distribuídos, ou violou alguma outra suposição, esse teste pode ser útil para contornar o problema.



Figura 13.5 Joseph Kruskal encontrando alguns erros na primeira edição de *Descobrir a estatística utilizando o SPSS*.

Li uma história recentemente em um jornal afirmando que cientistas tinham descoberto que a substância “*genistein*”, que ocorre naturalmente na soja, estava relacionada com uma contagem baixa de espermatozoides em machos ocidentais. Quando você lê a pesquisa real, sabe que na verdade ela foi realizada em ratos e não apresenta relação com uma baixa contagem de espermatozoides, mas existe uma evidência de um desenvolvimento sexual anormal nos ratos machos (provavelmente porque essa substância química age como estrogênio). O jornalista interpretou isso como uma relação com o declínio na contagem de espermatozoides em homens ocidentais (que jornalista!). De qualquer forma, como um vegetariano que come muitos produtos de soja e que provavelmente vai querer ter filhos um dia, quero testar essa ideia em humanos em vez de ratos. Peguei 80 homens e os dividi em quatro grupos que variavam no número de refeições com soja que eles comiam por semana em um período de um ano. O primeiro grupo era o controle e eles não tinham soja nas refeições (nada no ano inteiro); o segundo grupo tinha uma refeição com soja por semana (cerca de 52 refeições no ano); o terceiro grupo tinha quatro refeições por semana (aproximadamente 208 no ano); e o último grupo tinha sete refeições com soja por semana (ou 364 durante o ano). No final

do ano, todos os participantes foram testados (isto é, foram enviados a um laboratório tão longe quanto possível de mim).³

13.4.1 Teoria do teste de Kruskal-Wallis ②



A teoria do teste de Kruskal-Wallis é parecida com a do teste de Mann-Whitney (e da soma dos postos de Wilcoxon), assim, antes de continuar a leitura dê uma olhada na Seção 13.2.1. Da mesma forma que o teste de Mann-Whitney, o teste de Kruskal-Wallis tem como

base os postos dos valores. Para começar, simplesmente ordene os escores do menor para o maior, ignorando o grupo ao qual o escore pertence, e atribua ao menor o posto 1, ao próximo o posto 2 e assim por diante (veja a Seção 13.2.1 para mais detalhes). Quando você tiver atribuído postos a todos os dados coletados, retorne com os escores aos seus grupos e simplesmente adicione os postos de cada grupo. A soma dos postos de cada grupo é representada por R_i (onde i é utilizado para representar um grupo em particular). A Tabela 13.3 mostra os dados brutos para esse exemplo mais os postos (tente determinar os postos dos dados e veja se você obtém os mesmos resultados que eu!).

Depois que a soma dos postos foi calculada para cada grupo, a estatística teste H é calculada utilizando uma equação como a (13.1):

³ No caso de qualquer médico estar lendo isso, esses dados foram inventados e eu não tenho absolutamente ideia alguma do que seja uma verdadeira contagem de espermatozoides; assim os valores provavelmente são ridículos, peço desculpas e você pode rir da minha ignorância!

Tabela 13.3 Dados para o exemplo da soja com postos

Sem soja		Com uma refeição		Com 4 refeições		Com 7 refeições	
Espermatozoides (Milhões)	Posto	Espermatozoides (Milhões)	Posto	Espermatozoides (Milhões)	Posto	Espermatozoides (Milhões)	Posto
0,35	4	0,33	3	0,40	6	0,31	1
0,58	9	0,36	5	0,60	10	0,32	2
0,88	17	0,63	11	0,96	19	0,56	7
0,92	18	0,64	12	1,20	21	0,57	8
1,22	22	0,77	14	1,31	24	0,71	13
1,51	30	1,53	32	1,35	27	0,81	15
1,52	31	1,62	34	1,68	35	0,87	16
1,57	33	1,71	36	1,83	37	1,18	20
2,43	41	1,94	38	2,10	40	1,25	23
2,79	46	2,48	42	2,93	48	1,33	25
3,40	55	2,71	44	2,96	49	1,34	26
4,52	59	4,12	57	3,00	50	1,49	28
4,72	60	5,65	61	3,09	52	1,50	29
6,90	65	6,76	64	3,36	54	2,09	39
7,58	68	7,08	66	4,34	58	2,70	43
7,78	69	7,26	67	5,81	62	2,75	45
9,62	72	7,92	70	5,94	63	2,83	47
10,05	73	8,04	71	10,16	74	3,07	51
10,32	75	12,10	77	10,98	76	3,28	53
21,08	80	18,47	79	18,21	78	4,11	56
Total (R_i)	927		883		883		547

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) \quad (13.1)$$



Nessa equação, R_i é a soma dos postos para cada grupo, N é o total da amostra (nesse caso, 80) e n_i é o tamanho amostral de cada um dos grupos (nesse caso, temos todos os valores iguais a 20). Eleve a soma dos postos de cada grupo ao quadrado e divida esse valor pelo tamanho amostral de cada grupo. Depois, adicione esses valores. Isso resolve a parte central da equação; o resto dela envolve o cálculo de vários valores baseados no tamanho total da amostra. Para esses dados, temos:

$$\begin{aligned} H &= \frac{12}{80(81)} \left(\frac{927^2}{20} + \frac{883^2}{20} + \frac{883^2}{20} + \frac{547^2}{20} \right) \\ &\quad - 3(81) \\ &= \frac{12}{6480} (42966,45 + 38984,45 + 38984,45 \\ &\quad + 14960,45) - 243 \\ &= 0,0019 (135895,8) - 243 \\ &= 251,66 - 243 \\ &= 8,659 \end{aligned}$$

Essa estatística teste tem um tipo especial de distribuição conhecida como qui-quadrado (veja o Capítulo 16), e para essa distribuição existe um parâmetro denominado grau de liberdade, que é o número de grupos menos um ($k - 1$); nesse caso, 3.

13.4.2 Atribuindo dados e fazendo análise condicional ①

Quando os dados são coletados com diferentes participantes em cada grupo, precisamos inseri-los utilizando uma variável de código ou codificadora. Assim, o editor de dados terá duas colunas de dados. A primeira será a da variável de código (denominada aqui de **Soya**) que, nesse caso, terá quatro valores (por conveniência, sugiro 1 = sem soja, 2 = uma refeição com soja por semana, 3 = quatro refeições com soja se-

manais e 4 = sete refeições com soja semanais). A segunda coluna terá os valores da variável dependente (contagem de espermatozoides) mensurada no final do ano (chame essa variável de **sperm**). Quando você entrar com os dados no SPSS lembre de informar ao computador os códigos dos grupos que estão sendo utilizados (veja a Seção 2.4.4). Esses dados podem ser encontrados no arquivo **Soya.sav**.

Primeiro, rodamos algumas análises exploratórias nos dados; porque queremos verificar diferenças precisamos realizar essas análises exploratórias para cada grupo (Seções 3.5 e 3.6). Se você realizar essas análises deverá, então, encontrar algumas tabelas semelhantes às mostradas na Saída 13.7 do SPSS. A primeira tabela mostra que o teste K-S (veja a Seção 3.5) não foi significativo para o grupo-controle ($D(20) = 0,181$, $p > 0,05$), mas o teste de Shapiro-Wilk é significativo, e esse teste é, de fato, mais preciso (embora ele seja menos relatado) que o teste K-S (veja o Capítulo 3). Os dados do grupo que tiveram uma refeição de soja por semana foram significativamente diferentes da normal ($D(20) = 0,207$, $p < 0,05$), assim como os para quem teve quatro refeições semanais com soja ($D(20) = 0,267$, $p < 0,01$) e sete ($D(20) = 0,204$, $p < 0,05$). A segunda tabela mostra os resultados do teste de Levene. A hipótese de homogeneidade da variância foi violada, $F(3, 76) = 5,12$, $p < 0,01$ e isso é mostrado pelo fato de a significância do teste de Levene ser menor do que 0,05. Desse modo, esses dados violam duas hipóteses importantes: eles não são normalmente distribuídos e os grupos apresentam variâncias heterogêneas!

13.4.3 Executando o teste de Kruskal-Wallis no SPSS ①

Primeiro, acesse a caixa de diálogo principal utilizando o menu **Analyze**⇒**Nonparametric Tests**⇒**K Independent Samples...** (Analisar⇒Testes não-paramétricos⇒K Amostras Independentes...) (veja a Figura 13.6). Com essa caixa de diálogos ativada, selecione a variável dependente da lista (clique em **Sperm Count**

Saída 13.7 do SPSS

Tests of Normality (Testes de Normalidade)

Number of Soya Meals Per Week (Número de refeições com soja por semana)		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic (Estatística)	df (gl)	Sig. (Sig.)	Statistic (Estatística)	df (gl)	Sig. (Sig.)
Sperm Count (Millions) (Contagem de espermatozoides (Milhões))	No Soya Meals (Sem refeições com soja)	0.181	20	0.085	0.805	20	0.001
	1 Soya Meal per Week (1 refeição com soja por semana)	0.207	20	0.024	0.826	20	0.002
	4 Soya Meal per Week (4 refeições com soja por semana)	0.267	20	0.001	0.743	20	0.000
	7 Soya Meal per Week (7 refeições com soja por semana)	0.204	20	0.028	0.912	20	0.071

a Correção da significância de Lilliefors.

Test of Homogeneity of Variance (Teste de Homogeneidade das Variâncias)

		Levene Statistic (Estatística de Levene)	df1 (gl1)	df2 (gl2)	Sig. (Sig.)
Sperm Count (Millions) (Contagem de espermatozoides (Milhões))	Based on Mean (Com base na Média)	5.117	3	76	0.003
	Based on Median (Com base na mediana)	2.860	3	76	0.042
	Based on Median and with adjusted df (Com base na mediana e com gl ajustado)	2.860	3	58.107	0.045
	Based on on trimmed mean (Com base na média interna)	4.070	3	76	0.010

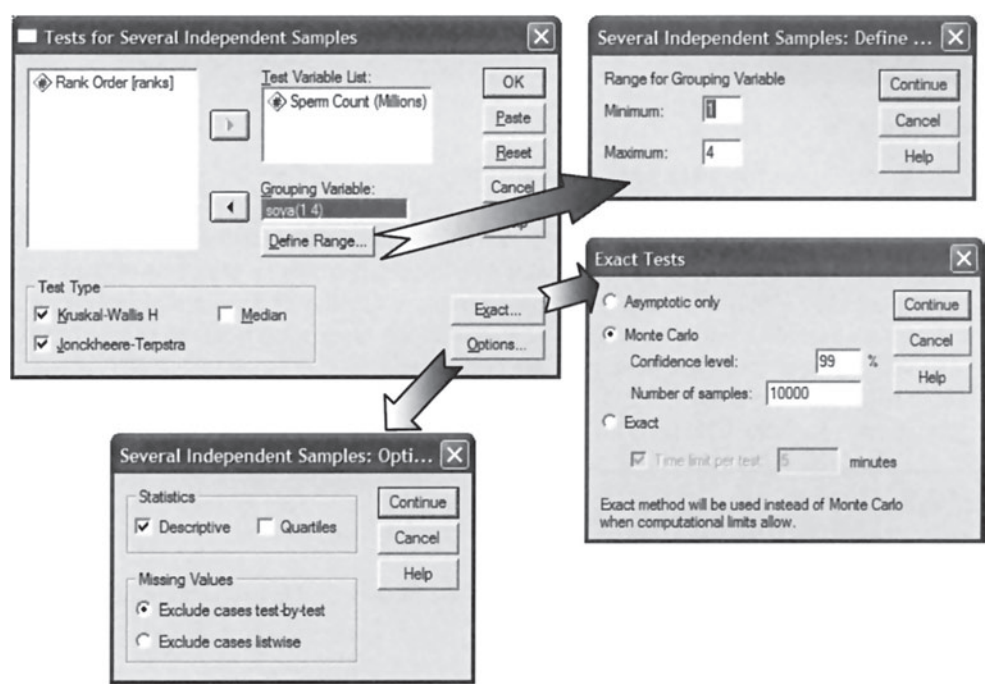







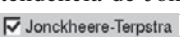


Figura 13.6 Caixas de diálogo para o teste de Kruskal-Wallis.

(**Milions**) (Contagem de Espermatozoides (Milhões)) e transfira-a para o quadro **Test Variable List** (Lista de Variáveis de Teste) clicando em . A seguir, selecione a variável independente (a variável de grupo), nesse caso, **Soya**, e transfira-a para o quadro **Grouping Variable** (Variável de Agrupamento). Quando a variável de agrupamento tiver sido selecionada, o botão  (Definir Amplitude) se torna ativo e você deve clicá-lo para ativar a caixa de diálogo **Define Range** (definir intervalo). O SPSS precisa saber a faixa de valores que você atribuiu aos grupos e existe um espaço para digitar os valores mínimo e máximo. Se você seguiu os códigos sugeridos, o valor mínimo deve ser 1 e o máximo 4, assim, digite esses números nos espaços apropriados. Quando tiver definido os grupos, clique em  (Continuar) para retornar à caixa de diálogo principal. A caixa de diálogo principal também fornece  (Opções) para realizar outros testes semelhantes ao de Kruskal-Wallis (veja o Quadro 13.4).

Se você clicar em  (Exato...), irá obter uma caixa de diálogo para selecionar o valor da significância exato para o teste de Kruskal-Wallis. Expliquei essa opção na Seção 13.2.3, portanto, não vou me repetir aqui. Vou apenas recapitular afirmando que quando os tamanhos amostrais são grandes, provavelmente você deve optar pelo método de Monte Carlo (como foi feito nesse exemplo), e quando tivermos pequenos tamanhos amostrais, a

opção deve ser pelo teste exato. Finalmente, se você clicar em  (Opções), uma caixa de diálogo fornece opções adicionais de análise. Essas opções não são muito úteis; por exemplo, a opção que fornece as estatísticas descritivas faz isso para todo o conjunto de dados (não dividindo os valores de acordo com o grupo a que eles pertencem). Por essa razão, recomendo obter as estatísticas descritivas utilizando o método que aprendemos nas Seções 3.4 e 3.5. Para rodar a análise, retorne à caixa de diálogo principal e clique em . A opção final é realizar o teste de tendência de Jonckheere-Terpstra (selecione ). Isso será útil se quisermos verificar se a média dos grupos apresenta uma tendência linear (veja a Seção 8.2.10.5).

13.4.4 Saídas para o teste de Kruskal-Wallis ①

A Saída 13.8 do SPSS mostra um resumo dos postos dos dados para cada condição, e precisamos disso para interpretar qualquer efeito.

A Saída 13.9 do SPSS mostra a estatística teste H para o teste de Kruskal-Wallis (embora o SPSS a chame de qui-quadrado, em virtude da sua distribuição, em vez de H), seu grau de liberdade associado, (nesse caso, temos quatro grupos, assim, o grau de liberdade é $4 - 1$ ou 3) e a significância. É importante observar que o valor da significância é 0,034; como esse

Quadro 13.4

Outras opções ②

Na caixa de diálogo principal existem outras opções que podem ser selecionadas:

- **Median** (Mediana): Testa se as amostras são retiradas de populações com a mesma mediana. Assim, faz o mesmo teste que o de Kruskal-Wallis. Ela funciona produzindo uma tabela de contingência dividida para cada grupo no número de escores que se encontram acima e abaixo da mediana observada do conjunto total de dados. Se os grupos forem da mesma população, devemos esperar que essas frequências sejam as mesmas em todas as condições (aproximadamente 50% acima e 50% abaixo).
- **Jonckheere-Terpstra**: Testa se as médias dos grupos seguem uma tendência linear (veja a Seção 13.4.6).

Saída 13.8 do SPSS 13.8

Ranks (Postos)

Number of Soya Meals (Número de refeições com soja)		N	Mean Rank (Média dos Postos)
Sperm Count (Millions) (Contagem de espermatozoides (Milhões))	No Soya Meals (Sem refeições com soja)	20	46.35
	1 Soya Meal per Week (1 refeição com soja por semana)	20	44.15
	4 Soya Meal per Week (4 refeições com soja por semana)	20	44.15
	7 Soya Meal per Week (7 refeições com soja por semana)	20	27.35
	Total (Total)	80	

Saída 13.9 do SPSS

Test Statistics^{a,b,c} (Estatísticas Teste)

			<i>Sperm Count</i> (millions) (Contagem de Espermatozoides (Milhões))
<i>Chi-Square</i> (Qui-Qua- drado)			8.659
<i>df</i> (gl)			3
<i>Asymp.</i> <i>Sig.</i> (Sig. Assint.)			0.034
<i>Monte Car-</i> <i>lo</i> (Monte Carlo)	<i>Sig. (Sig.)</i>		0.033 ^a
	<i>99% Con-</i> <i>fidence</i> <i>Interval</i> (Intervalo Inferior)	<i>Lower</i> <i>Bound</i> (Limite Inferior)	0.028
	<i>Sig. (Sig.)</i> de 99% de Confiança.)	<i>Upper</i> <i>Bound</i> (Limite Superior)	0.037

a Com base em 10000 tabelas amostradas com semente de partida 846668601.

b Teste Kruskal-Wallis.

c Variável de Agrupamento: Número de refeições com soja por semana.

levemente menor (0,033). Esse é o valor a ser verificado em vez do valor assintótico se eles apresentarem resultados diferentes. O intervalo de confiança para a significância também é útil: ele é 0,028 – 0,037 e o fato de que ele não contém o zero é importante porque significa que o intervalo pode conter o valor da estatística teste com 99% de probabilidade. Isso nos dá muita confiança de que o efeito é genuíno. Como a ANOVA de um fator, da mesma forma, esse teste nos diz apenas que a diferença existe; ele não informa onde a diferença está.

Uma forma de ver qual grupo difere é olhar para o diagrama de caixa e bigodes (veja a Seção 3.3.2) dos grupos (veja a Figura 13.7). Note que existem alguns valores atípicos (observe os círculos e asteriscos que estão acima do bigode superior) – esses são homens que produziram uma quantidade atípica de espermatozoides. Utilizando o controle como base, as medianas dos primeiros três grupos parecem semelhantes, contudo, a mediana do grupo que fez sete refeições semanais com soja parece um pouco abaixo, assim, talvez a diferença esteja aí. Entretanto, essas conclusões são subjetivas. O que realmente precisamos são alguns contrastes ou testes *post hoc* como os utilizados na ANOVA (veja as Seções 8.2.10 e 8.2.11).

13.4.5 Análise *post hoc* para o teste de Kruskal-Wallis ②

Existem duas maneiras de realizar procedimentos *post hoc* não-paramétricos, a primeira com a utilização de testes de Mann-Whitney (Seção 13.2). Contudo, se utilizarmos vários testes de Mann-Whitney, iremos inflacionar a taxa de erro do Tipo I (Seção 8.2.1) e é por esse motivo que não começamos fazendo muitos testes de Mann-Whitney! Contudo, podemos realizar vários testes de Mann-Whitney em seguida a um teste de Kruskal-Wallis



valor é menor do que 0,05, podemos concluir que a quantidade de refeições de soja ingeridas por semana afeta significativamente a contagem de espermatozoides. Note também que a estimativa por Monte Carlo da significância é

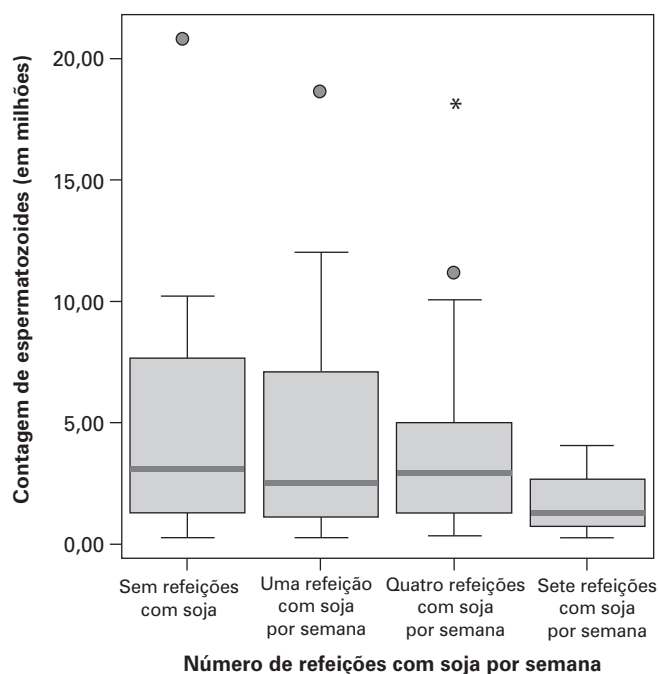


Figura 13.7 Diagrama de caixa e bigodes para a contagem de espermatozoides de pessoas ingerindo um número diferente de refeições com soja por semana.

desde que seja feito algum tipo de ajuste para assegurar que o Erro do Tipo I não ultrapasse o valor de 0,05. O método mais fácil é utilizar a correção de Bonferroni, que na sua forma mais simples significa que em vez de utilizarmos um valor crítico de 0,05 para cada teste, devemos utilizar o valor 0,05 dividido pelo número de testes que iremos realizar. Se você fizer isso, logo irá descobrir que está utilizando um valor para a significância tão pequeno que poderá causar problemas. Dessa forma, é recomendável ser bem seletivo nas comparações realizadas. Nesse exemplo, temos um grupo-controle que não tem refeições com soja; um bom e sucinto conjunto de comparações será comparar cada grupo contra esse grupo-controle:

- ▶ Teste 1: uma refeição com soja por semana comparado com refeições sem soja;
- ▶ Teste 2: quatro refeições com soja por semana comparado com refeições sem soja;

- ▶ Teste 3: sete refeições com soja por semana comparado com refeições sem soja.

Isso resulta em apenas três testes, assim, em vez de utilizarmos 0,05 como nosso valor crítico, devemos utilizar $0,05/3 = 0,0167$. Se não utilizarmos testes apropriados e fizermos todas as comparações, teremos um total de seis testes em vez de apenas três (sem soja vs. 1 refeição, sem soja vs. 4 refeições, sem soja vs. 7 refeições, 1 refeição vs. 4 refeições, 1 refeição vs. 7 refeições e 4 refeições vs. 7 refeições) e o nosso valor crítico seria de $0,05/6 = 0,0083$ em vez de 0,0167.

A Saída 13.10 do SPSS mostra as estatísticas teste da realização dos testes de Mann-Whitney nas três comparações que foram sugeridas. A melhor forma de fazer isso no SPSS é alterar os valores dos códigos para os Grupos 1 e 2 que foram digitados na caixa de diálogo **Define Groups** (Definir Grupos) (ver Figura 13.3). Lembre que agora estamos utilizando

do um valor crítico de 0,0167, assim, a única comparação significativa é a daqueles que tiveram sete refeições com soja por semana com aqueles que não tiveram refeições com soja na semana (porque o valor da significância observada foi de 0,009, que é menor do que 0,0167). As outras duas comparações produzem valores de significância que são maiores do que 0,0167 assim, devemos afirmar que elas não são significativas. Desse modo, o efeito que obtivemos parece refletir principalmente o fato de que comer sete refeições com soja por semana (sei disso pelas medianas vistas na Figura 13.7) reflete em uma contagem menor de espermatozoides quando comparado a quem não come

soja. Contudo, comer um pouco de soja (de uma a quatro refeições semanais) parece não afetar a contagem de espermatozoides.



A segunda forma de realizar testes *post hoc* é essencialmente a mesma de realizar testes de Mann-Whitney para todas as possíveis comparações, mas mesmo assim irei guiá-lo no processo! Ele é descrito por Siegel e Castellan (1988) e envolve tomar a diferença entre as médias dos postos dos diferentes grupos comparada com um valor *z* (corrigido para o número de comparações sendo feitas) e uma constante baseada no tamanho amostral total e as dos tama-

Saída 13.10 do SPSS

Sem soja *versus* uma refeição com soja por semana:

Test Statistics^b (Estatísticas Teste)

	Sperm Count (millions) (Contagem de Espermatozoides (em Milhões))
Mann-Whitney U (U de Mann-Whitney)	191.000
Wilcoxon W (W de Wilcoxon)	401.000
Z	−0.243
Asymp. Sig. (2-tailed) (Sig. Assint.) (Bilateral)	0.808
Exact Sig. [2*(1-tailed Sig.) (Sig. Exata) [2*(Sig. Unilateral)	0.820 ^a

a Não corrigido para empates.

b Variável de Agrupamento: Número de refeições com soja por semana.

Sem soja *versus* quatro refeições com soja por semana:

Test Statistics^b (Estatísticas Teste)

	Sperm Count (millions) (Contagem de Espermatozoides (em Milhões))
Mann-Whitney U (U de Mann-Whitney)	188.000
Wilcoxon W (W de Wilcoxon)	398.000
Z	−0.325
Asymp. Sig. (2-tailed) (Sig. Assint.) (Bilateral)	0.745
Exact Sig. [2*(1-tailed Sig.) (Sig. Exata) [2*(Sig. Unilateral)	0.758 ^a

a Não corrigido para empates.

b Variável de Agrupamento: Número de refeições com soja por semana.

Sem soja *versus* sete refeições com soja por semana:

Test Statistics^b (Estatísticas Teste)

	Sperm Count (millions) (Contagem de Espermatozoides (em Milhões))
Mann-Whitney U (U de Mann-Whitney)	104.000
Wilcoxon W (W de Wilcoxon)	314.00
Z	−2.597
Asymp. Sig. (2-tailed) (Sig. Assint.) (Bilateral)	0.009
Exact Sig. [2*(1-tailed Sig.) (Sig. Exata) [2*(Sig. Unilateral)	009 ^a

a Não corrigido para empates.

b Variável de Agrupamento: Número de refeições com soja por semana.

nhos amostrais dos dois grupos sendo comparados. Essa desigualdade é:

$$|\bar{R}_u - \bar{R}_v| \geq z_{\alpha/k(k-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_u} + \frac{1}{n_v} \right)} \quad (13.2)$$

O lado esquerdo dessa equação é apenas a diferença entre a média dos postos dos dois grupos sendo comparados, mas ignorando o sinal dessa diferença (as duas linhas verticais que englobam a diferença entre a média dos postos significam que se a diferença é negativa, nós devemos ignorar o sinal negativo e tratar a diferença como se ela fosse positiva). Para o resto da equação, k é o número de grupos (no exemplo da soja ele vale 4), N é o tamanho total da amostra (nesse caso, 80), n_u é o número de pessoas no primeiro grupo sendo comparado (como temos grupos de mesmo tamanho nesse exemplo, independentemente do grupo esse valor será igual a 20) e n_v é o tamanho do segundo grupo sendo comparado (novamente, esse valor será 20 independentemente do grupo sendo comparado, porque, nesse exemplo, todos os grupos apresentam o mesmo tamanho). A última coisa que precisamos saber é $z_{\alpha/k(k-1)}$, e para determinar esse valor precisamos decidir o nível para α , que é o nível de significância que queremos trabalhar. Nas ciências sociais tradicionalmente trabalhamos com um nível de 0,05 de significância, assim, α será 0,05. Calculamos, então $k(k-1)$, que para esses dados será $4(4-1) = 12$. Dessa forma, $\alpha/k(k-1) = 0,05/12 = 0,00417$. Assim, $z_{\alpha/k(k-1)}$ apenas significa “o valor da curva normal para o qual somente $\alpha/k(k-1)$ dos demais valores z são maiores” (ou, nesse caso, “o valor z (curva normal) para o qual somente 0,00417% dos valores da curva normal estão acima”). Em termos práticos, isso significa que devemos procurar na tabela no Apêndice A.1 a coluna denominada *Parte Menor* e encontrar o número 0,00417 (ou o valor mais próximo a ele, que, nesse caso, é 0,00415), e depois olhar na mesma linha na coluna z . Nesse caso, você deve encontrar o valor de z igual a 2,64. O próximo passo é calcular o lado direito da equação (13.2):

$$\begin{aligned} \text{Diferença Crítica} &= z_{\alpha/k(k-1)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_u} + \frac{1}{n_v} \right)} \\ &= 2,64 \sqrt{\frac{80(80+1)}{12} \left(\frac{1}{20} + \frac{1}{20} \right)} \\ &= 2,64 \sqrt{540(0,1)} = 2,64 \sqrt{54} \\ &= 19,40 \end{aligned} \quad (13.2)$$

Para esse exemplo, em virtude dos tamanhos amostrais dos grupos serem os mesmos, essa diferença crítica pode ser utilizada em todas as comparações. Contudo, quando os tamanhos dos grupos forem diferentes, a diferença crítica deverá ser calculada para cada comparação individualmente. A última etapa é calcular todas as diferenças entre as médias dos postos de todos os grupos (esses valores podem ser encontrados na Saída 13.8 do SPSS) (Tabela 13.4).



A equação (13.2) significa basicamente que se as diferenças entre as médias dos postos são maiores ou iguais ao valor

Tabela 13.4 Diferenças entre postos médios para os dados sobre soja

Comparação	\bar{R}_u	\bar{R}_v	$\bar{R}_u - \bar{R}_v$	$ \bar{R}_u - \bar{R}_v $
Sem soja – 1 refeição com soja	46,35	44,15	2,20	2,20
Sem soja – 4 refeições com soja	46,35	44,15	2,20	2,20
Sem soja – 7 refeições com soja	46,35	27,35	19,00	19,00
1 refeição com soja – 4 refeições com soja	44,15	44,15	0,00	0,00
1 refeição com soja – 7 refeições com soja	44,15	27,35	16,80	16,80
4 refeições com soja – 7 refeições com soja	44,15	27,35	16,80	16,80

da diferença crítica para aquela determinada comparação, então essa diferença é significativa. Nesse caso, como temos apenas uma diferença crítica, qualquer diferença que for maior do que 19,40 será significativa. Como você pode ver, todas as diferenças se encontram abaixo desse valor e, assim, concluímos que nenhum dos grupos foi significativamente diferente! Isso contradiz nosso achado anterior em que o teste de Mann-Whitney para o grupo sem refeições com soja comparado ao grupo de sete refeições com soja por semana foi significativo. Por que você acha que isso aconteceu? Para o nosso teste de Mann-Whitney, fizemos apenas três comparações e, assim, corrigimos o valor da significância para os três testes ($0,05/3 = 0,0167$). Antes nesta seção, afirmei que se comparássemos todos os grupos entre si, existiriam seis comparações, e que, nesse caso, uma diferença seria significativa somente se o valor da significância fosse menor do que $0,05/6 = 0,0083$. Se voltarmos ao nosso teste de Mann-Whitney significativo (Saída 13.10 do SPSS), o valor da significância foi de 0,009; portanto, se tivéssemos feito todas as seis comparações, elas não seriam significativas (porque 0,009 é maior do que 0,0083)! Isso ilustra o que eu disse antes sobre as vantagens de selecionar as comparações a serem feitas.

13.4.6 Testes para a tendência: o teste de Jonckheere-Terpstra ②

Voltando à Seção 13.4.3, selecionamos uma opção para o teste de Jonckheere-Terpstra, ☒ **Jonckheere-Terpstra** (Jonckheere, 1954, Terpstra, 1952). Essa estatística testa se as medianas dos grupos que estamos comparando seguem algum padrão de ordenação. Essencialmente, ela faz o mesmo que o teste de Kruskal-Wallis (isto é, testa a diferença das medianas dos grupos), mas incorpora informações sobre se a ordem dos grupos tem alguma tendência. Você deve utilizar essa estatística quando espera que os grupos sendo comparados tenham algum padrão específico com as medianas. Assim, no exemplo atual, esperamos que quanto mais soja a pessoa coma, mais a contagem

de espermatozoides decresça. Dessa forma, o grupo-controle deve ter a contagem mais alta, os com uma refeição por semana mais baixa, e os demais ainda menos. Assim, devemos perceber uma ordem nas nossas medianas: elas devem decrescer ao longo dos grupos. De forma contrária, devem existir situações onde podemos esperar que as medianas aumentem. Por exemplo, existe um fenômeno na psicologia conhecido como “efeito da mera exposição”, que significa basicamente que quanto mais você for exposto a algo, mais você gostará dele. Gravadoras utilizam esse efeito para assegurar que as músicas sejam tocadas nas rádios por pelo menos dois meses antes dos seus lançamentos, assim, no dia no lançamento, muitos gostarão da música e a levarão para os primeiros lugares nas paradas.⁴ De qualquer modo, se você pegar três grupos e expô-los a ouvir uma música 10, 20 e 30 vezes, respectivamente, e então medir quanto cada pessoa gostou da música, você espera que a mediana aumente. Aqueles que a ouviram 10 vezes vão gostar um pouco menos daqueles que a ouviram 20 vezes e esses, por sua vez, um pouco menos daqueles que a ouviram 30 vezes.

O teste de Jonckheere-Terpstra (na maioria das vezes chamado apenas de teste de Jonckheere) foi projetado para essas situações. No SPSS, ele funciona com base no princípio de a variável de código (a que define os grupos) especificar a ordem na qual se espera que as medianas mudem (não importa se esperamos que elas aumentem ou diminuam). Assim, no nosso exemplo da soja, os grupos foram codificados como 1 = sem soja, 2 = uma refeição com soja por semana, 3 = quatro refeições com soja por semana e 4 = sete refeições com soja por semana, assim, o teste irá verificar se as medianas das contagens de espermatozoides aumentam ou diminuem ao longo dos grupos quando eles são ordenados dessa forma. Obviamente, podemos alterar o esquema de códigos e testar se as medianas estão ordenadas de uma forma diferente. O funcionamento exato des-

⁴ Embora em muitos casos a mera exposição parece ter o efeito contrário para mim: quanto mais eu ouço certas músicas, mais quero que meu cérebro se livre delas.

se teste está descrito no *site* www.artmed.com.br no arquivo **Jonckheere.pdf**, mas você não precisa dele para utilizar o teste. O mais importante é testar se as medianas dos grupos estão em ordem crescente ou decrescente na *ordem especificada pela variável codificadora*.

Vimos como especificar o teste na Seção 13.4.3 e a Saída 13.11 do SPSS mostra o resultado do teste para os dados da soja. Essa tabela mostra o número de grupos sendo comparados, 4 (caso você ainda não tenha notado). Ela apresenta, também, o valor da estatística teste, J, que é 912. Em amostras grandes (mais do que oito elementos por grupo), essa estatística teste tem uma distribuição normal com a média e o desvio padrão facilmente definidos e calculados (veja o arquivo **Jonckheere.pdf** no *site* e você irá encontrar que como a saída informa a média é 1200 e o desvio padrão é 116,33). Sabendo esses valores, podemos converter a estatística teste para o escore-z, que nesse caso vale: $z = (912 - 1200)/116,33 = -2,476$. Isso é semelhante ao que fizemos

com os testes de Mann-Whitney e Wilcoxon. Esse escore-z pode então ser comparado com os valores da distribuição normal, e porque o teste de Jonckheere deve ser sempre unilateral (especificamos antes do experimento a ordem das medianas), estaremos procurando por um valor acima de 1,65 (quando ignoramos o sinal). O valor 2,47 é, dessa forma, significativo. O sinal do valor-z nos informa a ordem das medianas. Se ele for positivo, elas estão em uma tendência crescente, mas se ele é negativo, como aqui, as medianas estão em ordem decrescente (as medianas diminuem à medida que a variável de código aumenta). Nesse exemplo, as variáveis foram codificadas como 1 = sem soja, 2 = uma refeição com soja por semana, 3 = quatro refeições com soja por semana e 4 = sete refeições com soja por semana; isso significa que as medianas ficam menores à medida que avançamos de zero refeições com soja a sete refeições com soja.

Também podemos notar que existem valores significativos unilaterais e bilaterais es-

Saída 13.11 do SPSS

Jonckheere-Terpstra Test^b (Teste de Jonckheere-Terpstra)

			Sperm Count (Millions) (Contagem de Espermatozoides em Milhões)
Number of Levels in Number of Soya Meals Per Week (Número de níveis em número de refeições com soja por semana)			4
N			80
Observed J-T Statistic (Estatística J-T Observada)			912.000
Mean J-T Statistic (Média da Estatística J-T)			1200.000
Std. Deviation of J-T Statistic (Desvio Padrão da Estatística J-T)			116.333
Std J-T Statistic (Estatística J-T Padronizada)			-2.476
Asymp. Sig. (2-tailed) (Sig. Assint. (Bilateral))			0.013
Monte Carlo Sig. (2-tailed) (Sig. Monte Carlo (Bilateral))	Sig. (Sig.)		0.013 ^a
	99% Confidence Interval (Intervalo de 99% de Confiança.)	Lower Bound (Limite Inferior)	0.010
		Upper Bound (Limite Superior)	0.016
Monte Carlo Sig. (1-tailed) (Sig. Monte Carlo (Unilateral))	Sig. (Sig.)		0.006 ^a
	99% Confidence Interval (Intervalo de 99% de Confiança.)	Lower Bound (Limite Inferior)	0.004
		Upper Bound (Limite Superior)	0.008

a Com base em 10000 tabelas amostradas com semente de partida 846668601.
b Variável de Agrupamento: Número de refeições com soja por semana.

timados com o uso do método Monte Carlo. Eles apareceram porque escolhemos essa opção para o teste de Kruskal-Wallis. Eles confirmam o que já tínhamos encontrado.

13.4.7 Calculando o tamanho de efeito ②

Infelizmente, não existe uma forma simples de converter uma estatística qui-quadrado que tem mais do que um grau de liberdade em um tamanho de efeito r . Você pode utilizar o valor da significância da estatística teste de Kruskal-Wallis para encontrar um valor associado de z a partir de uma tabela dos valores da distribuição normal (como a encontrada no Apêndice A.1). A partir daí você pode utilizar a conversão para r utilizada na Seção 13.2.5. Contudo, esse tipo de tamanho de efeito é raramente útil (porque ele está representando um efeito geral). Em muitos casos, é mais interessante saber o tamanho de efeito para uma comparação específica (como a comparação de dois grupos). Por esse motivo, sugiro calcular o tamanho de efeito para os testes de Mann-Whitney que utilizamos para acompanhar a análise principal.

Para a primeira comparação (sem soja *versus* uma refeição com soja), a Saída 13.10 do SPSS mostra que z é $-0,243$, e porque ele teve por base a comparação de dois grupos, cada um contendo 20 observações, temos 40 observações no total. O tamanho de efeito será, portanto:

$$r_{0 \text{ Soja vs. 1 Soja}} = \frac{-0,243}{\sqrt{40}} \\ = -0,04$$

Isso representa um efeito muito pequeno: o valor está próximo de zero, o que informa que o efeito na contagem de espermatozoides de uma refeição com soja por semana é desprezível.

Para a segunda comparação (refeição sem soja *versus* quatro refeições com soja) a Saída 13.10 do SPSS mostra que z é $-0,325$, novamente tendo por base 40 observações. O tamanho de efeito será:

$$r_{0 \text{ Soja vs. 4 Soja}} = \frac{-0,325}{\sqrt{40}} \\ = -0,05$$

Mais uma vez temos um efeito desprezível, pois o valor está próximo de zero, significando que o efeito na contagem de espermatozoides de quatro refeições com soja por semana é quase nulo.

Para a comparação final (refeição sem soja *versus* sete refeições com soja) a Saída 13.10 do SPSS mostra que o valor de z é $-2,597$ com base em 40 observações. Assim, o tamanho de efeito é:

$$r_{0 \text{ Soja vs. 7 Soja}} = \frac{-2,597}{\sqrt{40}} \\ = -0,41$$

Isso representa um efeito médio, o que nos informa que sete refeições com soja por semana baixa a quantidade de espermatozoides (quando comparado a nenhuma refeição com soja por semana), o que é um achado significativo.

Podemos calcular ainda o tamanho de efeito para o teste de Jonckheere se quisermos utilizar a mesma equação. Isso envolve todos os dados, então teremos 80 como valor de N :

$$r_{\text{Jonckheere}} = \frac{-2,476}{\sqrt{80}} \\ = -0,28$$

13.4.8 Escrevendo e interpretando os resultados ①

Para o teste de Kruskal-Wallis, precisamos relatar apenas a estatística teste (que, como vimos antes, é representada por H), seus graus de liberdade e sua significância. Assim, podemos escrever que:

- ✓ A contagem de espermatozoides foi significativamente afetada pela ingestão de refeições com soja ($H(3) = 8,66$, $p < 0,05$).

Entretanto, precisamos relatar os testes de acompanhamento (incluindo os tamanhos de efeito):

- ✓ A contagem de espermatozoides foi significativamente afetada pela ingestão de refeições com soja ($H(3) = 8,66$, $p < 0,05$). Testes de Mann-Whitney foram utiliza-

dos para acompanhamento desse achado. Uma correção de Bonferroni foi aplicada e todos os efeitos foram testados no nível de 0,0167 de significância. Pareceu que a contagem de espermatozoides não apresentou diferenças com uma refeição com soja por semana ($U = 191$, $r = -0,04$) ou com quatro refeições por semana ($U = 188$, $r = -0,05$) quando comparadas com refeições sem soja. Contudo, quando sete refeições de soja por semana foram observadas, a contagem de espermatozoides foi significativamente menor do que quando comparada com refeições sem soja ($U = 104$, $r = -0,41$). Podemos concluir que se a soja for comida todo dia ela reduz significativamente a contagem de espermatozoides comparada com a não ingestão de soja; contudo comer soja frequentemente mas não todos os dias não apresenta efeitos significativos na contagem de espermatozoides!

Ou podemos relatar a tendência observada:

- ✓ Todos os efeitos são relatados ao nível $p < 0,05$. A contagem de espermatozoides é significativamente afetada pela ingestão de refeições com soja ($H(3) = 8,66$). O teste de Jonckheere revelou uma tendên-

cia significativa nos dados de quanto mais soja é ingerida mais a mediana da contagem de espermatozoides desce, $J = 912$, $z = -2,48$, $r = -0,28$.

13.5 DIFERENÇAS ENTRE VÁRIOS GRUPOS RELACIONADOS: A ANOVA DE FRIEDMAN ①

No capítulo 11, descobrimos uma técnica denominada ANOVA de um fator que pode ser utilizada para testar diferenças entre vários grupos relacionados. Embora, como vimos, a ANOVA seja robusta a violações de seus pressupostos, existe uma alternativa para o caso de medidas repetidas: a ANOVA de Friedman (1937). Ela é utilizada para testar diferenças entre condições experimentais quando existem mais do que duas condições e os mesmos participantes foram utilizados em todas as condições (cada pessoa contribuiu com várias medidas para os dados). Se tivermos dados não normalmente distribuídos ou tivermos outra suposição violada, esse teste pode ser útil para contornar o problema.

Pessoas jovens (mulheres principalmente) podem tornar-se obsessivas em relação ao peso e dietas, e em virtude da mídia estar in-

Dica da Samanta Ferrinho



- O teste de Kruskal-Wallis **compara várias condições quando diferentes pessoas participam em cada condição e os dados resultantes não são normalmente distribuídos ou não satisfazem a hipótese da ANOVA independente de um fator.**
- Observe a linha Sig. Assint. (*Asymp. Sig.*): se o valor for menor do que 0,05, os grupos são significativamente diferentes.
- Você pode realizar um acompanhamento da análise principal com testes em pares de Mann-Whitney, mas só deverá aceitá-los como significativos se eles tiverem uma significância inferior a 0,05/número de testes.
- Se você prevê que as médias irão aumentar ou diminuir junto com os grupos em certa ordem, utilize o teste de tendência de Jonckheere para confirmar ou não a previsão.
- Relate a estatística H, os graus de liberdade e o valor da significância para a análise principal. Para qualquer teste *post hoc*, relate a estatística e um tamanho de efeito se possível (você pode ainda apresentar o valor de z correspondente e o valor da significância). Relate ainda as medianas e suas amplitudes correspondentes (ou faça um diagrama de caixa e bigodes).

sistentemente veiculando imagens de celebridades ridiculamente magras e altas e fazendo lavagens cerebrais para que acreditemos que esses cabides são corpos atraentes, nós ficamos deprimidos com nossas imperfeições. Assim, outros parasitas se aproveitam da nossa vulnerabilidade ganhando rios de dinheiro com dietas que supostamente irão deixar nossos corpos maravilhosos! Não querendo perder essa grande oportunidade de explorar a insegurança das pessoas, apresento minha própria dieta: a “dieta Andikins.”⁵ O princípio básico é que você se comporte como eu: não coma carne, tome muito chá Darjeeling, coma muito queijo europeu, grande quantidade de pão crocante e fresquinho, massa e chocolate em qualquer oportunidade disponível (especialmente quando estiver escrevendo livros), deguste algumas cervejas no final de semana, jogue futebol e rúgbi pelo menos duas vezes por semana e toque bateria por pelo menos uma hora por dia até que sua namorada ameace cortar fora os seus braços e bater na sua cabeça com eles por fazer muito barulho. Para testar a eficácia da minha maravilhosa dieta, escolhi 10 mulheres que achavam que deviam perder algum peso e as coloquei na

dieta por dois meses. Os seus pesos foram mensurados em quilogramas no início da dieta, após um mês e no final do segundo mês.

13.5.1 Teoria da ANOVA de Friedman ②



A teoria da ANOVA de Friedman é semelhante a dos outros testes vistos neste capítulo: ela é baseada nos postos dos dados. Para começar, você simplesmente coloca os dados para as diferentes condições em colunas diferentes (nesse caso, existem três condições, portanto teremos três colunas). Os dados para o exemplo da dieta estão na Tabela 13.5; note que os dados estão em colunas diferentes, desse modo, cada linha representa o peso de uma pessoa diferente. Em seguida, precisamos determinar os postos dos dados para cada pessoa. Assim, começamos com a pessoa 1, olhamos para o seu peso ou escore (verificamos que ela pesava 63,75 kg no início, 65,38 no final do primeiro mês e 81,34 kg no final do segundo mês de dieta), então atribuímos ao menor valor o posto 1, ao seguinte o posto 2 e assim por diante (veja a Seção 13.2.1 para mais detalhes). Quando você tiver determinado todos os postos para a primeira pessoa, passe para a segunda e assim por diante. Faça

⁵ Não confundir com a dieta de Atkins, obviamente. ☺

Tabela 13.5 Dados para o exemplo da dieta com postos

	Peso			Postos do Início	Peso	
	Início	Mês 1	Mês 2		Postos do mês 1	Postos do mês 2
Pessoa 1	63,75	65,38	81,34	1	2	3
Pessoa 2	62,98	66,24	69,31	1	2	3
Pessoa 3	65,98	67,70	77,89	1	2	3
Pessoa 4	107,27	102,7	91,33	3	2	1
Pessoa 5	66,58	69,45	72,87	1	2	3
Pessoa 6	120,46	119,96	114,26	3	2	1
Pessoa 7	62,01	66,09	68,01	1	2	3
Pessoa 8	71,87	73,62	55,43	2	3	1
Pessoa 9	83,01	75,81	71,63	3	2	1
Pessoa 10	76,62	67,66	68,60	3	1	2
	R_j			19	20	21

isso para todas as pessoas para as quais os dados foram coletados. Depois, simplesmente adicione os postos de cada condição (R_i , onde i é utilizado para representar um grupo). A Tabela 13.5 mostra os dados brutos para esse exemplo mais os postos (tente determiná-los e veja se obtém os mesmos resultados apresentados aqui!).

Com a soma dos postos calculada para cada grupo, a estatística teste, F_r , é calculada pela equação (13.3):

$$F_r = \left[\frac{12}{Nk(k+1)} \sum_{i=1}^k R_i^2 \right] - 3N(k+1) \quad (13.3)$$

Nessa equação, R_i é a soma dos postos para cada grupo, N é o total da amostra (nesse caso, 10) e k é o número de condições (nesse caso, 3). Essa equação é bastante semelhante a do teste de Kruskal-Wallis (compare as equações (13.1) e (13.3)). Tudo o que precisamos fazer para cada condição é elevar a soma dos postos ao quadrado e somar esses valores. Isso resolve o meio da equação; o resto dela envolve calcular vários valores com base no tamanho total da amostra e no número de condições. Para esses dados, temos:

$$\begin{aligned} F_r &= \left[\frac{12}{(10 \times 3)(3+1)} (19^2 + 20^2 + 21^2) \right] \\ &\quad - (3 \times 10)(3+1) \\ &= \frac{12}{120} (361 + 400 + 441) - 120 \\ &= 0,1(1202) - 120 \\ &= 120,2 - 120 \\ &= 0,2 \end{aligned}$$



Quando o número de pessoas testadas é grande (maior do que 10), essa estatística teste, como a de Kruskal-Wallis na seção anterior, tem uma distribuição qui-quadrado (veja o Capítulo 16) e para essa distribuição existe um valor para os graus de liberdade, que é um menos que o número de grupos ($k - 1$); nesse caso, 2.

13.5.2 Atribuindo dados e fazendo análise condicional ①

Quando os dados são coletados utilizando os mesmos participantes em cada condição, eles precisam ser apresentados em diferentes colunas. Assim, o editor de dados terá três colunas de dados. A primeira coluna é para os dados do início da dieta (denominada **Start**), a segunda coluna terá os pesos após um mês de dieta (denominada **month1**) e a coluna final terá os pesos do final da dieta (denominada **month2**). Esses dados podem ser encontrados no arquivo denominado **Diet.sav**.

Primeiro execute uma análise exploratória nos dados (veja o Capítulo 3). Com um pouco de sorte, você irá obter uma tabela semelhante a da Saída 13.12 do SPSS, que mostra que o teste K-S (veja a Seção 3.5) não foi significativo para os pesos do início da dieta ($D(10) = 0,23$, $P > 0,05$), mas o teste de Shapiro-Wilk é significativo e esse teste é, de fato, mais preciso do que o teste K-S (veja o Capítulo 3). Os dados do final do primeiro mês de dieta foram significativamente diferentes da normal ($D(10) = 0,34$, $p < 0,01$), embora os dados do final do segundo mês da dieta pareçam normais ($D(10) = 0,20$, $p > 0,05$). Alguns desses dados não são normalmente distribuídos, assim, devemos utilizar testes não-paramétricos.

13.5.3 Executando a ANOVA de Friedman no SPSS ①

Acesse a caixa de diálogo principal utilizando o menu **Analyze** ⇒ **Nonparametric Tests** ⇒ **K Related Samples...** (Analisar ⇒ Testes Não-paramétricos ⇒ K Amostras Relacionadas...) (veja a Figura 13.8). Com a caixa de diálogos ativa, selecione as três variáveis dependentes da lista (clique em **start** e, então, mantendo o botão do *mouse* pressionado, arraste o mouse sobre **month1** e **month2**). Transfira as variáveis para o quadro **Test Variables** (Variáveis de Teste) clicando em . Se você clicar em (Exato), obterá uma caixa de diálogo para selecionar o valor exato da significância para a ANOVA de Friedman (veja a Seção 13.2.3). Lembre que

Saída 13.12 do SPSS

Tests of Normality (Testes de Normalidade)

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic (Estatística)	df (gl)	Sig. (Sig.)	Statistic (Estatística)	df (gl)	Sig. (Sig.)
Weight at Start (Kg) (Peso, em kg, no início)	0.228	10	0.149	0.784	10	0.009
Weight after 1 month (Kg) (Peso, em kg, após 1 mês)	0.335	10	0.002	0.685	10	0.001
Weight after 2 month (Kg) (Peso, em kg, após 2 meses)	0.203	10	0.200*	0.877	10	0.121

* Esse é um limite inferior da verdadeira significância.
a Correção da significância de Lilliefors.

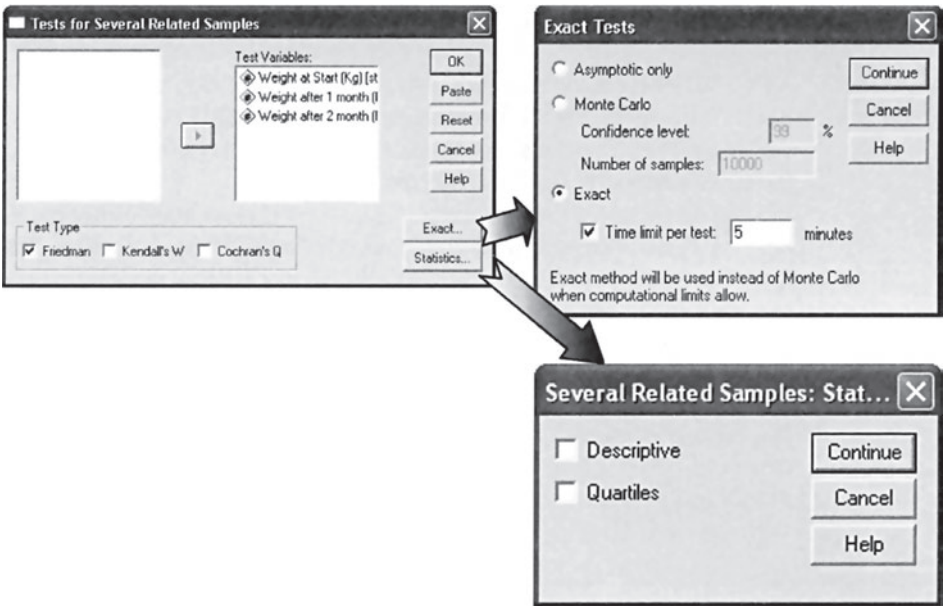


Figura 13.8 Caixas de diálogo para ANOVA de Friedman.

eu disse que quando as amostras são grandes, você deve provavelmente optar pelo método de Monte Carlo, mas com pequenas amostras vale a pena optar pelo teste exato. Você tem uma amostra de tamanho pequeno aqui, assim, pode optar pela significância exata. Finalmente, clique em **Statistics...** (Estatísticas) para selecionar algumas estatísticas descritivas. Para obter a análise, retorne à caixa de diálogos principal e clique em **OK**. Você pode também realizar alguns testes relacionados utilizando a mesma caixa de diálogo da ANOVA de Friedman (veja o Quadro 13.5).

13.5.4 Saídas para a ANOVA de Friedman ①

Como vimos, a ANOVA de Friedman, assim como todos os testes não-paramétricos deste capítulo, é baseada em postos e não nos dados mensurados. A Saída 13.13 do SPSS mostra a média dos postos em cada condição. Essas médias são importantes para mais tarde interpretarmos qualquer efeito: elas mostram que os postos são semelhantes ao longo das condições.

A Saída 13.14 do SPSS mostra a estatística teste (o SPSS a chama de qui-quadrado

Quadro 13.5

Outras opções ②

Na caixa de diálogo principal existem algumas outras opções que podem ser selecionadas:

- **Kendall's W:** (W de Kendall): Esse teste é bastante semelhante à ANOVA de Friedman, mas é utilizado especificamente para verificar a concordância entre razões (taxas). Assim, se, por exemplo, solicitássemos a 10 mulheres diferentes para avaliar a atratividade de Justin Timberlake, David Beckham e Tony Blair, poderíamos utilizar esse teste para verificar o quanto elas concordam. Esse teste é particularmente útil porque, como o coeficiente de correlação, o W de Kendall tem uma variação limitada: ele vai de 0 (nenhuma concordância entre os julgadores) a 1 (concordância completa entre os juízes).
- **Cochran's Q** (Q de Cochran). É uma extensão do teste de McNemar (veja o Quadro 13.3), e é basicamente um teste de Friedman para quando os dados são dicotômicos. Assim, imagine que você perguntou para 10 mulheres se elas gostariam de abraçar e beijar Justin Timberlake, David Beckham e Tony Blair e elas só poderiam responder sim ou não. Se as respostas fossem codificadas como 0 (não) e 1 (sim), poderíamos realizar um teste de Cochran nesses dados.

Saída 13.13 do SPSS

Descriptive Statistics (Estatísticas Descritivas)

	N	Mean (Média)	Std. Deviation (Desvio Padrão)	Minimum (Mínimo)	Maximum (Máximo)	Percentiles (Percentis)		
						25th (25 ^o)	50 th (Median) (50 ^o Mediana)	75th (75 ^o)
Weight at Start (Kg) (Peso, em kg, no início)	10	78.0543	20.23008	62.01	120.46	63.5549	69.2288	89.0709
Weight after 1 month (Kg) (Peso, em kg, após 1 mês)	10	77.4635	18.61502	65.38	119.96	66.2065	68.5728	82.5385
Weight after 2 month (Kg) (Peso, em kg, após 2 meses)	10	77.0668	16.10612	55.43	114.26	68.4525	72.2493	83.8365

Ranks (Postos)

	Mean Rank (Posto Médio)
Weight at Start (Kg) (Peso, em kg, no início)	1.90
Weight after 1 month (Kg) (Peso, em kg, após 1 mês)	2.00
Weight after 2 month (Kg) (Peso, em kg, após 2 meses)	2.10

em vez de F_r porque a F_r tem uma distribuição qui-quadrado). O valor dessa estatística é 0,20, o mesmo valor que calculamos anteriormente. Também já mencionamos os graus de liberdade da estatística teste (nesse caso, existem três grupos, assim, os graus de liberdade são: $3 - 1 = 2$) e a significância. Se você solicitar

a significância exata, esse valor também será calculado. O valor da significância é 0,905 (ou 0,974 se você ler a significância exata), o que está bem acima de 0,05; dessa forma, podemos concluir que a dieta Andikins não tem qualquer efeito: os pesos não mudaram significativamente ao longo do período da dieta.

Saída 13.14 do SPSS

Test Statistics^a (Estatísticas Teste)

N	10
Chi-Square (Qui-Quadrado)	0.200
df (gl)	2
Asymp. Sig. (Sig. Assint.)	0.905
Exact Sig. (Sig. Exata)	0.974
Point Probability (probabilidade do valor)	0.143

a Teste de Friedman.

13.5.5 Análise *post hoc* para a ANOVA de Friedman ②

Como para o teste de Kruskal-Wallis, existem duas formas de realizar os procedimentos não-paramétricos *post hoc*, que, na essência, são os mesmos. O primeiro é utilizar o teste dos postos com sinais de Wilcoxon (Seção 13.3), mas com correção para o número de testes que serão realizados (veja as Seções 8.2.1 e 13.4.5 para lembrar os motivos). A forma de corrigir para o número de testes é aceitar somente como significativo se a significância do resultado for inferior a α dividido pelo número de comparações realizadas (a correção de Bonferroni). Nas ciências sociais, isso significa normalmente 0,05 dividido pelo número de comparações. Nesse exemplo, temos somente três grupos, assim, se compararmos todos os grupos teremos três comparações:

- ▶ Teste 1: Peso no início da dieta comparado com o peso após um mês;
- ▶ Teste 2: Peso no início da dieta comparado com o peso após dois meses;
- ▶ Teste 3: Peso após um mês de dieta comparado com o peso após dois meses.

Dessa forma, em vez de utilizarmos o nível de significância de 0,05, vamos utilizar o valor $0,05/3 = 0,0167$. Não vamos nos incomodar com testes *post hoc* nesse exemplo, porque a ANOVA principal não foi significativa, mas iremos apresentar o processo para ilustrar como ele deve ser feito. Você pode realizar os testes de Wilcoxon selecionando os pares de variáveis de cada comparação e transferindo-as para o quadro **Test Pair(s) List** (Lista de pares do teste) (veja a Figura 13.4).



A Saída 13.15 do SPSS mostra a estatística teste da realização do teste dos postos com sinais de Wilcoxon para todas as três comparações. Lembre que agora estamos considerando um valor crítico de 0,0167 e, de fato, nenhuma das comparações é significativa pois elas apresentam valores de significância unilaterais de 0,500, 0,423 e 0,461 (isso não é surpreendente porque a análise principal não foi significativa).

A segunda forma de realizar os testes *post hoc* é muito semelhante ao que foi feito para o teste de Kruskal-Wallis na Seção 13.4.5 e é, também, descrita por Siegel e Castellan (1988). Novamente, tomamos as diferenças entre os postos médios dos diferentes grupos e comparamos essas diferenças a um valor baseado no valor z (corrigido para o número de comparações sendo realizadas) e uma constante com base no total da amostra, N (10 para esse exemplo) e o número total de condições, k (3, nesse caso). Essa desigualdade é:

$$|\bar{R}_u - \bar{R}_v| \geq z_{\alpha/k(k-1)} \sqrt{\frac{k(k+1)}{6N}} \quad (13.4)$$

O lado esquerdo da equação é a diferença entre as médias dos postos dos dois grupos sendo comparados, mas onde o sinal é ignorado. Como ocorreu com o teste de Kruskal-Wallis, precisamos saber o valor de $z_{\alpha/k(k+1)}$, e se ficarmos com o nosso valor tradicional de 0,05, sabendo que $k = 3$, teremos $\alpha/k(k-1) = 0,05/3(3-1) = 0,05/6 = 0,00833$. Assim, $z_{\alpha/k(k-1)}$ apenas significa “o valor de z para o qual somente $\alpha/k(k-1)$ dos demais valores são maiores” (ou, nesse caso, “o valor de z para o qual somente 0,833% dos demais valores são maiores”). Portanto, olhe a tabela no Apêndice A.1 e procure na coluna *Parte Menor* o valor 0,00833, ou o mais próximo, e depois busque o valor na mesma linha na coluna z . Nesse caso, existem os valores de 0,00842 e 0,00820 que fornecem os valores- z de 2,39 e 2,40, respectivamente; como o valor 0,00833 está entre esses dois valores que encontramos, podemos tomar a média dos escores- z , ou seja, 2,395, ou podemos ser

Saída 13.15 do SPSS

Ranks (Postos)

		<i>N</i>	<i>Mean Rank</i> (Posto Médio)	<i>Sum of Rank</i> (Soma dos Postos)
<i>Weight after 1 month (Kg) – Weight at Start (kg)</i> (Peso, em kg, após 1 mês – Peso, em kg, no início)	<i>Negative Ranks</i> (Postos Negativos)	4 ^a	7.00	28.00
	<i>Positive Ranks</i> (Postos Positivos)	6 ^b	4.50	27.00
	<i>Ties</i> (Empates)	0 ^c		
	<i>Total</i> (Total)	10		
<i>Weight after 2 month (Kg) – Weight at Start (kg)</i> (Peso, em kg, após 2 meses – Peso, em kg, no início)	<i>Negative Ranks</i> (Postos Negativos)	5 ^d	6.00	30.00
	<i>Positive Ranks</i> (Postos Positivos)	5 ^e	5.00	25.00
	<i>Ties</i> (Empates)	0 ^f		
	<i>Total</i> (Total)	10		
<i>Weight after 2 month (Kg) – Weight after 1 month (kg)</i> (Peso, em kg, após 2 meses – Peso, em kg, após 1 mês)	<i>Negative Ranks</i> (Postos Negativos)	4 ^g	7.25	29.00
	<i>Positive Ranks</i> (Postos Positivos)	6 ^h	4.33	26.00
	<i>Ties</i> (Empates)	0 ⁱ		
	<i>Total</i> (Total)	10		

a Peso, em kg, após 1 mês < Peso, em kg, no início.

b Peso, em kg, após 1 mês > Peso, em kg, no início.

c Peso, em kg, após 1 mês = Peso, em kg, no início.

d Peso, em kg, após 2 meses < Peso, em kg, no início.

e Peso, em kg, após 2 meses > Peso, em kg, no início.

f Peso, em kg, após 2 meses = Peso, em kg, no início.

g Peso, em kg, após 2 meses < Peso, em kg, após um mês.

h Peso, em kg, após 2 meses > Peso, em kg, após um mês.

i Peso, em kg, após 2 meses = Peso, em kg, após um mês.

Test Statistics^b (Estatísticas de Teste)

	<i>Weight after 1 month (kg) – Weight at Start (Kg)</i> (Peso, em kg, após 1 mês – Peso, em kg, no início)	<i>Weight after 2 month (Kg) – Weight at Start (Kg)</i> (Peso, em kg, após 2 meses – Peso, em kg, no início)	<i>Weight after 2 month (Kg) – Weight after 1 month (kg)</i> (Peso, em kg, após 2 meses – Peso, em kg, após 1 mês)
<i>Z</i>	–0.051 ^a	–0.255 ^a	–0.153 ^a
<i>Asymp. Sig. (2-tailed)</i> (Sig. Assint. (bilateral))	0.959	0.799	0.878
<i>Exact Sig. (2-tailed)</i> (Sig. Exata (unilateral))	1.000	0.846	0.922
<i>Exact Sig. (1-tailed)</i> (Sig. Exata (unilateral))	0.500	0.423	0.461
<i>Point Probability</i> (Probabilidade Pontual)	0.039	0.038	0.038

a Baseado em postos positivos.

b Testes de Postos com Sinais de Wilcoxon.

cautelosos e utilizar o valor 2,39, ou atrevidos e utilizar o valor 2,40. De qualquer forma, agora podemos calcular o lado direito da equação (13.4):

$$\begin{aligned} \text{Diferença crítica} &= z_{\alpha/k(k-1)} \sqrt{\frac{k(k+1)}{6N}} \\ &= 2,40 \sqrt{\frac{3(3+1)}{6(10)}} \end{aligned}$$

$$\begin{aligned} &= 2,40 \sqrt{\frac{12}{60}} \\ &= 2,40 \sqrt{0,2} \\ &= 1,07 \end{aligned}$$

Como as mesmas pessoas foram utilizadas, a mesma diferença crítica pode ser usada em todas as comparações. O próximo passo é calcular todas as diferenças entre as médias

dos postos de todos os grupos (as médias dos postos podem ser encontradas na Saída 13.13 do SPSS):

Tabela 13.6 Diferenças entre postos médios para os dados da dieta

Comparação	\bar{R}_u	\bar{R}_v	$\bar{R}_u - \bar{R}_v$	$ \bar{R}_u - \bar{R}_v $
Início – Primei-ro mês	1,90	2,00	-0,10	0,10
Início – Segun-do mês	1,90	2,10	-0,20	0,20
Primeiro mês – Segundo mês	2,00	2,10	-0,10	0,10



A equação (13.4) significa que se as diferenças entre os postos médios são maiores ou iguais à diferença crítica, essa diferença é significativa. Nesse caso, isso significa que se uma diferença for maior do que 1,07, ela será significativa. Todas as diferenças estão abaixo desse valor (Tabela 13.6), assim, podemos concluir que nenhum dos grupos difere significativamente e isso é consistente com a não-significância inicial da ANOVA.

13.5.6 Calculando o tamanho de efeito ②

Como mencionado anteriormente, não existe uma forma fácil de converter uma estatística qui-quadrado que tem mais de um grau de liberdade em um tamanho de efeito r e, de qualquer modo, nem sempre interessa ter um tamanho de efeito geral como o testado pela ANOVA de Friedman.⁶ Contudo, existe uma maneira mais sensata (em minha opinião, ao menos) de calcular tamanhos de efeito para qualquer comparação que tenhamos feito na ANOVA. Como vimos na Seção 13.3.5, é simples obtermos um tamanho de efeito r a partir

do teste de postos com sinais de Wilcoxon. De forma alternativa, podemos calcular tamanhos de efeito para os testes de Wilcoxon que foram utilizados como seguimento da análise principal. Esses efeitos serão bastante informativos na sua forma apropriada.

Para a primeira comparação (início *versus* final do primeiro mês), a Saída 13.15 do SPSS mostra que z é $-0,051$, e porque esse valor foi obtido com base na comparação de duas condições contendo 10 observações, temos 20 observações no total (lembre que o fato das observações terem sido feitas nas mesmas pessoas não importa aqui). O tamanho de efeito é:

$$\begin{aligned} r_{\text{Início-mês1}} &= \frac{-0,051}{\sqrt{20}} \\ &= -0,01 \end{aligned}$$

Para a segunda comparação (início *versus* final do segundo mês), a Saída 13.15 do SPSS mostra que z é $-0,255$ e isso teve como base 20 observações. O tamanho de efeito será:

$$\begin{aligned} r_{\text{Início-mês2}} &= \frac{-0,255}{\sqrt{20}} \\ &= -0,06 \end{aligned}$$

Para a comparação final (final do mês 1 *versus* final do mês 2) a Saída 13.15 do SPSS mostra que z é $-0,153$ e esse valor foi novamente baseado em 20 observações. Assim, teremos para o tamanho de efeito:

$$\begin{aligned} r_{\text{mês1-mês2}} &= \frac{-0,153}{\sqrt{20}} \\ &= -0,03 \end{aligned}$$

De forma não surpreendente, dada a não-significância dos testes de Wilcoxon, todas representam um efeito praticamente inexistente: todas estão muito próximos de zero.

13.5.7 Escrevendo e interpretando os resultados ①

Para a ANOVA de Friedman precisamos somente relatar a estatística teste (que, como

⁶ Se quiser, então, você pode (como no caso do teste de Kruskal-Wallis) utilizar o valor da significância da estatística teste qui-quadrado para encontrar um valor associado de z a partir de uma tabela de valores de probabilidade para a distribuição normal (veja o Apêndice A.1) e então convertê-lo para r como já foi visto neste capítulo.

já vimos, é representada por χ^2 ,⁷ os graus de liberdade e a significância. Assim, podemos escrever que:

- ✓ O peso dos participantes não mudou significativamente ao longo dos dois meses de dieta ($\chi^2(2) = 0,20$, $p > 0,05$).

Embora sem a significância da análise inicial, queremos relatar os testes *post hoc* para esses dados, caso você tenha que fazê-lo. Assim, você deve afirmar que (lembre que a estatística teste T é o menor valor das duas somas de postos para cada teste e esses valores estão na Saída 13.15 do SPSS):

- ✓ O peso dos participantes não mudou significativamente ao longo dos dois meses de dieta ($\chi^2(2) = 0,20$, $p < 0,05$). Testes de Wilcoxon foram utilizados para acompanhar esses achados. Uma correção de Bonferroni foi aplicada e todos os efeitos foram testados com um nível de significância de 0,0167. Pareceu que os pesos não sofreram alterações significativas do início da dieta até o primeiro mês $T = 27$, $r = -0,01$, do início da dieta até o segundo mês, $T = 25$, $r = -0,06$ e do primeiro

mês da dieta para o segundo mês, $T = 26$, $r = -0,03$. Podemos concluir que a dieta Andikins, assim como o seu criador, é um completo fracasso.

13.6 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Testes como o t e a ANOVA são robustos à violação de suas hipóteses, e correções podem ser feitas para assegurar que esses testes paramétricos sejam precisos quando as suposições são violadas. Este capítulo apresentou abordagens alternativas para a violação das hipóteses paramétrica, que consiste em utilizar testes tendo como base postos em vez dos dados reais mensurados. Começamos vendo o teste da soma dos postos de Wilcoxon e o teste de Mann-Whitney, utilizados para comparar dois grupos independentes. Os testes nos possibilitaram verificar alguns detalhes do processo de trabalhar com postos. A seguir, verificamos o teste dos postos com sinais de Wilcoxon, utilizado para comparar duas condições relacionadas. Examinamos situações complexas que envolvem mais de duas condições ou grupos (o teste para condições independentes de Kruskal-Wallis e a ANOVA de Friedman para condições ou grupos relacionados). Para cada um desses testes, focamos na teoria do teste

⁷ A estatística teste é frequentemente representada por χ^2_{F} , mas o guia de estilos oficial da American Psychology Association (APA) não reconhece esse termo.

Dica da Samanta Ferrinho



- A ANOVA de Friedman **compara várias condições quando as mesmas pessoas participam em cada condição e os dados resultantes não são normalmente distribuídos ou violam a hipótese da ANOVA de medidas repetidas de um fator.**
- Observe a linha Sig. Assint. (*Asymp. Sig.*): se o valor for menor do que 0,05, os grupos são significativamente diferentes.
- Você pode realizar um acompanhamento da análise principal com testes de postos com sinais de Wilcoxon, mas só deverá aceitá-los como significativos se eles tiverem uma significância inferior a 0,05/número de testes.
- Relate a estatística χ^2 , os graus de liberdade e o valor da significância para a análise principal. Para qualquer teste *post hoc*, relate a estatística T e um tamanho de efeito se possível (você pode ainda apresentar o valor de z correspondente e o valor da significância). Relate ainda as medianas e suas amplitudes correspondentes (ou faça um diagrama de caixa e bigodes).

(embora essa seção possa ser pulada) e depois detalhamos como conduzi-los com o auxílio do SPSS, como interpretar os resultados e como relatar os resultados de cada um dos testes apresentados. Nesse processo, descobrimos que drogas lhe deixam deprimido, que a soja reduz a contagem de espermatozoides e que o meu estilo de vida não leva à perda de peso! Isso conclui nosso estudo sobre a comparação de grupos. A última seção final do livro aborda alguns conteúdos pesados. Hora de parar e tomar um drinque.

13.7 TERMOS-CHAVE QUE DESCOBRIMOS

- Q de Cochran
- ANOVA de Friedman
- Teste de Jonkheere-Terpstra
- W de Kendall
- Z de Kolmogorov-Smirnov
- Teste de Kruskal-Wallis
- Teste de Mann-Whitney
- Teste de McNemar
- Teste da Mediana
- Método de Monte Carlo
- Reações extremas de Moses
- Testes não-paramétricos
- Postos
- Testes dos Sinais
- Corrida de Wald-Wolfowitz
- Teste da soma dos postos de Wilcoxon
- Teste dos postos com sinais de Wilcoxon

13.8 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Uma psicóloga estava interessada em verificar diferenças entre espécies de homens e cachorros. Ela observou um grupo de cães e outro grupo de homens em uma situação natural (20 de cada). Ela classificou vários comportamentos como sendo tipo-cão (urinar em árvores e postes, tentar copular com qualquer coisa que se mova e tentar lambe os próprios genitais). Para cada homem e cachorro, ela contou o número de comportamen-

tos tipo-cão observado num período de 24 horas. Ela apresentou a hipótese que os cachorros apresentam mais comportamentos do tipo-cão do que os homens. Os dados estão no arquivo **MenLikeDogs.sav**. Analise esses dados com um teste de Mann-Whitney. ①

- **Tarefa 2:** Muita especulação já foi feita sobre a influência de mensagens subliminares em discos. Por exemplo, tanto Ozzy Osbourne quanto Judas Priest foram acusados de colocar mensagens subliminares nos seus discos que influenciam adolescentes a fazer coisas como estourar a própria cabeça com uma arma. Um psicólogo está interessado em verificar se esse tipo de mensagem, de fato, tem algum efeito. Ele pegou a fita original da música “*Baby one more time*” (Amor mais uma vez) da Britney Spears e criou uma segunda versão que com a mensagem subliminar “*deliver your soul to the dark lord*” (entregue sua alma para o capeta) repetida no coro. Ele pegou essa versão e a original e tocou uma versão, ao acaso, para um grupo de 32 pessoas. Ele pegou o mesmo grupo seis meses depois e tocou a versão que a pessoa não tinha ouvido anteriormente. Assim cada pessoa ouviu tanto a versão original quanto a versão com a mensagem subliminar, mas em momentos diferentes. O psicólogo depois contou o número de bodes sacrificados na semana seguinte após cada versão ter sido ouvidas. Ele levantou a hipótese de que a mensagem subliminar levaria ao sacrifício de mais bodes. Os dados estão no arquivo **DarkLord.sav**; analise esses dados com o teste de postos com sinais de Wilcoxon. ①
- **Tarefa 3:** Uma psicóloga está interessada no efeito dos programas televisivos na vida doméstica. Ela lançou a hipótese que por meio de “aprender vendo”, certos programas podem incentivar as pessoas a se comportarem como os personagens do programa. Isso, por sua vez, pode afetar o relacionamento do telespectador (dependendo se o programa mostra relacionamentos harmoniosos ou conflituosos).

Ela pegou episódios de três programas de TV populares e os mostrou a 54 casais. Depois, cada casal foi deixado sozinho em uma sala por uma hora. O pesquisador mediu o número de vezes que o casal discutiu. Cada casal viu todos os três programas de TV em tempos diferentes (com uma semana de intervalo) e a ordem em que os programas foram vistos foi contrabalançada entre os casais. Os programas de TV selecionados foram *EastEnders** (que retrata as vidas de pessoas extremamente miseráveis, briguentas, que não gostam de nada além de bater um no outro, mentir um para o outro, dormir com as mulheres uns dos outros e não mostrar consideração alguma pelos demais seres humanos!), *Friends*** (que retrata um grupo não-realista de pessoas simpáticas e boas que se amam) e um programa da National Geographic sobre baleias (esse deve agir como controle). Os dados estão no arquivo **Eastenders.sav**; acesse o arquivo e realize uma ANOVA de Friedman. ①

- **Tarefa 4:** Uma pesquisadora está interessada em tentar prevenir a coulrofobia (medo de palhaços) em crianças. Ela decidiu realizar um experimento no qual diferentes grupos de crianças (15 em cada) foram expostas a diferentes formas de informações positivas sobre palhaços. O primeiro grupo assistiu a algumas propagandas do McDonald's em que Ronald McDonald é visto pulando com crianças e

dizendo que eles devem amar suas mães. A um segundo grupo foi contada uma história sobre um palhaço que ajudou algumas crianças que tinham se perdido na floresta (que diabos um palhaço fazia na floresta é um mistério). Um terceiro grupo foi entretido por um palhaço real que veio até a sala de aula e fez animais com balões para as crianças.⁸ Um grupo final foi deixado como controle e a eles nada foi oferecido. A pesquisadora fez uma escala própria da medida de quanto as crianças gostam de palhaços resultando em um escore que varia de 0 (sem qualquer medo de palhaços) até 5 (com muito medo de palhaços). Os dados estão no arquivo **coulrophobia.sav**; acesse os dados e realize um teste de Kruskal-Wallis. ①

As respostas para as questões estão no arquivo **Answers(Chapter13).pdf**. Como esses exemplos são do livro de Field e Hole (2003), você pode copiar o Capítulo 7 dessa obra para obter algumas respostas detalhadas.

13.9 LEITURAS COMPLEMENTARES

Siegel, S & Castellan, N. J. *Nonparametric statistics for the behavioral sciences* (2nd edition). New York: McGraw-Hill. Texto definitivo sobre estatística não-paramétrica. Talvez não seja um bom livro para os que têm fobia a estatística, mas se você foi bem-sucedido neste capítulo, esse livro é um próximo passo excelente.

* N. de T.: Série britânica, parecida com uma novela, popular e premiada que começou em 1985 no canal BBC1.

**N. de T.: Seriado americano transmitido pela rede NBC entre 1994 e 2004. Conta a vida de seis amigos que moram em Nova York.

⁸ Infelizmente, na primeira tentativa que o estudo foi feito o palhaço acidentalmente estourou um balão. O barulho assustou as crianças e elas associaram o medo ao palhaço. Todas as 15 crianças estão hoje fazendo terapia para a coulrofobia.

14

ANÁLISE DE VARIÂNCIA MULTIVARIADA (MANOVA)

14.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②



Nos últimos cinco capítulos, vimos como o Modelo Linear Geral (MLG) pode ser utilizado para detectar diferenças entre grupos em uma única variável dependente. Contudo,

existem muitas circunstâncias em que estamos interessados em diversas variáveis dependentes e, nesses casos, o modelo simples da ANOVA não é adequado. Em vez dele, podemos utilizar uma extensão dessa técnica conhecida como Análise de Variância Multivariada (ou MANOVA). A MANOVA pode ser pensada como uma ANOVA para situações em que existem diversas variáveis dependentes. Os princípios da ANOVA são estendidos à MANOVA no sentido de que podemos utilizar a MANOVA quando existe apenas uma variável independente ou quando existem várias, podemos olhar as interações entre variáveis independentes e podemos ainda realizar contrastes para ver quais grupos diferem entre si. A ANOVA pode ser utilizada somente em situações em que existe uma única variável dependente (ou saída),

por isso é conhecida como um teste univariado (univariado significa “uma variável”). A MANOVA trabalha com diversas variáveis dependentes (ou saídas) simultaneamente e, portanto, é um teste multivariado (multivariado quer dizer “muitas variáveis”). Este capítulo irá explicar o básico sobre a MANOVA, para aqueles que querem pular a teoria e ir direto para a prática. Mas para quem quiser saber mais, existe uma extensa seção teórica que explica como a técnica funciona. Depois, veremos um exemplo utilizando o SPSS e aprenderemos como as saídas do teste podem ser interpretadas. Isso vai nos levar a outro teste estatístico, conhecido como *Análise Discriminante*.

14.2 INTRODUÇÃO: SEMELHANÇAS E DIFERENÇAS COM A ANOVA ②

Se tivermos coletado dados sobre diversas variáveis dependentes, podemos simplesmente realizar uma ANOVA para cada variável dependente (e se você lê artigos de pesquisa, verá que os pesquisadores fazem isso com frequência!). No Capítulo 8, uma questão semelhante foi apresentada sobre o uso da ANOVA em vez de múltiplos testes *t*. A resposta de por que a MANOVA é utilizada em vez de várias

ANOVAs é a mesma: quanto mais testes realizarmos sobre os mesmos dados, mais estaremos inflacionando a taxa de erro de conjunto (veja a Seção 8.2.1). Quanto mais variáveis dependentes forem mensuradas, mais testes ANOVA serão necessários e maior será a probabilidade de cometermos um erro do Tipo I.

Por que não
fazer várias
ANOVAs?



Entretanto, existem outros motivos para preferirmos a MANOVA em vez de vários testes ANOVA. Existem informações adicionais importantes que são

obtidas da MANOVA. Se ANOVAs separadas forem realizadas para cada variável dependente, qualquer relacionamento entre todas as variáveis será ignorado. Desse modo, perdemos informações sobre as correlações que podem existir entre as variáveis dependentes. A MANOVA, ao incluir todas as variáveis dependentes na mesma análise, leva em conta o relacionamento entre as variáveis de saída. Relacionado a essa questão, a ANOVA pode nos informar se os grupos diferem ao longo de uma única dimensão enquanto a MANOVA tem o poder de detectar se os grupos diferem ao longo de uma combinação de dimensões. Por exemplo, a ANOVA nos informa como os escores de uma única variável dependente distinguem grupos de participantes (por exemplo, podemos distinguir motoristas, não motoristas e motoristas bêbados pelo número de pedestres que eles mataram). A MANOVA incorpora informações sobre várias medidas de saída e, dessa forma, nos informa se os grupos de participantes podem ser distinguidos por uma combinação de escores em várias medidas dependentes. Por exemplo, pode não ser possível distinguir motoristas, não motoristas e motoristas bêbados somente pelo número de pedestres que eles mataram, mas eles podem ser distinguidos por uma combinação do número de pedestres que mataram, o número de postes que eles derrubaram e o número de carros em que eles bateram. Nesse sentido, a MANOVA tem um poder maior de detectar um efeito, porque ela pode detectar se

os grupos diferem ao longo de uma combinação de variáveis, enquanto a ANOVA pode detectar somente diferenças entre grupos com base em uma única variável (veja Huberty e Morris, 1989 e a Seção 14.2.2). Por essas razões, é preferível realizar uma MANOVA do que vários testes ANOVA.

14.2.1 Advertência ②

Pela minha descrição, a MANOVA parece um teste rotineiro que permite medir centenas de variáveis dependentes e depois jogá-las na análise. Mas as coisas não são bem assim. Não é recomendável amontoar todas as variáveis dependentes juntas em uma MANOVA a menos que você tenha uma boa base teórica ou empírica para fazê-lo. Mencionei no começo do livro que os procedimentos estatísticos são apenas uma forma de triturar números; assim, mesmo que você coloque lixo na análise, ainda será possível obter conclusões estatisticamente significativas, mas que empiricamente não tem sentido. Em circunstâncias em que existe uma boa base teórica para incluir algumas, mas não todas, de suas variáveis dependentes, você deve executar análises separadas: uma para as variáveis sendo testadas de uma forma heurística e uma para as variáveis que sejam teoricamente significativas. O importante é não incluir montes de variáveis na MANOVA só porque elas foram coletadas.

14.2.2 Controvérsias ②

Da mesma forma que a ANOVA, a MANOVA é um teste de dois estágios em que um teste global é executado primeiro, seguido por testes mais específicos, aplicados para descobrir diferenças entre grupos. Infelizmente, existem controvérsias sobre os dois estágios do procedimento MANOVA. Para a análise principal, existem quatro formas comumente utilizadas de avaliar a significância global de uma MANOVA, e os debates giram em torno de qual dos métodos é o melhor considerando o poder e o tamanho da amostra. Além disso, existe um debate sobre a melhor forma de analisar e interpretar as diferenças entre os grupos

quando a MANOVA global é significativa. Existem duas abordagens principais para fazer a análise: a ANOVAs univariadas e a análise discriminante. Considerarei ambas.

Outra controvérsia está relacionada ao poder da MANOVA (isto é, a habilidade da MANOVA de detectar um efeito que de fato existe). Mencionei nas seções anteriores que a MANOVA tem mais poder que uma ANOVA para detectar efeitos porque ela pode levar em consideração as correlações entre as variáveis dependentes. Entretanto, a questão do poder é mais complexa que isso. Ramsey (1982) descobriu que à medida que as correlações entre as variáveis dependentes aumentam, o poder da MANOVA diminui. Isso levou Tabachnick e Fidell (2001) a afirmar que a MANOVA “apresenta melhores resultados com variáveis dependentes altamente correlacionadas de forma negativa e resultados aceitáveis quando as variáveis dependentes são moderadamente correlacionadas em qualquer direção, e que a MANOVA é uma perda de tempo quando as variáveis dependentes não são correlacionadas” (p. 357). Em contraste, Stevens (1980), investigando o efeito das correlações das variáveis dependentes no poder do teste, revelou que “o poder com altas intercorrelações é, em muitos casos, maior do que com intercorrelações moderadas e, em alguns casos, muito maior” (p. 736). Esses achados são contraditórios, o que nos deixa com a seguinte dúvida: qual é, exatamente, o relacionamento entre o poder e as intercorrelações das variáveis dependentes? Felizmente, Cole e colaboradores (1994) explicam muito bem esse relacionamento. Eles descobriram que o poder da MANOVA está associado a uma combinação das correlações entre variáveis e o *tamanho de efeito*. Resumindo, se você espera encontrar um grande efeito, então a MANOVA terá um grande poder se as medidas são de alguma forma diferentes (mesmo que negativamente correlacionadas) e se as diferenças entre os grupos estão na mesma direção para cada medida. Se você tiver duas variáveis dependentes, uma exibindo uma grande diferença de grupo e outra com uma pequena ou sem diferença

de grupo, então o poder aumentará se essas variáveis estiverem altamente correlacionadas. O que o trabalho de Cole e colaboradores nos diz é que se você está interessado em quão poderosa a MANOVA poderá ser, deve considerar não apenas as intercorrelações entre as variáveis dependentes mas também o tamanho e o padrão das diferenças entre os grupos que você espera obter. Contudo, lembre-se de que o trabalho de Cole e colaboradores está limitado a comparações entre dois grupos; as considerações sobre o poder do teste são mais complexas em situações de grupos múltiplos.

14.2.3 O exemplo do capítulo ②

Ao longo deste capítulo, utilizaremos um único exemplo para ilustrar como a MANOVA funciona e como realizá-la no SPSS. Imagine que estamos interessados nos efeitos da TCC (Terapia Cognitivo-Comportamental) sobre o TOC (Transtorno Obsessivo-Compulsivo). O TOC é uma doença caracterizada por ideias ou comportamentos que o próprio doente acha detestáveis (no meu caso, é o pensamento de alguém executando um teste *t* em dados que não são normalmente distribuídos ou a ideia da morte dos pais). Isso leva o doente a executar atividades para neutralizar esses pensamentos desagradáveis (essas atividades podem ser mentais, como executar uma MANOVA na minha cabeça para me fazer sentir melhor sobre o pensamento do teste *t*, ou físico, como tocar o chão 23 vezes para que os pais não morram). Podemos comparar um grupo de pacientes com TOC após o tratamento com TCC e após uma TC (Terapia Comportamental) com um grupo de pacientes com TOC que ainda estão esperando por tratamento (uma condição de ST – Sem Tratamento).¹ Muitas psicopatologias tem tanto elementos compor-

¹ A TC (Terapia Comportamental) é baseada na ideia de que se você parar os comportamentos mal-adaptados a desordem desaparecerá, enquanto a TCC (Terapia Cognitiva Comportamental) é baseada na ideia de .que tratar as cognições mal-adaptadas eliminará o transtorno. Se você quiser saber mais sobre o assunto, recomendo meu livro sobre psicologia clínica (Field, 2003), que está disponível nas seções de “enchufados” de qualquer livraria.

tamentais quanto cognitivos. Por exemplo, no TOC, se alguém tem uma obsessão por contaminação por germes, esse transtorno pode se manifestar como uma lavagem de mãos obsessiva e irá influenciar não apenas quantas vezes a pessoa irá lavar as mãos (comportamento), mas também o número de vezes que ela pensa em lavar as mãos (cognição). Da mesma forma, alguém obcecado por bolsas não apenas irá pensar muito sobre elas, mas também apresentar comportamentos relacionados (como dizer “bolsa” várias vezes ou comprar qualquer bolsa que ver pela frente). Se você estiver interessado em verificar quão bem-sucedida é a terapia, não é suficiente observar apenas expressões do comportamento (como se o comportamento obsessivo foi reduzido); também é importante determinar se as cognições estão sendo alteradas. Por isso, nesse exemplo, duas variáveis dependentes foram avaliadas: a ocorrência de comportamentos obsessivos (**actions**) e a ocorrência de obsessões cognitivas (**thoughts**). Essas variáveis dependentes foram mensuradas em um único dia e, assim,

representam o número de pensamentos/comportamentos obsessivos em um dia normal.

Os dados estão na Tabela 14.1 e podem ser encontrados no arquivo **OCD.sav**.^{*} Os participantes pertencem a um dos seguintes grupos: um (TCC), dois (TC) ou três (ST), e dentro dos grupos todos os participantes tiveram tanto as ações quanto os pensamentos mensurados.

14.3 TEORIA DA MANOVA ③



É complexo entender a teoria da MANOVA sem um bom conhecimento de álgebra matricial, mas a álgebra matricial está além do escopo deste livro (os interessados podem consultar Namboodiri, 1984; ou Stevens, 1992). Contudo, vou apresentar as bases conceituais da MA-

^{*} N. de T.: Se o arquivo **OCD.sav** for utilizado, os grupos e termos acima mencionados aparecerão em Inglês. Nesse caso, OCD (*Obsessive compulsive disorder*), CBT (*Cognitive Behaviour Therapy*), BT (*Behaviour Therapy*) e NT (*No-Treatment*).

Tabela 14.1 Dados do arquivo OCD.sav

Grupo	VD 1: Ações			VD 2: Pensamentos		
	TCC (1)	TC (2)	ST (3)	TCC (1)	TC (2)	ST (3)
	5	4	4	14	14	13
	5	4	5	11	15	15
	4	1	5	16	13	14
	4	1	4	13	14	14
	5	4	6	12	15	13
	3	6	4	14	19	20
	7	5	7	12	13	13
	6	5	4	15	18	16
	6	2	6	16	14	14
	4	5	5	11	17	18
\bar{X}	4,90	3,70	5,00	13,40	15,20	15,00
s	1,20	1,77	1,05	1,90	2,10	2,36
s^2	1,43	3,12	1,11	3,60	4,40	5,56
$\bar{X}_{\text{Geral(Ações)}} = 4,53$				$\bar{X}_{\text{Geral(Pensamentos)}} = 14,53$		
$S^2_{\text{Geral(Ações)}} = 2,1195$				$S^2_{\text{Geral(Pensamentos)}} = 4,8780$		

(*) VD = Variável Dependente

NOVA utilizando matrizes, sem exigir que você entenda exatamente como essas matrizes são utilizadas. Quem estiver interessado na teoria exata por trás da MANOVA deve ler Bray e Maxwell (1985), que é um ótimo texto sobre o assunto.

14.3.1 Introdução às matrizes ③

Primeiro, preciso explicar o que é uma matriz: uma matriz é simplesmente uma coleção de números arranjados em colunas e linhas. De fato, elas foram utilizadas ao longo deste livro sem que vocês se dessem conta: o editor de dados do SPSS. No editor de dados do SPSS, temos números arranjados em colunas e linhas e isso é exatamente o que é uma matriz. Uma matriz pode ter muitas colunas e linhas, e geralmente especificamos as dimensões de uma matriz utilizando números. Assim, uma matriz 2×3 é um arranjo de números com duas linhas e três colunas, e uma matriz 5×4 é uma com 5 linhas e 4 colunas. Por exemplo:

$$\begin{pmatrix} 2 & 5 & 6 \\ 3 & 5 & 8 \end{pmatrix} \quad \begin{pmatrix} 2 & 4 & 6 & 8 \\ 3 & 4 & 6 & 7 \\ 4 & 3 & 5 & 8 \\ 2 & 5 & 7 & 9 \\ 4 & 6 & 6 & 9 \end{pmatrix}$$

matriz 2×3

matriz 5×4

Você pode pensar nessas matrizes em termos de cada linha representando os dados de um único participante e cada coluna como sendo os dados de uma determinada variável. Assim, para uma matriz 5×4 podemos imaginar uma situação onde cinco participantes foram testados em quatro variáveis: desse modo, o primeiro participante teve um escore de 2 na primeira variável e 8 na quarta variável. Os valores no interior de uma matriz são referidos como *componentes* ou *elementos*.

Em uma **matriz quadrada**, o número de linhas é igual ao número de colunas. Nesse tipo de matriz é, às vezes, útil distinguir os componentes da diagonal (isto é, os valores que estão na linha diagonal a partir do elemento do canto superior esquerdo ao componente do canto inferior direito) dos componentes

fora da diagonal (valores que não estão na diagonal). Na matriz abaixo, os elementos da diagonal são 5, 12, 2 e 6, porque eles estão na linha diagonal. Os componentes fora da diagonal são todos os demais. Uma matriz quadrada na qual os elementos da diagonal são iguais a 1 e os fora da diagonal são iguais a 0 é conhecida como matriz identidade.

$$\begin{pmatrix} 5 & 3 & 6 & 10 \\ 3 & 12 & 4 & 6 \\ 6 & 4 & 2 & 7 \\ 10 & 6 & 7 & 6 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

matriz quadrada

matriz identidade

Espero que o conceito de matriz seja agora menos amedrontador do que parecia no começo: ela não é uma entidade matemática mágica, apenas uma maneira de representar um conjunto de dados – da mesma forma que uma planilha.

Existe um caso especial de uma matriz em que temos dados de somente uma pessoa; isso é conhecido como uma matriz linha ou *vetor linha*. Da mesma forma, se existe apenas uma coluna em uma matriz, temos um *vetor coluna*. Nos exemplos abaixo, o vetor linha pode ser pensado como os dados de uma única pessoa em quatro variáveis diferentes, e o vetor coluna pode ser pensado como os dados de cinco participantes avaliados em uma única variável:

$$\begin{pmatrix} 2 & 6 & 4 & 8 \end{pmatrix} \quad \begin{pmatrix} 8 \\ 6 \\ 10 \\ 15 \\ 6 \end{pmatrix}$$

vetor linha

vetor coluna

Com esse novo conhecimento sobre vetores, agora podemos dar uma breve olhada em como eles são utilizados para realizar a MANOVA.

14.3.2 Algumas matrizes importantes e suas funções ③

Assim como na ANOVA, estamos primeiramente interessados em quanta variância pode ser explicada pela manipulação experi-

mental (que em termos reais significa quanta variância é explicada pelo fato de que certos escores aparecem em certos grupos). Portanto, precisamos saber a soma dos quadrados devido à variável de agrupamento (a variação sistemática, SS_M), a soma dos quadrados devido a diferenças naturais entre os participantes (a variação residual, SS_R) e, é claro, a quantidade total de variação precisa ser explicada (SS_T). (Para mais detalhes sobre essas fontes de variação, releia os Capítulos 5 e 8.) Contudo, mencionei que a MANOVA também leva em conta diversas variáveis dependentes simultaneamente, e faz isso utilizando uma matriz que contém informações sobre as variâncias de responsabilidade de cada variável dependente.

Para o teste F univariado (por exemplo, a ANOVA), calculamos a razão da variância sistemática pela não-sistemática para uma única variável dependente. Na MANOVA, a estatística teste é derivada comparando a razão das variâncias sistemáticas pelas não-sistemáticas para diversas variáveis dependentes. Essa comparação é feita utilizando a razão de uma matriz representando as variâncias sistemáticas de todas as variáveis dependentes para uma matriz representando as variâncias não-sistemáticas de todas as variáveis dependentes. A matriz que representa as variâncias sistemáticas (ou a soma dos quadrados do modelo para todas as variáveis) é representada pela letra **H** e é denominada **hipótese de matriz produto cruzado e soma dos quadrados** (ou **hipótese SSCP**). A matriz que representa a variância não-sistemática (a soma dos quadrados residual para todas as variáveis) é representada pela letra **E**, e é denominada **matriz produto cruzado e erro da soma dos quadrados** (ou erro SSCP). Finalmente, existe uma matriz que representa a quantidade total de variância presente para cada variável dependente (a soma dos quadrados total para cada variável dependente), representada por **T** e denominada **matriz dos produtos cruzados e soma dos quadrados total** (ou **total SSCP**).

Mais tarde, mostrarei que essas matrizes são usadas exatamente da mesma forma que as

simples somas dos quadrados (SS_M , SS_R e SS_T) na ANOVA para derivar a estatística representando a razão das variâncias sistemáticas para as não-sistemáticas no modelo. Alguns de vocês devem ter notado que as matrizes que descrevi são todas denominadas matrizes somas dos quadrados e produto cruzado (SSCP). Deve ser óbvio o motivo de essas matrizes serem chamadas de soma dos quadrados, mas por que existe uma referência ao produto cruzado nos seus nomes? Aprendemos um pouco sobre desvios produto cruzado no Capítulo 4 e vimos que eles representavam o valor total do erro combinado entre duas variáveis (assim, de certa forma eles representavam uma estimativa padronizada da correlação total entre duas variáveis). Enquanto a soma dos quadrados de uma variável é a soma dos quadrados das diferenças entre os valores observados e a média dessa variável, o produto cruzado é o erro total combinado entre duas variáveis. Mencionei antes que a MANOVA tem o poder de levar em conta qualquer correlação entre as variáveis dependentes; ela faz isso utilizando esses produtos cruzados.

14.3.3 Calculando a MANOVA manualmente: um exemplo ③

Para começar, vamos executar ANOVAs univariadas em cada uma das duas variáveis dependentes no nosso exemplo TOC (veja a Tabela 14.1). Uma descrição do modelo da ANOVA pode ser encontrada no Capítulo 8 – eu vou me basear na hipótese de que você tenha lido esse capítulo, portanto, se você ainda não entendeu os detalhes do Capítulo 8 agora é uma boa hora para (re)ler as Seções 8.2.4 a 8.2.8.

14.3.3.1 ANOVA univariada para VD 1 (Ações) ②

Existem três somas de quadrados que precisam ser calculadas. Primeiro, precisamos avaliar quanta variabilidade existe dentro dos dados (SS_T) a ser explicada, depois precisamos ver quanto dessa variabilidade pode ser explicado pelo modelo (SS_M) e, finalmente, avaliar quanto erro existe no nosso modelo

(SS_R). No Capítulo 8, calculamos cada um desses valores:

- $SS_{T(Ações)}$: A soma total dos quadrados é obtida calculando as diferenças entre cada um dos 20 escores e a média desses escores, então elevando ao quadrado essas diferenças e somando todos os quadrados obtidos. Alternativamente, você pode deixar que o SPSS calcule a variância para os dados da ação (**action**) (sem levar em conta o grupo a que pertencem) e então multiplicar esse valor pelo número de escores menos 1:

$$\begin{aligned} SS_T &= s_{\text{Geral}}^2(n - 1) \\ &= 2,1195(30 - 1) \\ &= 2,1195 \times 29 \\ &= 61,47 \end{aligned}$$

- $SS_{M(Ações)}$: Esse valor é calculado tomando as diferenças das médias de cada grupo em relação à média geral e então elevando essas diferenças ao quadrado. Depois, esses valores são multiplicados pelo número de escores do grupo e então os resultados são somados:

$$\begin{aligned} SS_T &= 10(4,90 - 4,53)^2 + 10(3,70 - 4,53)^2 + 10(5,00 - 4,53)^2 \\ &= 10(0,37)^2 + 10(-0,83)^2 + 10(0,47)^2 \\ &= 1,37 + 6,89 + 2,21 \\ &= 10,47 \end{aligned}$$

- $SS_{R(Ações)}$: Esse valor é calculado tomando as diferenças entre cada escore e a média do grupo a qual o escore pertence. Essas diferenças são então elevadas ao quadrado e somadas entre si. Alternativamente, você pode solicitar que o SPSS calcule a variância dentro de cada grupo, multiplique a variância de cada grupo pelo número de escores menos um e então some os resultados entre si:

$$\begin{aligned} SS_R &= s_{\text{TCC}}^2(n_{\text{TCC}} - 1) + s_{\text{TC}}^2(n_{\text{TC}} - 1) \\ &\quad + s_{\text{ST}}^2(n_{\text{ST}} - 1) \\ &= (1,433)(10 - 1) + (3,122)(10 - 1) \\ &\quad + (1,111)(10 - 1) \end{aligned}$$

$$\begin{aligned} &= (1,433 \times 9) + (3,122 \times 9) \\ &\quad + (1,111 \times 9) \\ &= 12,9 + 28,1 + 10,0 \\ &= 51,0 \end{aligned}$$

O próximo passo é calcular a média das somas dos quadrados (o quadrado médio) dividindo o resultado obtido pelo número de graus de liberdade correspondente (Seção 8.2.7):

SS	gl	MS
$SS_{M(Ações)} = 10,47$	2	5,235
$SS_{R(Ações)} = 51,00$	27	1,889

O último estágio é calcular F pela divisão do quadrado médio do modelo pela média dos quadrados do erro do modelo:

$$F = \frac{MS_M}{MS_R} = \frac{5,235}{1,889} = 2,771$$

Esse valor pode então ser confrontado com o valor crítico da F . Você deve observar aqui o cálculo das várias somas dos quadrados e com que cada uma se relaciona.

14.3.3.2 ANOVA univariada para VD 2 (pensamentos) ②

Os cálculos aqui são realizados da mesma forma como ocorreram com a variável dependente (VD) 1.

$SS_{T(\text{Pensamentos})}$:

$$\begin{aligned} SS_T &= s_{\text{Geral}}^2(n - 1) \\ &= 4,878(30 - 1) \\ &= 4,878 \times 29 \\ &= 141,46 \end{aligned}$$

$SS_{M(\text{Pensamentos})}$:

$$\begin{aligned} SS_M &= 10(13,40 - 14,53)^2 + 10(15,20 - 14,53)^2 + 10(15,00 - 14,53)^2 \\ &= 10(-1,13)^2 + 10(0,67)^2 + 10(0,47)^2 \\ &= 12,77 + 4,49 + 2,21 \\ &= 19,47 \end{aligned}$$

$SS_{T(\text{Pensamentos})}$:

$$\begin{aligned} SS_R &= s_{TCC}^2(n_{TCC} - 1) + s_{TC}^2(n_{TC} - 1) \\ &\quad + s_{ST}^2(n_{ST} - 1) \\ &= 3,6(10 - 1) + 4,4(10 - 1) \\ &\quad + 5,56(10 - 1) \\ &= (3,6 \times 9) + (4,4 \times 9) \\ &\quad + (5,56 \times 9) \\ &= 32,4 + 39,6 + 50,0 \\ &= 122,0 \end{aligned}$$

O próximo passo é calcular a média das somas dos quadrados (o quadrado médio) dividindo o resultado obtido pelo número de graus de liberdade correspondente (Seção 8.2.7):

SS	gl	MS
$SS_{M(\text{Thoughts})} = 19,47$	2	9,735
$SS_{R(\text{Thoughts})} = 122,00$	27	4,519

O último estágio é calcular F pela divisão do quadrado médio do modelo pela média dos quadrados do erro do modelo:

$$F = \frac{MS_M}{MS_R} = \frac{9,735}{4,519} = 2,154$$

Esse valor pode então ser confrontado com o valor crítico da F. Novamente, o que deve ser notado aqui é o cálculo das várias somas dos quadrados e com o que cada uma se relaciona.

14.3.3.3 O relacionamento entre VDs: produtos cruzados ②

Já sabemos que a MANOVA utiliza a soma dos quadrados como uma ANOVA e na próxima seção, veremos exatamente como ela usa esses valores. Contudo, também mencionei que a MANOVA leva em conta o relacionamento entre as variáveis dependentes e que ela faz isso utilizando o produto cruzado. Para ser preciso, existem três produtos cruzados diferentes que são de interesse e esses três produtos cruzados se relacionam com as três somas dos quadrados que calculamos para as ANOVAs univariadas: isto é, existe o produ-

to cruzado total, o produto cruzado devido ao modelo e o produto cruzado dos resíduos. Vamos dar uma olhada no produto cruzado total (CP_T) primeiro.

Mencionei no Capítulo 4 que o produto cruzado é a diferença entre os escores e a média em um grupo multiplicado pela diferença entre os escores e a média de outro grupo. No caso do produto cruzado total, a média de interesse é a média geral de cada variável dependente (veja a Tabela 14.2). Consequentemente, podemos adaptar a equação do produto cruzado descrita no Capítulo 4 utilizando as duas variáveis dependentes. A equação resultante para o produto cruzado total está apresentada na equação (14.1). Portanto, para cada variável dependente, tomamos cada escore e subtraímos a média da variável. Isso deixa você com dois valores por participante (um para cada variável dependente) que devem ser multiplicados entre si para obter o produto cruzado de cada participante. O total pode ser obtido adicionando os produtos cruzados de todos os participantes. A Tabela 14.2 ilustra esse processo.

$$CP_T = \sum (x_{i(Ações)} - \bar{X}_{Geral(Ações)}) (x_{i(Pensamentos)} - \bar{X}_{Geral(Pensamentos)}) \quad (14.1)$$

O produto cruzado total é uma avaliação do relacionamento global entre as duas variáveis. Contudo, estamos, também, interessados em como o relacionamento entre as variáveis dependentes é influenciado pela manipulação experimental e esse relacionamento é mensurado pelo produto cruzado do modelo (CP_M). O CP_M é calculado de modo semelhante à soma dos quadrados do modelo. Primeiro, a diferença entre a média de cada grupo e a média geral é calculada para cada variável dependente. O produto cruzado é calculado multiplicando-se as diferenças encontradas para cada grupo. Cada produto é então multiplicado pelo número de escores dentro do grupo (como foi feito com a soma dos quadrados). Esse princípio é ilustrado na equação (14.2) e na Tabela 14.3:

$$CP_M = \sum n[(\bar{x}_{Grupo(Ações)} - \bar{X}_{Geral(Ações)}) (\bar{X}_{Grupo(Pensamentos)} - \bar{X}_{Geral(Pensamentos)})] \quad (14.2)$$

Tabela 14.1 Cálculos para obter o Produto Cruzado Total (CP_T)

Grupo	Actions (Ações)	Thoughts (Pensamentos)	Actions — $\bar{X}_{\text{Geral(Actions)}}$ (D ₁)	Thoughts — $\bar{X}_{\text{Geral(Thoughts)}}$ (D ₂)	D ₁ × D ₂
TCC	5	14	0,47	−0,53	−0,25
	5	11	0,47	−3,53	−1,66
	4	16	−0,53	1,47	−0,78
	4	13	−0,53	−1,53	0,81
	5	12	0,47	−2,53	−1,19
	3	14	−1,53	−0,53	0,81
	7	12	2,47	−2,53	−6,25
	6	15	1,47	0,47	0,69
	6	16	1,47	1,47	2,16
	4	11	−0,53	−3,53	1,87
TC	4	14	−0,53	−0,53	0,28
	4	15	−0,53	0,47	−0,25
	1	13	−3,53	−1,53	5,40
	1	14	−3,53	−0,53	1,87
	4	15	−0,53	0,47	−0,25
	6	19	1,47	4,47	6,57
	5	13	0,47	−1,53	−0,72
	5	18	0,47	3,47	1,63
	2	14	−2,53	−0,53	1,34
	5	17	0,47	2,47	1,16
ST	4	13	−0,53	−1,53	0,81
	5	15	0,47	0,47	0,22
	5	14	0,47	−0,53	−0,25
	4	14	−0,53	−0,53	0,28
	6	13	1,47	−1,53	−2,25
	4	20	−0,53	5,47	−2,90
	7	13	2,47	−1,53	−3,78
	4	16	−0,53	1,47	−0,78
	6	14	1,47	−0,53	−0,78
	5	18	0,47	3,47	1,63
\bar{X}_{Geral}	4,53	14,53			CP _T = $\sum(D_1 \times D_2) = 5,47$

Tabela 14.3 Cálculos para obter o Produto Cruzado do Modelo (CP_M)

Grupo	\bar{X}_{Grupo} Actions	$\bar{X}_{\text{Grupo}} - \bar{X}_{\text{Geral}}$ (D ₁)	\bar{X}_{Grupo} Thoughts	$\bar{X}_{\text{Grupo}} - \bar{X}_{\text{Geral}}$ (D ₂)	D ₁ × D ₂	N(D ₁ × D ₂)
TCC	4,9	0,37	13,4	−1,13	−0,418	−4,18
TC	3,7	−0,83	15,2	0,67	−0,556	−5,56
ST	5,0	0,47	15,0	0,47	0,221	2,21
\bar{X}_{Geral}	4,53		14,53		CP _M = $\sum N(D_1 \times D_2) = -7,53$	

Finalmente, precisamos, ainda, saber como o relacionamento entre as duas variáveis dependentes é influenciado pelas diferenças individuais do desempenho dos participantes. O produto cruzado residual (CP_R) nos informa sobre como o relacionamento entre as variáveis dependentes é afetado pelas diferenças individuais, ou erro no modelo. O CP_R é calculado de modo similar ao produto cruzado total exceto que a média dos grupos é utilizada em vez da média geral (veja a equação (14.3)). Assim, para calcular a diferença de cada escore, tomamos cada escore e subtraímos da média do grupo ao qual ele pertence (veja a Tabela 14.4):

$$CP_R = \sum (x_{i(Ações)} - \bar{X}_{Grupo(Ações)})(x_{i(Pensamentos)} - \bar{X}_{Grupo(Pensamentos)}) \quad (14.3)$$

Os observadores entre vocês irão notar que o produto cruzado dos resíduos pode também ser calculado pela subtração do produto cruzado do modelo do produto cruzado total:

$$\begin{aligned} CP_R &= CP_T - CP_M \\ &= 5,47 - (-7,53) = 13 \end{aligned}$$

Entretanto, é útil calcular o produto cruzado residual manualmente para o caso de algum erro no cálculo dos outros dois produtos cruzados. O fato que a soma dos produtos cruzados do modelo e dos resíduos deve fornecer o produto cruzado total pode ser utilizado como uma forma útil de contraprova.

Cada um dos diferentes produtos cruzados nos informa algo importante sobre o relacionamento entre as duas variáveis dependentes. Embora eu tenha utilizado um cenário simples para manter os cálculos relativamente simples, esses princípios podem ser facilmente estendidos para cenários mais complexos. Por exemplo, se tivéssemos mensurado três variáveis dependentes, então os produtos cruzados entre os pares de variáveis dependentes seriam calculados (como foram nesse exemplo) e colocados na matriz SSCP apropriada (veja a próxima seção). À medida que a complexidade da situação aumenta, também

cresce a quantidade de cálculos que precisam ser realizados. Nessas situações, a vantagem de um *software* como o SPSS se torna mais aparente!

14.3.3.4 A matriz SSCP total (T) ③

Nesse exemplo, temos somente duas variáveis dependentes e, assim, todas as matrizes SSCP serão matrizes do tipo 2x2. Se existissem três variáveis dependentes, as matrizes resultantes teriam dimensão 3x3. A matriz SSCP total, ou T, contém a soma dos totais dos quadrados para cada variável dependente e o produto cruzado total entre as duas variáveis dependentes. Você pode imaginar a primeira coluna e a primeira linha como representando uma variável dependente e a segunda coluna e linha como representando a segunda variável dependente:

	Coluna 1 <i>Actions</i> (Ações)	Coluna 2 <i>Thoughts</i> (Pensamentos)
Linha 1 <i>Actions</i> (Ações)	$SS_{T(Ações)}$	CP_T
Linha 2 <i>Thoughts</i> (Pensamentos)	CP_T	$SS_{T(Thoughts)}$

Já calculamos esses valores nas seções anteriores e assim podemos simplesmente colocá-los nas células adequadas da matriz:

$$T = \begin{pmatrix} 61,47 & 5,47 \\ 5,47 & 141,47 \end{pmatrix}$$

A partir dos valores da matriz (e do que eles representam) deve estar claro que a SSCP total representa tanto a quantidade de variação que existe dentro dos dados e a codependência total que existe entre as variáveis dependentes. Você deve notar ainda que os elementos fora da diagonal são iguais (eles são o produto cruzado total) porque esse valor é igualmente importante para as duas variáveis dependentes.

Tabela 14.4 Cálculos para obter o Produto Cruzado Residual (CP_R)

Grupo	Actions (Ações)	Actions — $\bar{X}_{\text{Grupo(Actions)}}$ (D_1)	Thoughts (Pensamentos)	Thoughts — $\bar{X}_{\text{Grupo(Thoughts)}}$ (D_2)	$D_1 \times D_2$
TCC	5	0,10	14	0,60	0,06
	5	0,10	11	-2,40	-0,24
	4	-0,90	16	2,60	-2,34
	4	-0,90	13	-0,40	0,36
	5	0,10	12	-1,40	-0,14
	3	-1,90	14	0,60	-1,14
	7	2,10	12	-1,40	-2,94
	6	1,10	15	1,60	1,76
	6	1,10	16	2,60	2,86
	4	-0,90	11	-2,40	2,16
\bar{X}_{TCC}	4,9		13,4		$\Sigma = 0,40$
TC	4	0,30	14	-1,20	-0,36
	4	0,30	15	-0,20	-0,06
	1	-2,70	13	-2,20	5,94
	1	-2,70	14	-1,20	3,24
	4	0,30	15	-0,20	0,06
	6	2,30	19	3,80	8,74
	5	1,30	13	-2,20	-2,86
	5	1,30	18	2,80	3,64
	2	-1,70	14	-1,20	2,04
	5	1,30	17	1,80	2,34
\bar{X}_{TC}	3,7		15,2		$\Sigma = 22,60$
ST	4	-1,00	13	-2,00	2,00
	5	0,00	15	0,00	0,00
	5	0,00	14	-1,00	0,00
	4	-1,00	14	-0,00	1,00
	6	1,00	13	-2,00	-2,00
	4	-1,00	20	5,00	-5,00
	7	2,00	13	-2,00	-4,00
	4	-1,00	16	1,00	-1,00
	6	1,00	14	-1,00	-1,00
	5	0,00	18	3,00	0,00
\bar{X}_{ST}	5		15		$\Sigma = -10,00$
					$CP_R = \Sigma(D_1 \times D_2) = 13$

14.3.3.5 A matriz SSCP Residual (E) ③

A matriz SSCP residual (ou erro), E, contém a soma dos quadrados dos resíduos para cada variável dependente e o produto cruza-

do residual entre as duas variáveis dependentes. Essa matriz SSCP é semelhante à matriz SSCP total, exceto que a informação está relacionada aos erros do modelo:

	<i>Coluna 1</i> <i>Actions (Ações)</i>	<i>Coluna 2</i> <i>Thoughts</i> (Pensamentos)
<i>Linha 1</i> <i>Actions (Ações)</i>	SS _{R(Actions)}	CP _R
<i>Linha 2</i> <i>Thoughts</i> (Pensamentos)	CP _R	SS _{R(Thoughts)}

Já calculamos esses valores nas seções anteriores e, assim, podemos simplesmente colocá-los nas células adequadas da matriz:

$$E = \begin{pmatrix} 51 & 13 \\ 13 & 122 \end{pmatrix}$$

A partir dos valores da matriz (e do que eles representam), deve estar claro que a SSCP residual representa a quantidade de variação não-sistemática que existe em cada variável dependente e a codependência que existe entre variáveis dependentes devida apenas a fatores casuais. Como antes, os elementos fora da diagonal são iguais (ambos são produtos cruzados residuais).

14.3.3.6 A matriz SSCP do modelo (H) ③

A matriz SSCP do modelo (ou hipótese), H, contém a soma dos quadrados do modelo para cada variável dependente e o produto cruzado residual entre as duas variáveis dependentes:

	<i>Coluna 1</i> <i>Actions (Ações)</i>	<i>Coluna 2</i> <i>Thoughts</i> (Pensamentos)
<i>Linha 1</i> <i>Actions (Ações)</i>	SS _{M(Actions)}	CP _M
<i>Linha 2</i> <i>Thoughts</i> (Pensamentos)	CP _M	SS _{M(Thoughts)}

Esses valores foram calculados nas seções anteriores e assim podemos simplesmente colocar os números apropriados nas células adequadas da matriz (veja abaixo). A partir desses valores na matriz (e o que eles representam), deve estar claro que a SSCP do modelo representa tanto a variação sistemática que existe

para cada variável dependente quanto a codependência entre as variáveis dependentes devido ao modelo (isto é, devido a manipulação experimental). Como antes, os elementos fora da diagonal são os mesmos (ambos representam o produto cruzado do modelo):

$$H = \begin{pmatrix} 10,47 & -7,53 \\ -7,53 & 19,47 \end{pmatrix}$$

Matrizes são aditivas, o que significa que podemos somar (ou subtrair) duas matrizes pela adição (ou subtração) dos elementos correspondentes. Quando calculamos a ANOVA univariada, vimos que a soma total dos quadrados era a soma dos quadrados do modelo com a soma dos quadrados dos resíduos (isto é, $SS_T = SS_M + SS_R$). O mesmo é verdadeiro para a MANOVA, exceto que estamos adicionando matrizes e não simplesmente valores:

$$\begin{aligned} T &= H + E \\ T &= \begin{pmatrix} 10,47 & -7,53 \\ -7,53 & 19,47 \end{pmatrix} + \begin{pmatrix} 51 & 13 \\ 13 & 122 \end{pmatrix} \\ &= \begin{pmatrix} 10,47 + 51 & -7,53 + 13 \\ -7,53 + 13 & 19,47 + 122 \end{pmatrix} \\ &= \begin{pmatrix} 61,47 & 5,47 \\ 5,47 & 141,47 \end{pmatrix} \end{aligned}$$

A demonstração de que essas matrizes podem ser somadas ajuda você a entender (eu espero!) que os cálculos da MANOVA são conceitualmente os mesmos da ANOVA univariada, com a diferença sendo que são utilizadas matrizes em vez de valores simples.

14.3.4 Princípios do teste estatístico MANOVA ④

Na ANOVA univariada, calculamos a razão da variância sistemática pela não-sistemática (isto é, dividimos SS_M por SS_R).² O equivalente conceitual será, portanto, dividir a matriz H pela matriz E. Existe, entretanto,

² Na realidade, utilizamos a média dos quadrados, mas esses valores são meramente as somas dos quadrados corrigidos para os graus de liberdade.

o problema de que matrizes não são divisíveis por outras matrizes! Contudo, existe uma operação matricial equivalente à divisão, que é multiplicar uma das matrizes pelo inverso da outra. Assim, se queremos dividir a matriz H pela E , temos que multiplicar a matriz H pela matriz inversa da E (representada por E^{-1}). Assim, a estatística teste é baseada na matriz que resulta da multiplicação da SSCP do modelo pela inversa da SSCP residual. Essa matriz é representada por HE^{-1} .

Calcular a inversa de uma matriz é uma operação difícil e trabalhosa e você não precisa entender como isso é feito, pois o SPSS faz automaticamente. Contudo, os leitores interessados podem consultar Stevens (1992) ou Namboodiri (1984) – esses textos fornecem material bastante acessível sobre como calcular a inversa de uma matriz. Para os leitores que irão consultar essas fontes, os cálculos desse exemplo estão incluídos no arquivo do *site* www.artmed.com.br **Appendix Chapter 14.pdf** (você pode querer verificar os meus cálculos!). Já os leitores não interessados devem acreditar que:

$$E^{-1} = \begin{pmatrix} 0,0202 & -0,0021 \\ -0,0021 & 0,0084 \end{pmatrix}$$

$$HE^{-1} = \begin{pmatrix} 0,2273 & -0,0852 \\ -0,1930 & 0,1794 \end{pmatrix}$$

Lembre que HE^{-1} representa a razão entre a variância sistemática do modelo para a variância não-sistemática e, assim, a matriz resultante é conceitualmente a mesma razão F da ANOVA univariada. Esse é outro problema, também. Na ANOVA, quando dividimos a variância sistemática pela não-sistemática, obtemos um único número: a razão F . Na MANOVA, quando dividimos a variância sistemática pela não-sistemática obtemos uma matriz contendo vários valores. Nesse exemplo, a matriz contém quatro valores, mas se tivéssemos três variáveis dependentes a matriz teria nove valores. De fato, a matriz resultante sempre irá conter p^2 valores, onde p é o número de variáveis dependentes. O problema é como converter esses valores da matriz em um único

valor que tenha sentido. Esse é o ponto onde teremos que abandonar qualquer esperança de entender a matemática por trás do teste e falar apenas conceitualmente.

14.3.4.1 A função discriminante ④

O problema de ter vários valores para determinar a significância estatística pode ser simplificado pela conversão das variáveis dependentes em dimensões subjacentes ou fatores (esse processo será discutido com mais detalhes no Capítulo 15). No Capítulo 5, vimos como a regressão múltipla funciona ajustando um modelo linear ao conjunto de dados para prever uma variável de saída (a variável dependente na terminologia da ANOVA). Esse modelo linear é construído por uma combinação das variáveis previsoras (ou variáveis independentes), onde cada uma tem uma contribuição única para o modelo linear. Podemos fazer algo similar aqui, exceto que estamos interessados no problema oposto (isto é, prever uma variável independente a partir de um conjunto de variáveis dependentes). Assim, é possível calcular as dimensões subjacentes das variáveis dependentes. Essas combinações lineares das variáveis dependentes são conhecidas como *variates* (ou às vezes denominadas *variáveis latentes* ou *fatores*). Nesse contexto, queremos utilizar essas combinações lineares de variáveis (*variates*) para prever a que grupo uma pessoa pertence (isto é, se ele está no grupo TCC, TC ou ST), assim, nós as estamos utilizando para discriminar grupos de pessoas. Portanto, essas *variates* são denominadas *funções discriminantes*. Embora eu tenha feito um paralelo entre essas funções discriminantes e o modelo de regressão múltipla, existe uma diferença no sentido de que podemos extrair várias funções discriminantes de um conjunto de variáveis dependentes, enquanto na regressão múltipla todas as variáveis independentes estão incluídas em um único modelo.

Essa é a teoria em termos simplistas, mas como descobrimos essas funções discriminantes? Sem entrar em muitos detalhes, utilizamos um procedimento matemático de

maximização, de modo que a primeira função discriminante (V_1) seja a combinação linear das variáveis dependentes que maximizam as diferenças entre grupos.

Segue disso que a razão entre as variâncias sistemáticas e não-sistemáticas (SS_M/SS_R) será maximizada para a primeira *variate*, mas *variates* subsequentes terão valores menores dessa razão. Lembre que essa razão é um valor análogo do que a razão F representa na ANOVA univariada e, assim, obtemos o valor máximo possível da razão F quando olhamos para a primeira função discriminante. Essa *variate* pode ser descrita em termos de uma equação de regressão linear (porque ela é uma combinação linear das variáveis dependentes):

$$\begin{aligned} Y &= b_0 + b_1X_1 + b_2X_2 \\ V_1 &= b_0 + b_1VD_1 + b_2VD_2 \\ V_1 &= b_0 + b_1Ações + b_2Pensamentos \end{aligned} \quad (14.4)$$

A Equação (14.4) mostra as equações de regressão múltiplas para dois previsores e estende isso para mostrar como uma forma semelhante da equação pode ser descrita como funções discriminantes. Os valores b na equação são pesos (como na regressão) que nos informam algo sobre a contribuição de cada variável dependente para a combinação linear (*variate*) em questão. Na regressão, os valores de b são obtidos pelo método dos mínimos quadrados; na análise da função discriminante, os valores b são obtidos a partir dos auto-vetores (veja o Quadro 5.2) da matriz HE^{-1} . Podemos ignorar b_0 porque ele serve somente para localizar a *variate* no espaço geométrico, o que não é necessário quando estivermos utilizando-a para discriminar grupos.

Numa situação na qual existem apenas duas variáveis dependentes e dois grupos para a variável independente, existirá apenas uma combinação linear (*variate*). Isso torna o cenário bem simples: olhando para a função discriminante das variáveis dependentes, em vez de olhar para as próprias variáveis dependentes, podemos obter um único valor do SS_M/SS_R para a função discriminante e então avaliar a significância desse valor. Contudo, em

casos mais complexos quando existirem mais do que duas variáveis dependentes ou mais do que três níveis da variável independente (como é o caso do nosso exemplo), existirá mais do que uma combinação linear (CL) ou *variate*. O número de CL obtidos será menor do que p (o número de variáveis dependentes) ou $k - 1$ (onde k é o número de níveis da variável independente). No nosso exemplo, tanto p quanto $k - 1$ são iguais a 2 e, assim, devemos ser capazes de encontrar duas CLs (ou *variates*). Mencionei antes que os valores b que descrevem as CLs são obtidos pelo cálculo dos autovetores da matriz HE^{-1} e, de fato, existem dois autovetores derivados dessa matriz; um com os valores b para a primeira CL e outro com o b da segunda CL. Falando conceitualmente, os autovetores são vetores associados com uma dada matriz que não mudam quando a matriz é diagonalizada (veja o Quadro 5.2 para uma explicação visual de autovalores e autovetores). Uma matriz diagonal ou diagonalizada é simplesmente uma matriz em que os elementos fora da diagonal são todos iguais a zero, e transformando HE^{-1} em uma matriz diagonal, eliminamos todos os elementos fora da diagonal (reduzindo assim o número de valores para os quais devemos testar a significância. Entretanto, calculando os autovetores e autovalores, ainda estaremos com vários valores que representam a razão entre as variâncias sistemática e não-sistemática (porque eles não mudam pela diagonalização da matriz), mas agora existem menos valores. Os cálculos para obter os autovetores são complexos (estudantes insanos podem tentar a leitura de Strang, 1980 ou Nambodiri, 1984), assim, para a matriz HE^{-1} os autovetores são:

$$\begin{aligned} \text{Autovetor}_1 &= \begin{pmatrix} 0,603 \\ -0,335 \end{pmatrix} \\ \text{Autovetor}_2 &= \begin{pmatrix} 0,425 \\ 0,339 \end{pmatrix} \end{aligned}$$

Substituindo esses dois valores nas duas equações para as CLs e lembrando que podemos ignorar o valor de b_0 , obtemos os modelos apresentados na equação (14.5)

$$V_1 = 0,603Ações - 0,335Pensamentos \quad (14.5)$$

$$V_2 = 0,425Ações + 0,339Pensamentos$$

É possível utilizar as equações para cada CL a fim de calcular um escore para cada pessoa na CL. Por exemplo, o primeiro participante no grupo TCC apresentou 5 ações obsessivas e 14 pensamentos obsessivos. Portanto, seu escore na CL₁ será -1,675:

$$V_1 = (0,603 \times 5) + (0,335 \times 14) = -1,675$$

O escore para a segunda CL será 6,87:

$$V_2 = (0,425 \times 5) + (0,339 \times 14) = 6,871$$

Se calcularmos esses escores CL para cada participante e depois calcularmos as matrizes SSCP (por exemplo, H, E, T e HE⁻¹) que utilizamos previamente, encontraríamos todos eles como tendo produtos cruzados iguais a zero. Isso porque as CLs extraídas dos dados são ortogonais, o que significa que elas não são correlacionadas. Em resumo, as CLs extraídas são dimensões independentes construídas como uma combinação linear das variáveis dependentes que foram mensuradas.

Essa redução dos dados apresenta propriedades úteis no sentido de que se olharmos para a matriz HE⁻¹ calculada a partir dos escores CLs (em vez das variáveis dependentes), encontraremos que todos os elementos fora da diagonal (os produtos cruzados) serão zero. Os elementos da diagonal dessa matriz representam a razão da variância sistemática para a não-sistemática (isto é, SS_M/SS_T) para cada uma das CLs subjacentes. Assim, para os dados do nosso exemplo, isso significa que em vez de ter quatro valores representando a razão entre as variâncias sistemáticas e não-sistemáticas, teremos apenas dois. Essa redução pode não parecer muito. Contudo, em geral, se tivermos variáveis dependentes p, normalmente acabaríamos tendo valores p² representando a razão entre as variâncias sistemáticas e não-sistemáticas; pela utilização das funções discriminantes, reduzimos esse número para p. Se existirem quatro variáveis dependentes, acabaríamos com quatro valores

em vez de 16 (o que destaca as vantagens desse processo).

Para os dados do nosso exemplo, a matriz HE⁻¹ calculada a partir dos escores das CLs é:

$$HE_{CLs}^{-1} = \begin{pmatrix} 0,335 & 0,000 \\ 0,000 & 0,073 \end{pmatrix}$$

Fica claro a partir dessa matriz que temos dois valores para serem considerados quando avaliarmos a significância das diferenças entre grupos. Isso provavelmente pareceu um procedimento complicado para reduzir os dados dessa forma: contudo, ele mostra que os valores ao longo da diagonal da matriz para as CLs (ou seja, 0,335 e 0,073) são os *autovalores* da matriz HE⁻¹ original. Portanto, esses valores podem ser calculados diretamente dos dados coletados sem a necessidade de primeiro determinar os autovetores. Os cálculos desses autovalores estão incluídos nos apêndices, mas não é necessário que você entenda como eles são obtidos. Esses autovalores são conceitualmente equivalentes à razão F na ANOVA e, assim, o passo final é avaliar quão grande eles são quando comparados com o que se esperaria apenas pelo acaso. Existem quatro formas de avaliar esses valores.

14.3.4.2 Traço de Pillai-Bartlett (V) ④

O traço de Pillai-Bartlett (também conhecido como traço de Pillai) é dado na equação (14.6), na qual λ representa os autovalores para cada uma das funções discriminantes e s representa o número dessas funções. O traço de Pillai é a soma das proporções da variância explicada nas funções discriminantes. Como tal, ele é semelhante à razão SS_M/SS_T:

$$V = \sum_{i=1}^s \frac{\lambda_i}{1 + \lambda_i} \quad (14.6)$$

Para os nossos dados, esse traço é 0,319, que pode ser transformado em um valor que tem uma distribuição aproximada F.

$$V = \frac{0,335}{1 + 0,335} + \frac{0,073}{1 + 0,073} = 0,319$$

14.3.4.3 T^2 de Hotelling ④

O traço de Hotelling-Lawley ou T^2 de Hotelling (Figura 14.1) é simplesmente a soma dos autovalores para cada CL (veja a Equação (14.7)), e para esses dados o seu valor é 0,408 ($0,335 + 0,073$). Essa estatística teste é a soma da SS_M/SS_T para cada CL e, desse modo, ela se compara diretamente à razão F da ANOVA:

$$T = \sum_{i=1}^s \lambda_i \quad (14.7)$$

14.3.4.4 Lambda de Wilks (Λ) ④

O *lambda de Wilks* (Λ) é o produto da variância *não-explicada* em cada CL (veja a Equação (14.8)) – o símbolo Π é semelhante ao símbolo para a soma (Σ) que já encontramos, exceto que ele significa *multiplicar* em vez de somar). Assim, o lambda de Wilks representa a razão da variância do erro pela variância total (SS_R/SS_T) para cada CL:

$$\Lambda = \prod_{i=1}^s \frac{1}{1 + \lambda_i} \quad (14.8)$$

Para os dados, nesse exemplo, o valor é 0,698 e deve ficar claro que valores de autovalores grandes (que representam um grande efeito experimental) levam a pequenos valores do lambda de Wilks, portanto, existe significância estatística quando o lambda de Wilks for pequeno:

$$\Lambda = \left(\frac{1}{1 + 0,335} \right) \left(\frac{1}{1 + 0,073} \right) = 0,698$$

14.3.4.5 Maior raiz de Roy ④

A *maior raiz de Roy* sempre me faz pensar em um estatístico barbudo com uma pá de jardim arrancando uma enorme raiz de mandioca (ou outra semelhante); contudo, não é uma mandioca como o nome sugere, mas simplesmente o autovalor para a primeira CL. Assim, em certo sentido, ela é igual ao traço de Hotelling-Lawley, mas apenas para a primeira CL (veja a Equação (14.9)):

$$\text{Maior raiz} = \lambda_{\text{máximo}} \quad (14.9)$$



Figura 14.1 Harold Hotelling saboreando um chá, minha atividade favorita.



Como tal, a maior raiz de Roy representa a proporção da variância explicada pela não-explicada (SS_M/SS_R) para a primeira função discriminante.³ Para os dados nesse exemplo, o valor da maior raiz de Roy é simplesmente 0,335 (o autovalor para a primeira função CL). Assim, esse valor é conceitualmente equivalente à razão F na ANOVA univariada. Deve ser aparente, a partir do que aprendemos sobre as propriedades de maximização das funções discriminantes, que a raiz de Roy representa a diferença máxima possível entre grupos dos dados coletados. Contudo, em muitos casos essa estatística deve ser a mais poderosa.

³ Essa estatística é algumas vezes caracterizada como $\lambda_{\text{máximo}}/(1 + \lambda_{\text{máximo}})$, mas não é a estatística apresentada pelo SPSS.

14.4 HIPÓTESES DA MANOVA ③

A MANOVA tem suposições semelhantes à ANOVA, porém estendidas para o caso multivariado:

- **Independência:** As observações devem ser estatisticamente independentes.
- **Amostragem aleatória:** Os dados devem ser aleatoriamente selecionados da população de interesse e mensurados a um nível de intervalo pelo menos.
- **Normalidade multivariada:** Na ANOVA, assumimos que nossa variável dependente é normalmente distribuída dentro de cada grupo. No caso da MANOVA, assumimos que as variáveis dependentes (coletivamente) têm normalidade multivariada nos grupos.
- **Homogeneidade das matrizes de covariância:** Na ANOVA, assumimos que as variâncias em cada grupo são aproximadamente iguais (homogeneidade das variâncias). Na MANOVA devemos assumir que isso é verdadeiro para cada variável dependente, mas também que as correlações entre quaisquer duas variáveis dependentes sejam as mesmas para todos os grupos. Essa hipótese é examinada testando se a população das matrizes das variâncias-covariâncias dos grupos sendo analisados são iguais.⁴

Quando essas hipóteses são violadas, a precisão dos resultados da estatística teste multivariada fica comprometida. Assim, o que fazer? Bem, o SPSS não oferece uma versão não-paramétrica da MANOVA; contudo, algumas ideias já foram adiantadas com base em postos (muito semelhante aos testes não-paramétricos vistos no Capítulo 13). Embora as discussões desses testes estejam além dos propósitos desse texto, Zwick (1985) sugere algumas ideias que podem ser vantajosas

quando a normalidade multivariada ou a homogeneidade das matrizes de covariância não podem ser assumidas.

14.4.1 Verificando as hipóteses ③

Muitas das hipóteses podem ser testadas da mesma forma que nos testes univariados (veja o Capítulo 8); as suposições adicionais da normalidade multivariada e da igualdade das matrizes das covariâncias requerem procedimentos diferentes. A hipótese de normalidade multivariada não pode ser testada no SPSS e, assim, a única solução prática é testar a hipótese da normalidade univariada para cada variável dependente (veja o Capítulo 3). Essa solução é prática (porque é fácil de ser implementada) e útil (porque a normalidade univariada é uma condição necessária para a normalidade multivariada), mas ela não *garante* a normalidade multivariada. Assim, embora essa abordagem seja o melhor que podemos fazer, sugiro aos leitores que consultem Stevens (1992, Capítulo 6), que fornece algumas soluções alternativas.

A suposição de igualdade das matrizes de covariâncias é mais fácil de ser verificada. Primeiro, para que essa hipótese seja verdadeira, os testes univariados de igualdade de variâncias entre grupos devem ser satisfeitos. Essa hipótese é facilmente verificada pelo teste de Levene (veja a Seção 3.6). Como um teste preliminar, o de Levene não deve ser significativo para qualquer uma das variáveis dependentes. Entretanto, o teste de Levene não leva em conta as covariâncias e assim as matrizes de variâncias-covariâncias devem ser comparadas entre os grupos com a utilização do teste de Box. Esse teste não deve ser significativo se as matrizes são semelhantes. Contudo, o teste de Box é muito suscetível a desvios da normalidade multivariada e, assim, pode não ser significativo não porque as matrizes sejam semelhantes, mas porque a hipótese de normalidade multivariada não foi satisfeita. Por isso, é crucial saber se os dados têm uma distribuição normalmente multivariada antes de interpretar os resultados do teste de Box.

⁴ Para vocês que leem sobre matrizes SSCP, se pensarem sobre o relacionamento entre as somas dos quadrados com variâncias e produtos cruzados com correlações, deve ficar claro que a matriz de variância-covariância é basicamente uma forma padronizada da matriz SSCP.

14.4.2 Escolhendo o teste estatístico ③

Que teste estatístico devo usar?



Somente quando existe uma CL (*variate*) subjacente serão necessários que os quatro testes estatísticos sejam os mesmos. Portanto, é importante saber que teste é melhor em termos de poder e robustez. Tanto Olson (1974; 1976; 1979) quanto Stevens (1979) fizeram um extensivo trabalho sobre o poder de quatro testes MANOVA. Olson (1974) observou que para tamanhos amostrais pequenos e moderados, as quatro estatísticas diferem pouco em termos de poder. Se as diferenças entre os grupos estiverem concentradas na primeira CL (como muitas vezes é o caso na pesquisa em ciências sociais), a estatística de Roy deverá ser a mais poderosa (porque ela leva em conta apenas a primeira CL), seguido pelo traço de Hotelling, λ de Wilks e o traço de Pillai. Contudo, quando os grupos diferem ao longo de mais de uma CL (Combinação Linear ou *variate*), o poder é contrário (isto é, o traço de Pillai é o mais poderoso e a raiz de Roy, o menos). Uma questão final com respeito ao poder dos testes é quanto ao tamanho da amostra e o número de variáveis dependentes. Stevens (1980) recomenda a utilização de um número pequeno de variáveis dependentes (menos do que 10), exceto se os tamanhos amostrais forem grandes.

Em termos de robustez, os quatro testes estatísticos são relativamente robustos a violações da normalidade multivariada (embora a raiz de Roy seja afetada por distribuições platicúrticas – ver Olson, 1976). A raiz de Roy também não é robusta quando a homogeneidade das matrizes de covariâncias não se verifica (Stevens, 1979). O trabalho de Olson e Stevens levaram Bray e Maxwell (1985) a concluir que quando tamanhos amostrais são iguais, o traço de Pillai-Bartlett é o mais robusto a violações das hipóteses. Contudo, quando os tamanhos amostrais são diferentes, essa estatística é afetada pela violação da hipótese de igualdade das matrizes de covariâncias. Como regra, com grupos desiguais, verifique a hipótese

de homogeneidade das matrizes de covariâncias utilizando o teste de Box; se ele não for significativo, e se a hipótese de normalidade multivariada é verdadeira (o que nos permite assumir que o teste de Box é preciso), assuma que o traço de Pillai é o melhor.

14.4.3 Análise adicional ③

Na Seção 14.2.2.1, mencionei que existiam algumas controvérsias sobre o quanto é útil uma análise adicional na MANOVA. A abordagem tradicional é seguir uma MANOVA significativa com ANOVAs separadas em cada uma das variáveis dependentes. Se essa abordagem for tomada, você poderá se perguntar por que se preocupar com a MANOVA (anteriormente, disse que múltiplas ANOVAs não era uma boa abordagem). Bem, a ANOVA que dá prosseguimento a uma MANOVA significativa é dita “protegida” pela MANOVA inicial (Bock, 1975). A ideia é que o teste multivariado global protege contra erros do Tipo I inflados porque se o teste inicial é não-significativo (isto é, a hipótese nula é verdadeira), qualquer teste subsequente é ignorado (qualquer significância deve ser um erro do Tipo I porque a hipótese nula é verdadeira). Contudo, a noção de proteção é um tanto falaciosa porque uma MANOVA significativa, frequentemente, reflete uma diferença significativa para uma, mas não todas, as variáveis dependentes. ANOVAs subsequentes são então executadas em todas as variáveis dependentes, mas a MANOVA protege somente a variável dependente para a qual existe uma diferença de grupo genuína (veja Bray e Maxwell, 1985, p. 40-41). Portanto, você pode considerar a aplicação da correção de Bonferroni as ANOVAs posteriores (Harris, 1975).

A abordagem da ANOVA como sequência da MANOVA implicitamente assume que uma MANOVA significativa não é devida às variáveis dependentes representando um conjunto de dimensões subjacentes que diferenciam os grupos. Portanto, alguns pesquisadores advogam o uso da análise discriminante, que encontra as combinações lineares das variáveis dependentes que melhor *separa* (ou

discrimina) os grupos. Esse procedimento está mais de acordo com o espírito da MANOVA porque engloba os relacionamentos que existem entre as variáveis dependentes e é certamente útil para iluminar os relacionamentos entre as variáveis dependentes e a pertinência aos grupos. A maior vantagem dessa abordagem sobre múltiplas ANOVAs é que ela reduz e explica as variáveis dependentes em termos de um conjunto de dimensões subjacentes que podem refletir dimensões psicológicas ou teóricas substantivas. Por omissão, o procedimento MLG padrão no SPSS fornece ANOVAs univariadas, mas não a análise discriminante.⁵ Entretanto, a análise discriminante pode ser acessada via diferentes menus, e no restante deste capítulo iremos utilizar os dados do TOC (Transtorno Obsessivo-Compulsivo) para ilustrar como essas análises podem ser feitas (aqueles que pularam a seção teórica devem verificar a Tabela 14.1).

⁵ Usuários de versões anteriores a 7.5 devem notar que a análise discriminante pode ser obtida utilizando o menu **Statistics⇒General Linear Model** (ou **ANOVA Models**, como era conhecido na versão 6)⇒**Multivariate...** (Estatística⇒Modelos Lineares Generalizados⇒Multivariada...)

14.5 MANOVA NO SPSS ②

14.5.1 A análise principal ②

Carregue o arquivo **OCD.sav** ou insira os dados manualmente. Se você entrar com os dados manualmente, irá precisar de três colunas: uma coluna deverá ser da variável de código para os grupos (utilizei o seguinte: TCC (CBT) = 1, TC (BT) = 2, ST (NT) = 3),* e nas duas colunas restantes entre com os escores para cada variável dependente, respectivamente. Depois de digitar os dados, acesse a caixa de diálogo principal da MANOVA utilizando o menu **Analyze⇒General Linear Model⇒Multivariate...** (Analisar⇒Modelos Lineares Generalizados)⇒Multivariada...) (veja a Figura 14.2).

As ANOVAs (e várias comparações múltiplas) executadas após a MANOVA principal são o mesmo que executar ANOVAs separadas no SPSS para cada variável dependente. Consequentemente, a caixa de diálogo principal e as opções para a MANOVA são muito semelhantes ao procedimento da ANOVA fatorial

* N. de T.: As siglas entre parênteses são as originais. Elas podem ser utilizadas caso você decida usar o arquivo **OCD.sav**.

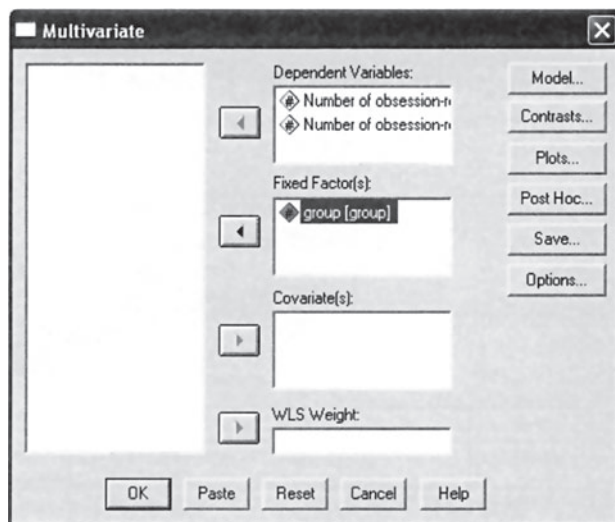







Figura 14.2 Caixa de diálogo para a MANOVA.

que vimos no Capítulo 10. A principal diferença dessa caixa de diálogo é que o quadro **Dependent Variables** (Variáveis Dependentes) tem espaço para diversas variáveis. Selecione as duas variáveis dependentes da lista de variáveis clicando em . Selecione **group** (grupo) da lista de variáveis e transfira-a para o quadro **Fixed Factor(s)** (Fator(es) Fixo(s)) clicando em . Existe também um quadro em que você pode colocar as covariáveis. Para essa análise, não existem covariáveis, contudo, você pode aplicar os princípios da ANCOVA para o caso multivariado e depois realizar uma análise de covariância multivariada (MANCOVA). Depois de especificar as variáveis da análise, você pode selecionar qualquer uma das outras caixas de diálogo clicando nos botões no lado direito:

 (Modelo) Esse botão abre uma caixa de diálogo para particularizar a análise e selecionar o tipo de soma dos quadrados que será utilizada (veja a Seção 10.3.2).

 (Diagramas) Esse botão abre uma caixa de diálogo para selecionar diagramas de interação. Essa opção é útil somente quando mais de duas variáveis independentes tenham sido mensuradas (veja a Seção 10.3.3).

 (Salvar) Esse botão abre uma caixa de diálogo para salvar os resíduos do MLG (isto é, diagnósticos da regressão). Essas opções são úteis para verificar quão bem o modelo aderiu aos dados (veja o Capítulo 5).

14.5.2 Comparações múltiplas na MANOVA ②




A opção padrão para dar prosseguimento a uma MANOVA é olhar para as ANOVAs univariadas individuais para cada variável dependente. Para esses testes, o SPSS tem as mesmas opções que o procedimento da ANOVA univariada (veja o Capítulo 8). O botão  (Contrastes) abre uma caixa de diálogo para especificar uma das várias opções de contrastes disponíveis para a(s) variável(is) independente(s) na análise. A

Tabela 8.6 descreve o que cada um desses testes compara, mas para esse exemplo é possível utilizar um *simple* (contraste simples) que compara cada um dos grupos-experimentais ao grupo dos sem tratamento (controle). O grupo-controle (ST = Sem Tratamento) foi codificado como a última categoria (ele tem o código mais alto no editor de dados), assim, precisamos selecionar a variável de grupo e mudar o contraste para um contraste simples utilizando a última categoria como a categoria de referência (veja a Figura 14.3). Para mais detalhes sobre contrastes, veja a Seção 8.2.10.

Em vez de rodar um contraste, podemos executar testes *post hoc* na variável independente para comparar cada grupo a todos os demais. Para acessar a caixa de diálogo para os testes *post hoc*, clique em . Essa caixa de diálogo é a mesma que para a ANOVA fatorial (veja a Figura 10.5) e a escolha do teste deve ter por base o mesmo critério que foi visto na Seção 8.2.11. Para os objetivos desse exemplo, sugiro selecionar duas das minhas recomendações usuais: REGWQ e Games-Howell. Depois de selecionar os testes *post hoc*, retorne à caixa de diálogo principal.

14.5.3 Opções adicionais ③

Para acessar a caixa de diálogo **options** (Opções), clique em  na caixa de diálogo principal (veja a Figura 14.4). A caixa de diálogo resultante é semelhante à da ANOVA fatorial (veja a Seção 10.3.6); contudo, existem algumas opções adicionais que vale a pena mencionar.

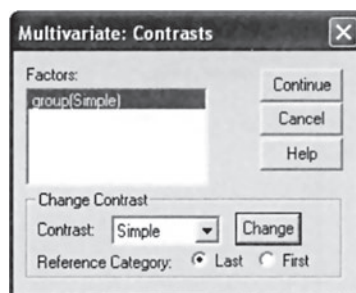


Figura 14.3 Contrastes para a(s) variável(is) independente(s) na MANOVA.

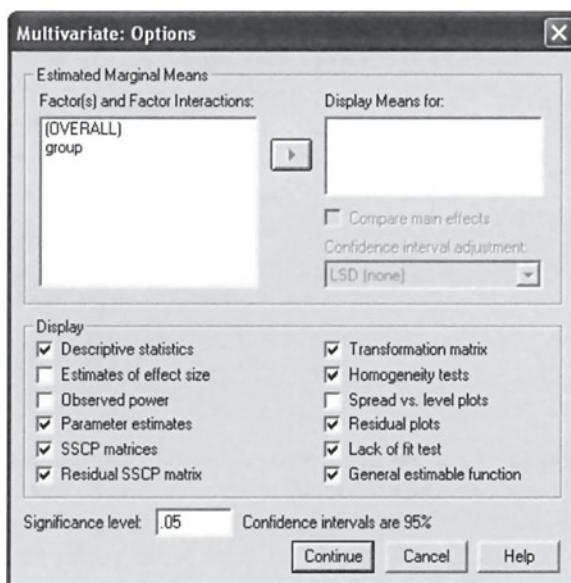


Figura 14.4 Opções adicionais na MANOVA.

- **SSCP Matrices** (Matrizes SSCP): Se essa opção for selecionada, o SPSS produzirá uma matriz SSCP para o modelo, uma para os erros e uma do total. Essa opção pode ser útil para entender os cálculos da MANOVA. Contudo, se você não leu a seção teórica, não selecione essa opção e fique feliz por não ter que se preocupar com essas matrizes!
- **Residual SSCP Matrix** (Matriz SSCP residual): Se essa opção for selecionada, o SPSS irá apresentar uma matriz dos erros, uma dos erros das variâncias-covariâncias e uma dos erros das correlações. A matriz dos erros das variâncias-covariâncias é a matriz sobre a qual o *teste de esfericidade de Bartlett* é baseado. O teste de Bartlett examina se a matriz é proporcional à matriz identidade (isto é, que as covariâncias são zero e as variâncias – os valores ao longo da diagonal – são aproximadamente iguais).

As opções restantes são as mesmas da ANOVA fatorial (e, dessa forma, foram descritas no Capítulo 10); recomendo reler esse capítulo antes de decidir quais são as opções úteis.

14.6 SAÍDAS DA MANOVA ③

14.6.1 Análise preliminar e checagem de hipóteses ③

A Saída 14.1 do SPSS mostra uma tabela inicial das estatísticas descritivas produzida clicando na opção estatísticas descritivas (**descriptives statistics**) na caixa de diálogo opções (**options**) (Figura 14.4). Essa tabela contém as médias gerais dos grupos e os desvios padrão para cada variável dependente. Esses valores correspondem àqueles calculados manualmente na Tabela 14.1; olhando para essa tabela deve ficar claro o que essa parte da saída nos informa. Também é possível ver, a partir das médias, que os participantes tiveram muito mais pensamentos obsessivos do que comportamentos desse tipo.

A Saída 14.2 do SPSS mostra o teste de Box para a hipótese de igualdade das matrizes de covariâncias (veja a Seção 14.4.1). Essa estatística testa a hipótese nula de que as matrizes de variâncias-covariâncias são as mesmas nos três grupos. Portanto, se as matrizes são iguais (e, dessa forma, a hipótese de homogeneidade é satisfeita), essa estatística deve ser

Saída 14.1 do SPSS

Descriptive Statistics (Estatísticas Descritivas)

	Group (Grupo)	Mean (Média)	Std. Deviation (Desvio Padrão)	N
Number of obsession-related thoughts (Número de pensamentos obsessivos)	CBT (TCC)	13,40	1,90	10
	BT (TC)	15,20	2,10	10
	No Treatment Control (Sem Tratamento – Controle)	15,00	2,36	10
	Total (Total)	14,53	2,21	30
Number of obsession-related behaviours (Número de comportamentos obsessivos)	CBT (TCC)	4,90	1,20	10
	BT (TC)	3,70	1,77	10
	No Treatment Control (Sem Tratamento – Controle)	5,00	1,05	10
	Total (Total)	4,53	1,46	30

Saída 14.2 do SPSS

Box's Test of Equality of Covariance Matrices^a (Teste de Box de Igualdade das matrizes de covariâncias)

Box's M (M de Box)	9,959
F	1,482
df1 (gl1)	6
df2 (gl2)	18169
Sig. (Sig.)	0,180

a Delineamento: Intercepto + GRUPO)

Bartlett's Test of Sphericity^a (Teste de Esfericidade de Bartlett)

Likelihood Ratio (Razão de Verossimilhança)	0,042
Approx. Chi-Square (Qui-Quadrado Aproximado)	5,511
df (gl)	2
Sig. (Sig.)	0,064

a Delineamento: Intercepto + GRUPO)

não-significativa. Para esses dados, $p = 0,18$ (que é maior do que 0,05): assim, as matrizes de covariâncias devem ser aproximadamente iguais e a hipótese é verificada.

Se o valor do teste de Box for significativo ($p < 0,05$), as matrizes de covariâncias são significativamente diferentes e a hipótese de homogeneidade terá sido violada. O efeito da violação dessa suposição não é clara. Haskstian, Roed e Lind (1979) relatam que o T^2 de Hotelling é robusto na situação de dois grupos quando os tamanhos amostrais são iguais. Como regra geral, se os tamanhos amostrais são iguais, desconsidere o teste de Box, por-

que ele é altamente instável, e assuma que as estatísticas de Hotelling e Pillai sejam robustas (veja a Seção 14.4.2). Contudo, se os tamanhos dos grupos forem diferentes, a robustez não pode ser assumida (especialmente se o teste de Box é significativo a $p < 0,001$). Quanto mais variáveis dependentes tenham sido mensuradas e quanto maior forem as diferenças nos tamanhos amostrais, mais distorcidos se tornarão os valores das probabilidades produzidos pelo SPSS. Tabachnick e Fidell (2001), contudo, sugerem que se amostras grandes produzem grandes variâncias e covariâncias, os valores probabilísticos serão conservadores (e as significâncias encontradas serão confiáveis). Contudo, se forem amostras pequenas que produzirem grandes variâncias e covariâncias, os valores probabilísticos serão liberais e, assim, as diferenças significativas devem ser tratadas com cautela (entretanto, os efeitos não-significativos podem ser confiáveis). Como tal, o teste de Box precisa de fato ser examinado quando os tamanhos amostrais diferem: não devemos acreditar nele quando a normalidade multivariada não puder ser assumida (ou esteja sendo questionada), e as matrizes de variâncias-covariâncias de amostras devam ser inspecionadas para avaliar se as probabilidades apresentadas são conservadoras ou liberais. Na situação em que não se pode acreditar nas probabilidades obtidas, existe muito pouco a ser feito exceto equalizar as amostras eliminando casos ao acaso nos grupos maiores

(embora com essa perda de informação ocorra também uma perda de poder).

O teste de Bartlett de esfericidade testa se a hipótese de esfericidade foi satisfeita. Ele é útil somente nos delineamentos univariados de medidas repetidas porque a MANOVA não requer essa suposição.

14.6.2 Teste estatístico MANOVA ③

A Saída 14.3 do SPSS mostra a tabela principal dos resultados. As estatísticas teste são citadas para o intercepto do modelo (até mesmo a MANOVA pode ser caracterizada como um modelo de regressão, embora como isso é feito esteja fora do escopo deste livro) e para a variável de grupo. Para os nossos propósitos, os efeitos de grupo são de interesse porque eles nos informam se as terapias tiveram ou não um efeito sobre os pacientes com TOC (OCD). Você verá que o SPSS lista as quatro estatísticas multivariadas e seus valores correspondem àqueles que foram calculados nas Seções 14.3.4.2 e 14.3.4.5. Na próxima coluna, esses valores são transformados em uma razão F. A coluna que realmente interessa, contudo, é aquela que contém os valores das significâncias das razões F. Para esses dados, o traço de Pillai ($p = 0,049$), o lambda de Wilks ($p = 0,050$) e a maior raiz de Roy ($p = 0,020$) alcançaram o critério para significância de 0,05. Entretanto, o traço de Hotelling (p

$= 0,051$) é não-significativo por esse critério. Esse cenário é interessante, porque a estatística teste que escolhemos determina se rejeitaremos ou não a hipótese nula de que não existem diferenças entre os grupos. Entretanto, dado que sabemos sobre a significância do traço de Pillai quando os tamanhos amostrais são iguais, podemos acreditar no resultado dessa estatística teste, que indica uma diferença significativa. Esse exemplo destaca o poder associado à raiz de Roy (você deve notar que essa estatística tem uma significância maior do que as demais) quando as suposições do teste foram satisfeitas.

A partir desse resultado, provavelmente concluímos que o tipo de terapia empregado tem um efeito significativo no TOC (OCD). A natureza desse efeito não está clara a partir da estatística teste multivariada. Primeiro, ela nada informa sobre quais grupos diferem de quais, segundo, não diz nada sobre se o efeito da terapia foi nos pensamentos ou nos comportamentos obsessivos ou na combinação dos dois. Para determinar a natureza do efeito, o SPSS nos oferece os testes univariados.

14.6.3 Teste estatístico Univariado ②

A Saída 14.4 do SPSS inicialmente mostra uma tabela resumo do teste de Levene de igualdade das variâncias para cada variável dependente. Esses testes são os mesmos que

Saída 14.3 do SPSS

Multivariate Tests^a (Testes Multivariados)

Effect (Efeito)	Value (Valor)	Value (Valor)	F	Hypothesis df (Gl da Hipótese)	Error df (Gl do Erro)	Sig, (Sig.)
Intercept (Intercepto)	Pillai's Trace (Traço de Pillai)	0.983	745.230 ^c	2.000	26.000	0.000
	Wilks' Lambda (Lambda de Wilks)	0.017	745.230 ^c	2.000	26.000	0.000
	Hotelling's Trace (Traço de Hotelling)	57.325	745.230 ^c	2.000	26.000	0.000
	Roy Largest Root (Maior Raiz de Roy)	57.325	745.230 ^c	2.000	26.000	0.000
GROUP (Grupo)	Pillai's Trace (Traço de Pillai)	0.318	2.557	4.000	54.000	0.049
	Wilks' Lambda (Lambda de Wilks)	0.699	2.555 ^c	4.000	52.000	0.050
	Hotelling's Trace (Traço de Hotelling)	0.407	2.546	4.000	50.000	0.051
	Roy Largest Root (Maior Raiz de Roy)	0.335	4.520	2.000	27.000	0.020

a. Delineamento: Intercepto + GRUPO.

b Calculado utilizando alfa = 0,05.

c Estatística exata.

Saída 14.4 do SPSS

Levene's Test of Equality of Error Variances^a
(Teste de Levene para a Igualdade das Variâncias dos Erros)

	<i>F</i>	<i>df1</i> (gl1)	<i>df2</i> (gl2)	<i>Sig.</i> (Sig.)
<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	0.76	2	27	0.927
<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	1.828	2	27	0.180

Testa a hipótese de que as variâncias dos erros da variável dependente são iguais entre os grupos.
a Delineamento: Intercepto + GRUPO.

Tests of Between-Subjects Effects (Testes dos efeitos entre participantes)

<i>Source</i> (Fonte)	<i>Dependent Variable</i> (Variável Dependente)	<i>Type III Sum of Squares</i> (Soma dos Quadrados do Tipo III)	<i>df</i> (gl)	<i>Mean Square</i> (Média dos Quadrados)	<i>F</i>	<i>Sig.</i> (Sig)
<i>Corrected Model</i> (Modelo Corrigido)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	19.467 ^b	2	9.733	2.154	0.136
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	10.467 ^c	2	5.233	2.771	0.080
<i>Intercept</i> (Intercepto)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	6336.533	1	6336.533	1402.348	0.000
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	616.533	1	616.533	326.400	0.000
<i>GROUP</i> (Grupo)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	19.467	2	9.733	2.154	0.136
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	10.467	2	5.233	2.771	0.080
<i>Error</i> (Erro)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	122.000	27	4.519		
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	51.000	27	1.889		
<i>Total</i> (Total)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	6478.000	30			
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	678.000	30			
<i>Corrected Total</i> (Total Corrigido)	<i>Number of obsession-related thoughts</i> (Número de pensamentos obsessivos)	141.467	29			
	<i>Number of obsession-related behaviours</i> (Número de comportamentos obsessivos)	61.467	29			

a Calculado utilizando alfa = 0,05.
b R ao quadrado = 0.138 (R ao quadrado ajustado = 0,074).
c R ao quadrado = 0.170 (R ao quadrado ajustado = 0,109).

seriam encontrados se uma ANOVA de um fator tivesse sido realizada em cada variável dependente (veja o Capítulo 8). O teste de Levene deve ser não-significativo para todas as variáveis dependentes se as suposições de homogeneidade das variâncias forem válidas.

Os resultados para esses dados mostram claramente que as suposições foram satisfeitas. Esse resultado não apenas nos dá confiança nos testes univariados que seguem, mas também reforça a ideia de que a estatística teste multivariada é robusta.

A próxima parte da saída contém uma tabela resumo da ANOVA para as variáveis dependentes. A linha de interesse é a denominada **GROUP** (Grupo) (você notará que os valores nessa linha são os mesmos da linha **Corrected Model** (Modelo Corrigido); isto é, porque o modelo ajustado aos dados contém somente uma variável independente: **group** (grupo)). A linha **GROUP** (GRUPO) contém um resumo da ANOVA para cada uma das variáveis dependentes que são valores para a soma dos quadrados tanto para ações (*actions*) quanto para pensamentos (*thoughts*) (esses valores correspondem aos valores do SS_M calculado nas Seções 14.3.3.1 e 14.3.3.2, respectivamente). A linha rotulada **Error** (Erro) contém informações sobre a soma dos resíduos dos quadrados e a média dos quadrados para cada uma das variáveis dependentes: esses valores do SS_R foram calculados nas Seções 14.3.3.1 e 14.3.3.2; retorne a essas seções para confirmar o significado desses valores. A linha **Corrected Total** (Total Corrigido) contém os valores da soma total dos quadrados para cada variável dependente (novamente, esses valores SS_T foram calculados nas Seções 14.3.3.1 e 14.3.3.2). As partes importantes dessa tabela são as colunas denominadas **F** e **Sig.** nas quais as razões **F** para cada ANOVA univariada e o valor das suas significâncias são apresentados. O que deve estar claro na Saída 14.4 do SPSS e nos cálculos feitos nas Seções 14.3.3.1 e 14.3.3.2 é que os valores associados com as ANOVAs univariadas realizadas após a MANOVA são *idênticos* aos valores obtidos na ANOVA de um fator que foi realizada em cada variável dependente. Esse fato ilustra que a MANOVA oferece somente uma proteção hipotética sobre a inflação da taxa de erro do Tipo I: não existe um ajustamento real que possa ser feito nos valores obtidos.

Os valores de **p** na Saída 14.4 do SPSS indicam que existe uma diferença não-significativa entre os grupos de terapia tanto em termos de pensamentos obsessivos ($p = 0,136$) quanto de comportamentos obsessivos ($p = 0,080$). Esses dois resultados nos levam a concluir que não é possível afirmar que o tipo de terapia

tenha um efeito significativo nos níveis de TOC (OCD) experimentado pelos pacientes. Vocês podem ter notado algo estranho nesse exemplo: o teste estatístico multivariado nos levou a concluir que a terapia tem um impacto significativo no TOC, enquanto os resultados univariados indicam que a terapia não foi bem-sucedida. Antes de continuar, pense um pouco sobre o porquê disso ocorrer.

A razão desse tipo de resultado é simples: o teste multivariado leva em conta as correlações entre as variáveis dependentes, assim, para esses dados ele tem mais poder para detectar diferenças entre os grupos. Com esse conhecimento em mente, os testes univariados não são úteis para interpretação, porque os grupos diferem por uma combinação das variáveis dependentes. Para ver como variáveis dependentes interagem, precisamos realizar uma análise da função discriminante, que será descrita na Seção 14.7.

14.6.4 Matrizes SSCP ③

Se você selecionou as duas opções para obter matrizes SSCP (Seção 14.5.3), o SPSS produzirá as tabelas mostradas na Saída 14.5 do SPSS e na Saída 14.6 do SPSS. A primeira tabela (Saída 14.5 do SPSS) mostra a SSCP do modelo (H), denominada **Hypothesis GROUP** (Hipóteses de GRUPO) (utilizei um sombreamento para destacar essa matriz), e a SSCP erro (E), rotulada de **Error** (Erro) (utilizei um sombreamento mais forte para destacar essa matriz). A matriz para o intercepto também é mostrada, mas essa matriz não é importante para os nossos propósitos. Deve estar bem claro que os valores nas matrizes do modelo e do erro apresentadas na Saída 14.5 do SPSS correspondem aos valores que calculamos nas Seções 14.3.3.6 e 14.3.3.5, respectivamente. Essas matrizes são úteis, portanto, para compreender o padrão dos dados, especialmente se olharmos para os valores dos produtos cruzados que indicam o relacionamento entre as variáveis dependentes. Nesse exemplo, os valores da matriz SSCP, somas dos quadrados para os erros, são substancialmente maiores do que os valores da matriz SSCP modelo (ou

Saída 14.5 do SPSS

Between-Subjects SSCP Matrix (Matriz SSCP entre participantes)

			Number of obsession-related behaviours (Número de comportamentos obsessivos)	Number of obsession- related thoughts (Número de pensamentos obsessivos)
Hypotesis (Hipótese)	Intercept (Intercepto)	Number of obsession-related behaviours (Número de compor- tamentos obsessivos)	616.533	1976.533
		Number of obsession-related thoughts (Número de pensa- mentos obsessivos)	1976.533	6336.533
	GROUP (Grupo)	Number of obsession-related behaviours (Número de compor- tamentos obsessivos)	10.467	-7.533
		Number of obsession-related thoughts (Número de pensa- mentos obsessivos)	-7.533	19.467
Error (Erro)	Number of obsession-related behaviours (Número de compor- tamentos obsessivos)		51.000	13.000
	Number of obsession-related thoughts (Número de pensa- mentos obsessivos)		13.000	122.000

Com base na soma dos quadrados to Tipo III.

Saída do SPSS 14.6

Residual SSCP Matrix (Matriz SSCP Residual)

		Number of obsession-related behaviours (Número de comportamentos obsessivos)	Number of obsession-related thoughts (Número de pensamentos obsessivos)
Sum-of-Squares and Cross-Products (Soma dos quadrados dos pro- dutos cruzados)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	51.000	13.000
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	13.000	122.000
Covariance (Covariância)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	1.889	0.481
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	0.481	4.519
Correlation (Correlação)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	1.000	0.165
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	0.165	1.000

Com base na soma dos quadrados to Tipo III.

grupo), enquanto os valores absolutos dos produtos cruzados são bem parecidos. Esse padrão sugere que se a MANOVA é significativa: portanto, talvez o relacionamento entre as variáveis dependentes seja importante, e não as variáveis dependentes individualmente.

A Saída 14.6 do SPSS mostra a matriz SSCP residual novamente, mas dessa vez ela inclui a matriz de variâncias-covariâncias e a matriz das correlações. Essas matrizes estão todas relacionadas. Se você voltar ao Capítulo 4, verá que a covariância é calculada pela divisão do produto cruzado pelo número de observações (isto é, a covariância é a média dos produtos cruzados). Da mesma forma, a variância é calculada pela divisão das somas dos quadrados pelo número de graus de liberdade (e, assim, de forma semelhante representa a média da soma dos quadrados). Por conseguinte, a matriz de variâncias-covariâncias representa a forma média da matriz SSCP. Finalmente, vimos no Capítulo 4 que a correlação era uma versão padronizada da covariância (onde o desvio padrão é também levado em consideração), assim, a matriz de correlações representa a forma padrão da matriz de variâncias-covariâncias. Como ocorreu com a matriz SSCP, essas outras matrizes são úteis para determinar a extensão do erro do modelo.

A matriz de variâncias-covariâncias é especialmente útil porque o teste de esfericidade de Bartlett é baseado nessa matriz. O teste de Bartlett verifica se a matriz é proporcional a uma matriz identidade. Na Seção 14.3.1, vimos que na matriz identidade os elementos da diagonal são iguais a 1 e os fora da diagonal são 0. Entretanto, o teste de Bartlett testa se os elementos da diagonal da matriz de variâncias-covariâncias são iguais (isto é, as variâncias dos grupos são as mesmas) e se os elementos fora da diagonal são aproximadamente zero (isto é, as variáveis dependentes não estão correlacionadas). Nesse caso, as variâncias são bastante diferentes (1,89 comparado com 4,52) e as covariâncias são levemente diferentes de zero (0,48); assim, o teste de Bartlett deve apresentar um resultado significativo (veja a Saída 14.2 do SPSS). Embora essa discussão seja irrelevante para os testes multivariados, espero que com ela você possa relacionar essas ideias com as questões de esfericidade apresentadas no Capítulo 11 e ver com maior clareza como essa hipótese é testada.

14.6.5 Contrastes ③

Na Seção 14.5.2, sugeri a realização de contrastes simples (*simple*) que comparam cada um dos grupos submetidos à terapia com o grupo-controle (não tratados). A Saída 14.7 do SPSS mostra os resultados desses contrastes. Essa tabela está dividida em duas seções denominadas *Level 1 vs. Level 3* (Nível 1 vs. Nível 3) e *Level 2 vs. Level 3* (Nível 2 vs. Nível 3), onde os números correspondem aos códigos da variável de grupo (isto é, 1 representa o código mais baixo utilizado no editor de dados e 3, o mais alto). Se você codificou a variável de grupo utilizando os mesmos valores que eu, então esses contrastes representam TCC vs. ST (CBT vs. NT) e TC vs. ST (BT vs. NT), respectivamente. Cada contraste é executado nas duas variáveis dependentes separadamente e, assim, eles são idênticos aos contrastes que seriam obtidos de uma ANOVA univariada. Essa tabela fornece valores para a estimativa dos contrastes e o valor hipotético (que será sempre zero porque estamos testando a hipótese nula de que a diferença entre grupos é zero). A diferença estimada observada é então testada para ver se ela é significativamente diferente de zero tendo por base o erro padrão (você poderá reler os Capítulos 1 e 7 para ver alguma teoria nesse tipo de teste de hipóteses). Um intervalo de 95% de confiança é produzido para a diferença estimada.

A partir dos valores da *Sig.*, você deve ter notado que quando comparamos TCC com ST (CBT com NT), não existem diferenças nos pensamentos ($p = 0,104$) ou comportamentos ($p = 0,872$) porque os dois valores estão acima do ponto de corte que é, nesse caso, 0,05. Contudo, comparando TC com ST (BT com NT), não existe diferenças em pensamentos ($p = 0,835$), mas existe uma diferença significativa nos comportamentos entre os grupos ($p = 0,044$, que é menor que 0,05). Os intervalos de confiança confirmam esses achados: vimos anteriormente que um intervalo de 95% de confiança que deve conter a verdadeira diferença entre os grupos em 95% das vezes. Se esses limites cruzarem zero (isto é, o inferior é um número negativo e o superior, positivo), isso

Saída 14.7 do SPSS

Contrast Results (K Matrix) (Resultados dos Contrastes – Matriz K)

Group Simple Contrast (Contrastes simples entre os grupos)		Dependent Variable (Variável dependente)	
		Number of obsession-related thoughts (Número de pensamentos obsessivos)	Number of obsession-related behaviours (Número de comportamentos obsessivos)
Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Contrast Estimate (Estimativa do Contraste)	–1.600	–0.100
	Hypothesized Value (Valor Hipotético)	0	0
	Difference (Estimate – Hypothesized) (Diferença (Estimativa – Hipotético))	–1.600	–0.100
	Std. Error (Erro Padrão)	0.951	0.615
	Sig (Sig.)	0.104	0.872
	95% Confidante Interval for Difference (Intervalo de 95% de Confiança para a Diferença)	–3.551	–1.361
	Lower Bound (Limite Inferior)		
Level 1 vs. Level 3 (Nível 1 vs. Nível 3)	Contrast Estimate (Estimativa do Contraste)	0.200	–1.300
	Hypothesized Value (Valor Hipotético)	0	0
	Difference (Estimate – Hypothesized) (Diferença (Estimativa – Hipotético))	0.200	–1.300
	Std. Error (Erro Padrão)	0.951	0.615
	Sig (Sig.)	0.835	0.044
	95% Confidante Interval for Difference (Intervalo de 95% de Confiança para a Diferença)	–1.751	–2.561
	Lower Bound (Limite Inferior)		
	Upper Bound (Limite Superior)	2.151	–0.039

a Reference category = 3 (Categoria de referência = 3).

Dica da Samanta Ferrinho



- A MANOVA é utilizada para testar diferenças entre grupos com diversas variáveis dependentes simultaneamente.
- O teste de Box verifica a hipótese da igualdade das matrizes de covariâncias. Esse teste pode ser ignorado quando os tamanhos amostrais são os mesmos, pois quando isso acontece a MANOVA é robusta a violações dessa hipótese. Se os tamanhos dos grupos diferem, esse teste deve ser considerado. Se o valor da *Sig.* for menor do que 0,001, os resultados da análise não são confiáveis (veja a Seção 14.6.1).
- A tabela Testes Multivariados (*Multivariate Tests*) apresenta os resultados da MANOVA. Existem quatro estatísticas testes listadas (Traço de Pillai, Lambda de Wilks, Traço de Hotelling e Maior Raiz de Roy). Recomendo a utilização do traço de Pillai. Se o valor da *Sig.* dessa estatística for menor que 0,05, os grupos diferem significativamente em relação às variáveis dependentes.
- Testes ANOVA podem ser utilizados para dar seguimento à MANOVA (uma ANOVA diferente para cada variável dependente). Os resultados desses testes estão listados na tabela Testes de Efeitos Entre Participantes (*Tests of Between-Subjects Effects*). Essas ANOVAs podem, por sua vez, serem complementadas pela utilização de contrastes (veja os Capítulos 8 a 12). Pessoalmente, não recomendo essa abordagem e sugiro que seja feita uma *análise discriminante*.

nos informa que o valor da verdadeira diferença pode ser zero (ou seja, não existirá diferença entre os grupos). Entretanto, não podemos estar confiantes de que a diferença entre grupos observada é significativa porque a verdadeira diferença entre os grupos na população pode ser zero. Se, entretanto, o intervalo de confiança não contém o zero (isto é, os dois limites ou são positivos ou negativos), podemos ficar confiantes de que o verdadeiro valor da diferença entre grupos é diferente de zero. Desse modo, podemos estar certos de que existe uma diferença genuína entre os grupos. Para esses dados, todos os intervalos de confiança incluem o zero (os limites inferiores são negativos e os superiores são positivos), exceto para o contraste TC *versus* ST para comportamentos; assim somente esse contraste é significativo. Isso é inesperado, pois a ANOVA univariada para comportamentos não foi significativa e, assim, não esperávamos que existisse alguma diferença entre os grupos.

14.7 MANOVA ADICIONAL COM ANÁLISE DISCRIMINANTE ③

Mencionei anteriormente que uma MANOVA significativa pode ser seguida pela utilização tanto de ANOVAs univariadas quanto da análise discriminante. No exemplo deste capítulo, as ANOVAs univariadas não foram uma forma útil de verificar o que

o teste multivariado estava mostrando porque o relacionamento entre as variáveis dependentes obviamente foi responsável pelo efeito. Entretanto, esses dados foram projetados especificamente para ilustrar como as ANOVAs univariadas devem ser tratadas com cautela e, na vida real, é bem provável que uma MANOVA significativa seja acompanhada de pelo menos uma ANOVA significativa. Contudo, isso não significa que o relacionamento entre as variáveis dependentes não seja importante, e ele é ainda vital para investigar a natureza do relacionamento. A análise discriminante é a melhor maneira de obter isso, e eu recomendo que você dê prosseguimento a uma MANOVA tanto com os testes univariados quanto com uma análise discriminante se quiser entender os seus dados totalmente.

É simples realizar uma análise discriminante no SPSS: acesse a caixa de diálogo principal pelo caminho **Analyze**⇒**Classify**⇒**Discriminant...** (Analisar⇒Classificar⇒Discriminante...) (veja a Figura 14.5).

A caixa de diálogo principal irá listar as variáveis no editor de dados no lado esquerdo e apresentar dois espaços no direito: um para a variável de grupo e outro para as previsoras. Na análise discriminante, tentamos separar (ou discriminar) um conjunto de grupos utilizando vários previsores (assim, é mais ou menos como uma regressão logística, mas onde existem vários grupos em vez de ape-

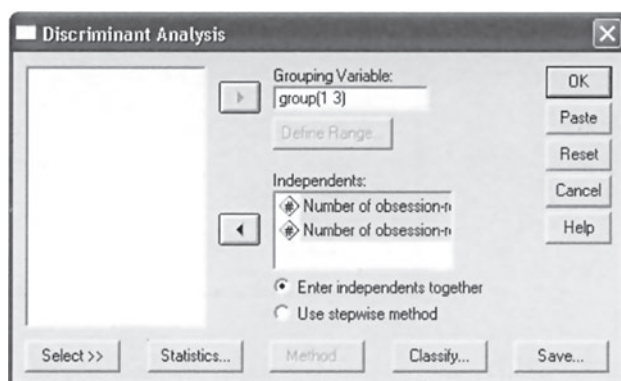




Figura 14.5 Caixa de diálogo principal para a análise discriminante.

nas dois).⁶ Pode ser confuso pensar ações e pensamentos como variáveis independentes (mesmo porque elas eram variáveis dependentes na MANOVA!), por isso estou me referindo a elas como previsores – essa é outra razão para não rotular as variáveis como dependentes e independentes em uma análise de correlação.

Para executar a análise, selecione a variável **group** (grupo) e transfira-a para o quadro **Grouping Variable** (Variável de Grupo) clicando em . Uma vez que essa variável tenha sido transferida, o botão **Define Range** (Definir intervalo...) se tornará ativo e você deverá clicá-lo para obter a caixa de diálogo na qual poderá especificar os valores extremos dos códigos utilizados (1 e 3, nesse caso). Uma vez que os códigos utilizados pela variável de grupo tenham sido especificados, você deve selecionar as variáveis **actions** (ações) e **thoughts** (pensamentos) e transferi-las para a caixa **Independents** (Independentes) clicando em . Existem duas opções disponíveis para determinar como os previsores são fornecidos ao modelo. O **Default** (Padrão) é que ambos sejam colocados juntos e essa é a opção que precisamos (porque na MANOVA as variáveis dependentes são analisadas simultaneamente). É possível entrar com as variáveis dependentes passo a passo (**Stepwise**), e se essa opção estiver selecionada, o botão **Method...** se tornará ativo, o que abrirá a caixa de diálogo para especificar um critério de entrada para os previsores. Para prosseguir com a MANOVA, precisamos somente estar preocupados com as opções restantes.

Clique em **Statistics...** (Estatísticas) para ativar a caixa de diálogo semelhante à da Figura 14.6. Essa caixa de diálogo nos permite requerer o cálculo das médias dos grupos, a ANOVA univariada e o teste de Box da igualdade das matrizes de covariâncias, todos já

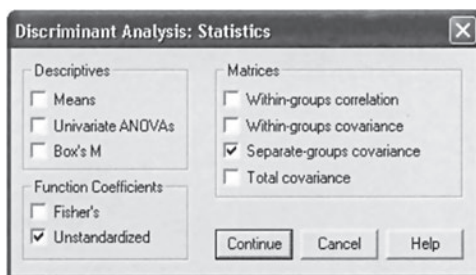


Figura 14.6 Opções para a análise discriminante.

fornecidos na saída da MANOVA (assim, não precisamos solicitá-los novamente). Além disso, podemos solicitar as correlações entre grupos e as matrizes de covariâncias, que são as mesmas correlações residuais e a matriz de covariância vistas na Saída 14.6 do SPSS. Existe, também, uma opção para apresentar uma matriz de covariâncias para grupos separados, que pode ser útil para entender o relacionamento entre as variáveis dependentes para cada grupo (essa matriz é um resultado que a MANOVA não fornece e recomendo selecioná-la). Finalmente, podemos solicitar a matriz de covariâncias total, que apresenta as variâncias e covariâncias de todas as variáveis dependentes. Outra opção útil é selecionar os coeficientes não-padronizados (**Unstandardized**) da função. Essa opção irá produzir os **bs** não-padronizados para cada CL (Combinação Linear – **Variate**) (veja a Equação (14.5)). Quando você tiver terminado com essa caixa de diálogo, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

Se você clicar em **Classify...** (Classificar), terá acesso a uma caixa de diálogo semelhante à da Figura 14.7. Esse local apresenta várias opções disponíveis. Primeiro, você pode selecionar como as probabilidades a priori são determinadas: se os tamanhos dos grupos forem os mesmos, você não deve alterar os valores padrão que são apresentados; contudo, se você tem um delineamento não-balanceado, é recomendável basear as probabilidades a priori nos tamanhos dos grupos observados. A opção por omissão (**Default**) para realizar a análise

⁶ Eu poderia ter apresentado a análise discriminante no Capítulo 6 em vez da regressão logística porque elas são maneiras diferentes de chegar aos mesmos resultados. Entretanto, a regressão logística tem algumas hipóteses restritivas e é geralmente mais robusta, por isso apresentei a análise discriminante somente neste capítulo.

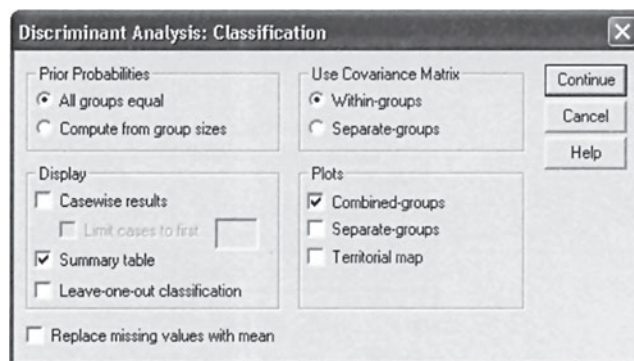


Figura 14.7 Opções de classificação para a análise discriminante.

lise da matriz de covariâncias dentro do grupo é boa (porque essa é a matriz sobre a qual a MANOVA está baseada). Devemos, também, requerer os diagramas de grupos combinados, que irão apresentar gráficos separados da CLs para cada participante agrupados de acordo com a terapia que receberam. O diagrama dos grupos separados mostra a mesma coisa, mas utilizando gráficos diferentes para cada um dos grupos; quando o número de grupos é pequeno, é melhor selecionar um diagrama combinado porque ele é mais fácil de interpretar. As opções restantes são de pouco interesse quando análise discriminante é utilizada como um prosseguimento da MANOVA. A única opção útil é a tabela resumo, que fornece uma medida geral de quão bem as combinações lineares (CL) discriminantes classificam os participantes reais. Quando você tiver terminado

com as opções, clique em **Continue** para retornar à caixa de diálogo principal.

As opções finais podem ser acessadas clicando em **Save...** (Salvar) para acessar uma caixa de diálogo semelhante à da Figura 14.8. Existem três opções disponíveis, duas das quais estão relacionadas às probabilidades e a que grupo os valores previstos pertencem. Esses valores são comparáveis àqueles obtidos com a análise de regressão logística (veja o Capítulo 6). A opção final é o fornecimento dos **escores discriminantes**. Esses são os escores para cada pessoa, em cada CL (*variate*), obtidos a partir da equação (14.5). Esses escores podem ser úteis porque as CLs que a análise identifica podem representar construtos sociais ou psicológicos subjacentes. Se esses construtos são identificáveis, para interpretá-los é útil saber o escore de cada participante em cada dimensão.

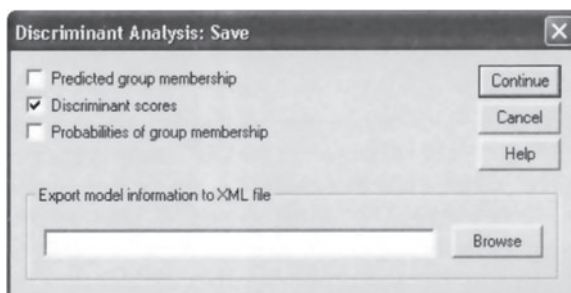


Figura 14.8 Caixa de diálogo *Save new variables* (Salvar novas variáveis) da análise discriminante.

14.8 SAÍDAS DA ANÁLISE DISCRIMINANTE ④

A Saída 14.8 do SPSS mostra as matrizes de covariância para grupos separados (selecionadas na Figura 14.6) Essas matrizes são construídas com as variâncias de cada variável dependente para cada grupo (esses valores são mostrados na Tabela 14.1). As covariâncias são obtidas tomando o produto cruzado entre as variáveis dependentes para cada grupo (mostrados na Tabela 14.4 como 0,40, 22,6 e -10) e dividindo cada um por 9 (os graus de liberdade), $N - 1$ (onde N é o número de observações). Os valores dessa tabela são úteis porque eles nos dão uma ideia de como o relacionamento entre as variáveis dependentes muda de grupo para grupo. Por exemplo, no grupo TCC, os comportamentos e pensamentos não têm praticamente relacionamento linear porque a covariância é praticamente zero. No grupo TC, pensamentos e ações estão positivamente relacionados, assim, quando o número de comportamentos diminui, o mesmo ocorre com os pensamentos. Na condição ST (Sem Tratamento), existe um relacionamento negativo, desse modo, se o número de pensamentos aumenta, o número de comportamentos (ações) diminui. É importante notar que essas matrizes não informam sobre

a importância do relacionamento porque elas não estão padronizadas (veja o Capítulo 4); elas apenas fornecem uma indicação básica.

A Saída 14.9 do SPSS mostra as estatísticas iniciais da análise discriminante. Primeiro conhecemos os autovetores para cada CL (*variate*), e você deve notar que os valores correspondem aos valores dos elementos da diagonal da matriz HE^{-1} (calculados no arquivo **AppendixChapter14.pdf**, disponível no site www.artmed.com.br). Esses autovetores são convertidos em percentuais da variância da CL e a primeira delas é responsável por 82,2% da variância comparada com a segunda CL, responsável por apenas 17,8%. A próxima parte da saída mostra o lambda de Wilks, que tem o mesmo valor (0,699), os graus de liberdade (4) e o valor da significância (0,05) como na MANOVA (veja a Saída 14.3 do SPSS). O ponto importante a ser destacado dessa tabela é que somente uma das CLs (*variates*) é significativa (a segunda não é significativa, $p = 0,173$). Portanto, as diferenças entre os grupos mostradas pela MANOVA podem ser explicadas em termos de apenas *uma* dimensão subjacente.

As tabelas da Saída 14.10 do SPSS são as mais importantes para interpretação. A primeira tabela mostra os coeficientes padro-

Saída 14.8 do SPSS

Covariance Matrices (Matrizes de Covariância)

Group (Grupo)		Number of obsession-related behaviours (Número de comportamentos obsessivos)	Number of obsession-related thoughts (Número de pensamentos obsessivos)
CBT (TCC)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	1.433	4.4444E-02
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	4.444E-02	3.600
BT (TC)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	3.122	2.511
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	2.511	4.400
No Treatment Control (Sem Tratamento Controle)	Number of obsession-related behaviours (Número de comportamentos obsessivos)	1.111	-1.111
	Number of obsession-related thoughts (Número de pensamentos obsessivos)	-1.111	5.556

Saída 14.9 do SPSS

Eigenvalues (Autovalores)

Function (Função)	Eigenvalue (Autovalor)	% of Variance (% da Variância)	Cumulative % (% Acumulado)	Canonical Correlation (Correlação Canônica)
1	0.335 ^a	82.2	82.2	0.501
2	0.073 ^a	17.8	100.0	0.260

a As primeiras duas funções discriminantes canônicas foram utilizadas na análise.

Wilks's Lambda (Lambda de Wilks)

Test of Function(s) (Teste de Função(ões))	Wilks's Lambda (Lambda de Wilks)	Chi-square (Qui-quadrado)	df (gl)	Sig. (Sig.)
1 through 2 (1 até 2)	0.699	9.508	4	0.050
2	0.932	1.856	1	0.173

Saída 14.10 do SPSS

Standardized Canonical Discriminant Function
Coefficients (Coeficientes padronizados
da função discriminante canônica)

	Function (Função)	
	1	2
Number of obsession-related behaviours (Número de comportamentos obsessivos)	0.829	0.584
Number of obsession-related thoughts (Número de pensamentos obsessivos)	-0.713	0.721

Structure Matrix (Matriz de Estrutura)

	Function (Função)	
	1	2
Number of obsession-related thoughts (Número de pensamentos obsessivos)	0.711*	0.703
Number of obsession-related behaviours (Número de comportamentos obsessivos)	-0.576	0.817*

Correlações entre grupos agrupadas entre variáveis discriminantes e funções discriminantes canônicas padronizadas Variáveis ordenadas pelas correlações absolutas entre cada variável e cada função discriminante.

* Maior correlação absoluta entre cada variável e cada função discriminantes.

nizados da função discriminante para as duas CLs. Esses valores são versões padronizadas dos valores dos autovetores calculados na Seção 14.3.4.1. Lembre que as CLs podem ser expressas em termos de uma equação de regressão linear (veja a Equação (14.4)) e os coeficientes padronizados da função discriminante são equivalentes aos betas padronizados da regressão. Por conseguinte, esses coeficientes nos informam a contribuição relativa de cada variável para a CL. Está claro, a partir do tamanho dos valores para esses dados, que o número de comportamentos obsessivos tem uma contribuição maior para a primeira CL do que o número de pensamentos, mas que o contrário é verdadeiro para a segunda CL. Também, lembrando que os coeficientes beta padronizados variam no intervalo de -1 a 1 , vale notar que as duas variáveis tem uma grande contribuição para a primeira CL (isto é, ambas são importantes) porque seus valores estão próximos de 1 e -1 , respectivamente. Tendo em mente que somente a primeira CL é importante, podemos concluir que é necessário reter as duas variáveis dependentes no conjunto de discriminadores (porque suas ponderações padronizadas são praticamente de mesma magnitude). O fato de que uma variável dependente tem um peso negativo e um positivo indica que as diferenças de grupos são explicadas pelas diferenças entre as variáveis dependentes.

Outra forma de olhar para os relacionamentos entre variáveis dependentes e CL dis-

criminentes é verificar a matriz estrutura, que fornece os coeficientes de correlação canônicos da CL. Esses valores são comparáveis às cargas dos fatores (veja o Capítulo 15) e indicam a natureza das CLs. Bargman (1970) argumenta que algumas variáveis dependentes apresentam altas correlações canônicas da CL enquanto outras têm valores baixos, as que têm valores altos contribuem mais para a separação do grupo. Desse modo, elas representam a contribuição relativa de cada variável dependente para a separação do grupo (veja Bray e Maxwell, 1985, p. 42-45). Novamente, estamos interessados somente na primeira CL (porque a segunda não foi significativa) e, olhando para a matriz de estrutura, podemos concluir que o número de comportamentos foi levemente mais importante em diferenciar os três grupos (porque 0,711 é maior do que 0,576). Contudo, o número de pensamentos é ainda muito importante porque o valor da correlação é bastante alto. Como ocorreu com as ponderações padronizadas, o fato que uma variável dependente tem uma correlação positiva e a outra tem uma negativa indica que a separação do grupo é determinada pela diferença entre as variáveis dependentes.

A próxima parte da saída (Saída 14.11 do SPSS) informa os coeficientes da função discriminante canônica, que são versões não-padronizadas dos coeficientes padronizados descritos acima. Desse modo, esses são os valores de *b* na equação (14.4) e você irá notar que eles correspondem aos valores nos autovetores derivados na Seção 14.3.4.1 e utilizados na equação (14.5). Esses valores não são tão úteis quanto à versão padronizada, mas mostram de onde surgiu a versão padronizada. A próxima tabela fornece as centroides para cada grupo. As centroides são simplesmente a média dos escores da CL para cada grupo. Devemos olhar o sinal da centroide (positivo ou negativo), e desses dados é possível ver que a CL 1 discrimina o grupo TC (BT) dos outros dois (notavelmente o grupo TCC (CBT) porque as diferenças entre centroides são as maiores para esses grupos). A segunda CL (que não foi significativa) parece discriminar o grupo ST

Saída 14.11 do SPSS

Canonical Discriminant Function Coefficients
(Coeficientes da função discriminante canônica)

	Function (Função)	
	1	2
Number of obsession-related behaviours (Número de comportamentos obsessivos)	0.603	0.425
Number of obsession-related thoughts (Número de pensamentos obsessivos)	-0.335	0.339
(Constant) (Constante)	2.139	-6.857

Functions at Group Centroids (Funções nas Centroides dos Grupos)

Group (Função)	Function (Função)	
	1	2
CBT (TCC)	0.601	-0.229
BT (TC)	-0.726	-0.128
No Treatment Control (Controle - Sem Tratamento)	0.125	0.357

Funções discriminante canônica não-padronizadas avaliadas na média dos grupos

(NT) dos dois grupos-experimentais (mas não significativamente). O relacionamento entre as CLs e os grupos fica mais bem ressaltado utilizando um diagrama de grupos combinados (solicitado pela utilização da caixa de diálogo como a da Figura 14.7). Esse gráfico apresenta os escores das CLs para cada pessoa, agrupados de acordo com a condição experimental a qual a pessoa pertence. Além disso, as centroides dos grupos são ressaltadas como o escore médio da CL para cada grupo.

A Figura 14.9 mostra esse diagrama para os dados do TOC, e fica claro a partir da posição das centroides (os grandes círculos rotulados com as iniciais do grupo) que a CL 1 discrimina o grupo TC do grupo TCC (olhe para a distância horizontal entre essas centroides). A segunda CL não diferencia qualquer grupo: já sabíamos disso porque ela foi não-significativa, mas o diagrama mostra que as distâncias verticais entre as centroides dos grupos são pequenas, indicando que não existe separação entre grupos para essa CL.

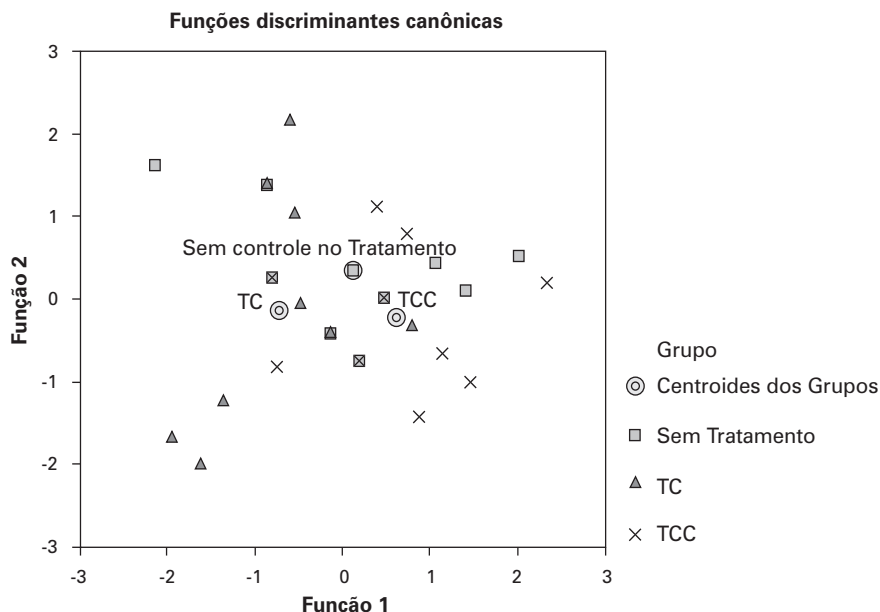


Figura 14.9 Diagrama combinado dos grupos.

Dica da Samanta Ferrinho



- A Análise da Função Discriminante (AFD) pode ser utilizada após a MANOVA para ver como as variáveis dependentes discriminam os grupos.
- A AFD identifica CLs (combinações das variáveis dependentes); para descobrir quantas CLs são significativas, olhe para as tabelas *Lambda de Wilks*: se o valor *Sig.* é menor do que 0,05, a CL discrimina significativamente os dois grupos.
- Uma vez que as CLs significativas tenham sido identificadas, utilize a tabela *Coefficientes padronizados da função discriminante canônica* (Standardized Canonical Discriminant Function Coefficients) para descobrir como as variáveis dependentes contribuem para as CLs. Escores altos indicam que uma variável dependente é importante para a CL e variáveis com coeficientes positivos ou negativos estão contribuindo para a CL de forma oposta.
- Finalmente, para descobrir que grupos são discriminados pela CL, olhe para a tabela *Funções nas Centroides dos Grupos* (Functions at Group Centroids): para uma dada CL, grupos com valores de sinais opostos estão sendo discriminados por essa CL.

14.9 ALGUMAS CONSIDERAÇÕES FINAIS ④

14.9.1 A interpretação final ④

Até agora, colhemos muitas informações sobre os nossos dados, mas como podemos

juntar todas para poder responder nossa questão de pesquisa: a terapia pode melhorar o TOC, e se sim qual delas é a melhor? A MANOVA, nos informa que a terapia pode ter um efeito significativo nos sintomas do TOC (Transtorno Obsessivo-Compulsivo), mas as

ANOVAs univariadas não-significativas sugeriram que essa melhoria não é simplesmente em termos de pensamentos ou comportamentos. A análise discriminante sugere que a separação dos grupos pode ser melhor explicada em termos de uma dimensão subjacente. Nesse contexto, a dimensão é provavelmente o próprio TCC (o qual podemos presumir ser constituído tanto de pensamentos quanto de comportamentos). Assim, a terapia não necessariamente muda comportamentos ou pensamentos *per se*, mas ela influencia a dimensão subjacente do TOC. Assim, a resposta para a primeira questão parece ser: sim, a terapia pode influenciar o TOC, mas a natureza dessa influência não está clara.

A próxima questão é mais complexa: qual terapia é a melhor? A Figura 14.10 mostra os gráficos dos relacionamentos entre as variáveis dependentes e as médias dos grupos dos dados originais. O gráfico das médias mostra que, para ações, a TC (Terapia Comportamental) reduz o número de comportamentos obsessivos, enquanto a TCC (Terapia Cognitivo Comportamental) e ST (Sem Tratamento), não. Para pensamentos, a TCC reduz o número de pensamentos obsessivos, enquanto a TC e ST, não (confira o padrão das barras). Olhando agora para o relacionamento entre pensamentos e ações no grupo de TC, existe relacionamento positivo quase linear entre pensamentos e ações, assim, quanto mais pensamentos obsessivos a pessoa apresenta, mais comportamentos obsessivos ela terá. No grupo TCC não existe relacionamento (pensamentos e ações variam de forma totalmente independente). No grupo dos sem tratamento (ST), existe um relacionamento negativo (e incidentalmente não-significativo) entre pensamentos e ações. Descobrimos, da análise discriminante, que os comportamentos são mais importantes em termos do TOC (como um construto) que pensamentos (sabemos disso a partir das correlações da CL canônica). Entretanto, a terapia comportamental parece ser a melhor porque ela se direciona a comportamentos em vez de pensamentos (comparada à TCC ela é preferível – a despeito da distância entre as

centroídes dos grupos na Figura 14.9). Contudo, a significância da função discriminante não necessariamente nos diz que o grupo TC foi significativamente menor que o grupo ST (assim, fazer terapia não é necessariamente melhor do que não fazer terapia). Resumindo, a TC tem a maior influência no TOC como um construto, em virtude da importância relativa dos comportamentos nesse construto comparado à cognição.

14.9.2 ANOVA univariada ou Análise discriminante? ③

Esse exemplo deve ter tornado claro que a ANOVA univariada e a análise discriminante são formas de responder diferentes questões que surgem de uma MANOVA significativa. Se a ANOVA univariada for escolhida, as correções de Bonferroni devem ser aplicadas no nível de aceitação ou significância. Você deve executar as duas análises para ter uma visão ampla do que está acontecendo com os dados. A vantagem da análise discriminante é que ela lhe informa algo sobre as dimensões subjacentes dentro dos dados (o que é especialmente útil se você tiver utilizado várias medidas dependentes numa tentativa de capturar algum construtor social ou psicológico). Mesmo que a ANOVA univariada seja significativa, a análise discriminante fornece algum conhecimento adicional dos dados e deve ser utilizada. Espero que este capítulo o tenha convencido dessa recomendação!

14.10 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

Neste capítulo, discutimos obsessivamente sobre a MANOVA e a análise da função discriminante, e descobrimos, para nosso horror, que o Roy tem uma raiz grande. Descobrimos que para situações em que várias saídas foram medidas em grupos diferentes, a técnica da ANOVA pode ser estendida e é denominada MANOVA (Análise de Variância Multivariada). Utilizamos essa técnica em vez de executar várias ANOVAs porque

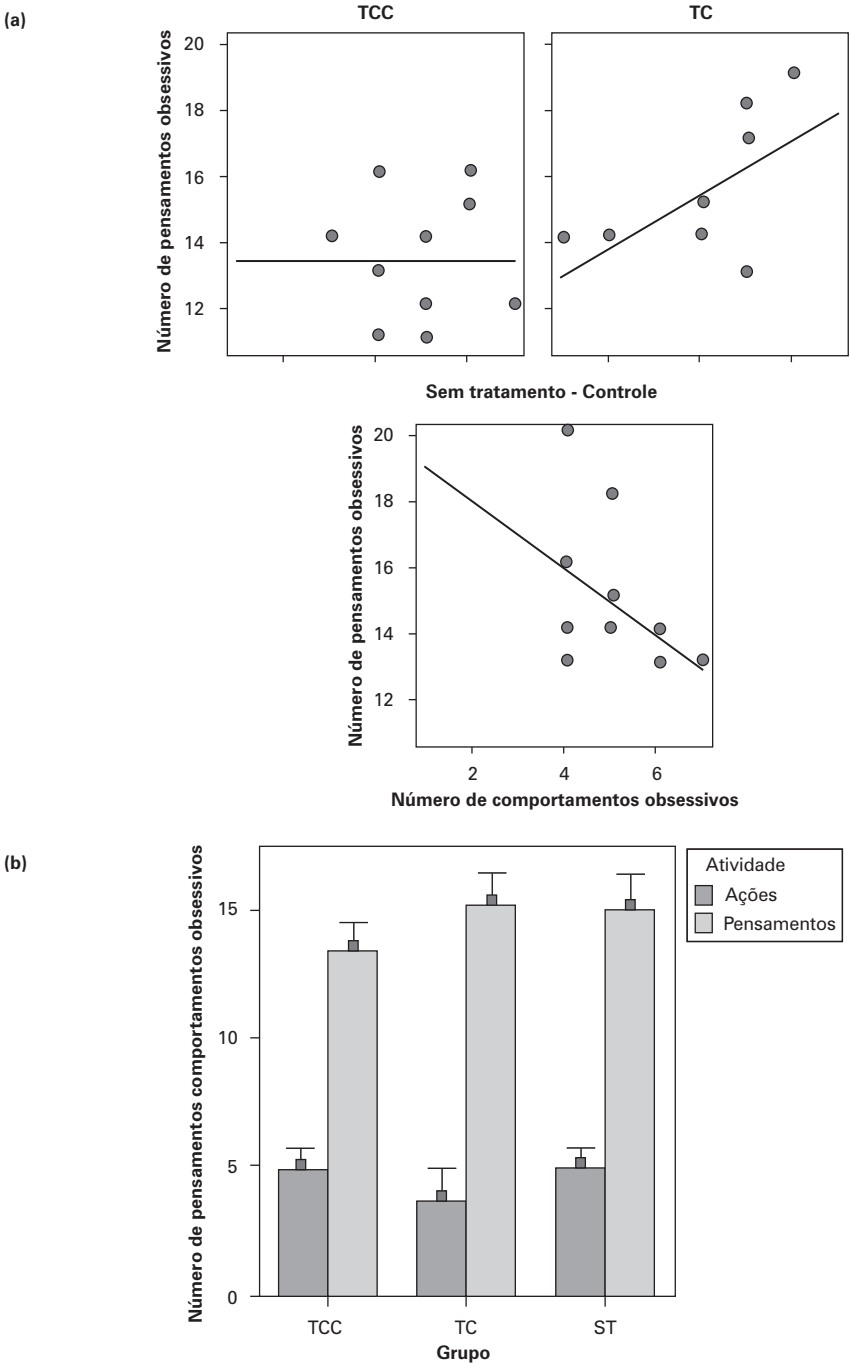


Figura 14.10 Gráficos mostrando (a) os relacionamentos e (b) as médias entre as variáveis dependentes em cada grupo da terapia.

assim mantemos controle sobre a taxa de erro do Tipo I e podemos incorporar o relacionamento entre variáveis de saída na análise. A MANOVA funciona de forma muito parecida à ANOVA, mas com matrizes em vez de números. Apresentamos um exemplo de MANOVA no SPSS e descobrimos que, para tornar a vida ainda mais complicada do que já é, obtemos não uma, mas quatro estatísticas relacionadas ao mesmo efeito! Dessas, tentei convencê-los de que o traço de Pillai é a opção mais segura. Finalmente, analisamos duas opções para dar prosseguimento à MANOVA: executar várias ANOVAs ou uma análise da função discriminante. Dessas, a análise da função discriminante é a que nos fornece mais informação, mas pode ser difícil de interpretar.

14.11 TERMOS-CHAVE QUE DESCOBRIMOS

- Testes de esfericidade de Bartlett
- Teste de Box
- Análise discriminante
- Escores discriminantes
- SSCP do erro (E)
- HE^{-1}
- Homogeneidade da matriz de covariâncias
- Traço de Hotelling-Lawley (T^2)
- SSCP das Hipóteses (H)
- Matriz Identidade
- Matriz
- Multivariada
- Análise Multivariada da Variância (ou MANOVA)
- Normalidade multivariada
- Traço de Pillai-Bartlett (V)
- Maior raiz de Roy
- Matriz quadrada
- Matrizes da soma dos quadrados e dos produtos cruzados (SSCP)
- SSCP total (T)
- Univariada
- Matriz de variâncias-covariâncias
- Lambda de Wilks

14.12 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Um psicólogo clínico notou que vários de seus pacientes maníacos psicóticos personificavam galinhas em público. Ele quis saber se esse comportamento poderia ser utilizado para diagnosticar esse transtorno e então decidiu comparar seus pacientes com uma amostra de pessoas normais. Ele observou 10 de seus pacientes durante um dia normal. Ele também precisou observar as 10 pessoas mais normais que conseguisse achar: naturalmente, ele escolheu observar professores da Universidade de Sussex. Ele mensurou todos os participantes utilizando duas variáveis dependentes: a primeira foi o número de personificações que eles fizeram nas ruas de Brighton durante um dia e a segunda avaliou o quanto foi bom o desempenho (medido numa escala até 10 por um criador de galinhas experiente). Os dados estão no arquivo **chicken.sav**; utilize a MANOVA e a AD (Análise discriminante) a fim de verificar se essas variáveis podem ser utilizadas para distinguir pacientes maníacos psicóticos daqueles que não tem o transtorno.
- **Tarefa 2:** Estava interessado em saber se o conhecimento dos diferentes aspectos da psicologia melhorava ao longo do curso. Peguei uma amostra de alunos do primeiro, segundo e terceiro ano e apliquei a eles cinco testes (numa escala até 15) envolvendo diferentes aspectos da psicologia: **Exper** (psicologia experimental, como psicologia cognitiva e neuropsicologia); **Stats** (Estatística); **Social** (psicologia social); **Develop** (Psicologia do desenvolvimento); **Person** (personalidade). Sua tarefa é: (1) executar uma análise geral apropriada para determinar se existe uma diferença global entre os grupos ao longo dessas cinco medidas; (2) verificar os valores das diferenças dos grupos

nas saídas e interpretar os resultados de acordo; (3) selecionar contrastes para testar a hipótese de que o segundo e o terceiro ano terão resultados melhores, em todas as escalas, do que o primeiro ano; (4) selecionar testes que comparem todos os grupos entre si – compare brevemente esses resultados com os dos contrastes; e (5) executar uma análise separada na qual você testa se a combinação das medidas pode discriminar os grupos com sucesso (comente brevemente essa análise). Inclua somente aquelas escalas que revelaram diferenças entre os grupos para os

contrastes. Como esses resultados poderão ajudá-lo a explicar os achados da sua análise inicial? Os dados estão no arquivo **psychology.sav**. ④

14.13 LEITURAS COMPLEMENTARES

BRAY, J. H., MAXWELL, S. E. *Multivariate analysis of variance*. Sage university paper series on quantitative applications in the social sciences. Newbury Park (CA): Sage. 1985, p. 07-54. Essa monografia sobre a MANOVA é soberba; nada há de melhor que eu possa recomendar.

ANÁLISE DE FATORES EXPLORATÓRIA

15.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ②

Nas ciências sociais, geralmente tentamos medir coisas que não podem ser medidas diretamente (as chamadas *variáveis latentes ou construtos*). Por exemplo, estudantes de administração (ou mesmo psicólogos) podem estar interessados em medir a exaustão (*burnout*), que é quando alguém que tem trabalhado muito num projeto (um livro, por exemplo) por um longo período de tempo subitamente se encontra destituído de motivação e inspiração e quer bater com a cabeça várias vezes no computador. Você não pode medir a exaustão diretamente: ela tem muitas facetas. Entretanto, pode medir vários aspectos da exaustão: a motivação, o nível de estresse, se a pessoa tem alguma ideia nova e assim por diante. Feito isso, seria útil saber se essas medidas realmente refletem uma única variável. Colocando de outra forma, essas diferentes variáveis são manifestações de uma mesma variável básica? Este capítulo irá apresentar a *análise de fatores* (e a *análise de componentes principais*), uma técnica para identificar grupos ou agrupamentos de variáveis. Essa técnica tem três usos principais: (1) entender a estrutura de um conjunto de variáveis (por exemplo, pioneiros

da inteligência como Spearman e Thurstone utilizaram a análise de fatores para tentar entender a estrutura da variável latente “inteligência”); (2) construir um questionário para medir uma variável subjacente (por exemplo, você pode fazer um questionário para medir a exaustão); e (3) reduzir um conjunto de dados a um tamanho mais manejável enquanto se retém o máximo da informação original possível (por exemplo, vimos no Capítulo 5 que a multicolinearidade pode ser um problema na regressão múltipla e a análise de fatores pode ser utilizada para resolver esse problema combinando variáveis colineares). Ao longo deste capítulo, iremos descobrir o que são fatores, como os encontramos e o que eles nos informam (se informam algo) sobre o relacionamento entre as variáveis que medimos.

15.2 FATORES ②

Se medirmos diversas variáveis, ou fizermos várias perguntas para alguém sobre eles mesmos, a correlação entre cada par de variáveis (ou perguntas) pode ser organizada em uma *matriz-R*. Uma *matriz-R* é uma matriz de cor-



relações: uma tabela de coeficientes de correlações entre variáveis (vimos versões menores dessa matriz no Capítulo 4). Os elementos diagonais de uma matriz-R são iguais a 1, porque cada variável irá se correlacionar perfeitamente com ela mesma. Os elementos fora da diagonal são os coeficientes de correlação entre os pares de variáveis ou questões.¹ A existência de vários coeficientes de correlação altos entre subconjuntos de variáveis sugere que essas variáveis podem estar medindo aspectos de uma mesma dimensão subjacente. Essas dimensões subjacentes são conhecidas como *fatores* (ou *variáveis latentes*). Pela redução de um conjunto de dados a partir de um grupo de variáveis inter-relacionadas em um conjunto menor, a análise de fatores obtém a parcimônia explicando a quantidade máxima da variância comum em uma matriz de correlação utilizando um número menor de conceitos explanatórios.

Existem inúmeros exemplos do uso da análise de fatores nas ciências sociais. Os teóricos da personalidade em psicologia usam a análise de fatores para medir traços de personalidade. A maioria dos leitores está familiarizada com a extroversão-introversão e traços neuróticos medidos por Eysenck (1953). Muitos outros questionários sobre a personalidade estão baseados na análise de fatores (notadamente, Cattell's, 1966a, questionário de 16 fa-

tores de personalidade), e esses levantamentos são frequentemente usados para propósitos de recrutamento em indústrias e até em alguns grupos religiosos. Entretanto, embora a análise de fatores seja mais conhecida por ser adotada por psicólogos, seu uso, de maneira alguma, está restrito à mensuração de dimensões da personalidade. Economistas, por exemplo, podem usar a análise de fatores para verificar se a produtividade, lucro e força de trabalho podem ser reduzidos a uma dimensão subjacente do crescimento da companhia, e Jeremy Miles me contou que um bioquímico usou a análise de fatores para analisar amostras de urina!

Vamos colocar algumas dessas ideias em prática imaginando que queremos medir os diferentes aspectos que tornam uma pessoa popular. Podemos aplicar várias medidas que acreditamos explorar diferentes aspectos da popularidade. Assim, podemos medir as habilidades sociais de uma pessoa (*Social Skills*), seu egoísmo (*Selfish*), quão interessante os outros a acham (*Interest*), a proporção de tempo que gastam falando sobre a outra pessoa durante uma conversa (*Talk 1*), a proporção de tempo que gastam falando delas mesmas (*Talk 2*) e sua propensão para mentir a outras pessoas (*the Liar Scale*). Podemos calcular os coeficientes de correlação para cada par de variáveis e criar uma matriz-R. A Tabela 15.1 mostra essa matriz. Todos os coeficientes de correlação significativos estão em negrito. Está claro que existem dois conjuntos de variáveis interrelacionadas. Portanto, essas variáveis podem estar medindo algumas dimen-

¹ Essa matriz é chamada de matriz-R, ou R, porque ela contém coeficientes de correlação e o *r* geralmente representa o coeficiente de correlação de Pearson (veja o Capítulo 4) – o *r* se torna maiúsculo quando representa uma matriz.

Tabela 15.1 Uma matriz-R

	Conversa 1 (Talk 1)	Habilidades Sociais (Social Skills)	Interessante (Interest)	Conversa 2 (Talk 2)	Egoísta (Selfish)	Mentiroso (Liar)
Conversa 1 (Talk 1)	1,000					
Habilidades Sociais (Social Skills)	0,772	1,000				
Interesse (Interest)	0,646	0,879	1,000			
Conversa 2 (Talk 2)	0,074	-0,120	0,054	1,000		
Egoísta (Selfish)	-0,131	0,031	-0,101	0,441	1,000	
Mentiroso (Liar)	0,068	0,012	0,110	0,361	0,277	1,000

sões comuns subjacentes. A quantidade de tempo que alguém fala de outra pessoa durante uma conversa parece se correlacionar com o nível de habilidade social e quão interessante os outros acham aquela pessoa. Habilidades sociais também se correlacionam com quão interessante os outros acham aquela pessoa. Esses relacionamentos indicam que quanto melhores suas habilidades sociais, mais interessante e tagarela é provável que você seja. Entretanto, existe um segundo conjunto de variáveis. A quantidade que as pessoas falam delas mesmas numa conversa se correlaciona com o quão egoísta elas são e o quanto elas mentem. Ser egoísta também se correlaciona com o grau que uma pessoa conta mentiras. Resumindo, é mais provável que pessoas egoístas mintam e falem sobre elas mesmas.

Na análise de fatores, nos esforçamos para reduzir a matriz-R à sua dimensão subjacente olhando para as variáveis que parecem se aglomerar de maneira significativa. Essa redução de dados é alcançada olhando as variáveis que se correlacionam altamente com um grupo de outras variáveis, mas que não se correlacionam com as variáveis fora daquele grupo. Nesse

exemplo, parece haver dois aglomerados que se encaixam nessa ideia. O primeiro fator parece se relacionar com a sociabilidade em geral, enquanto o segundo fator parece se relacionar com a maneira que a pessoa trata os outros socialmente (podemos chamá-lo de Consideração). Portanto, podemos assumir que a popularidade depende não somente da sua sociabilidade, mas também de se você é autêntico com os outros.

15.2.1 Representação gráfica dos fatores ②

Fatores (não confundir com variáveis independentes da ANOVA fatorial) são entidades estatísticas que podem ser visualizadas como eixos de um sistema de coordenadas onde as variáveis podem ser representadas. De forma simples, essa afirmação significa que se você imaginar que os fatores sejam os eixos de um sistema de coordenadas, é possível representar as variáveis ao longo desses eixos. As coordenadas das variáveis ao longo de cada eixo representam a força do relacionamento entre a variável e cada fator. A Figura 15.1 mostra o gráfico para os dados de popularidade (nos

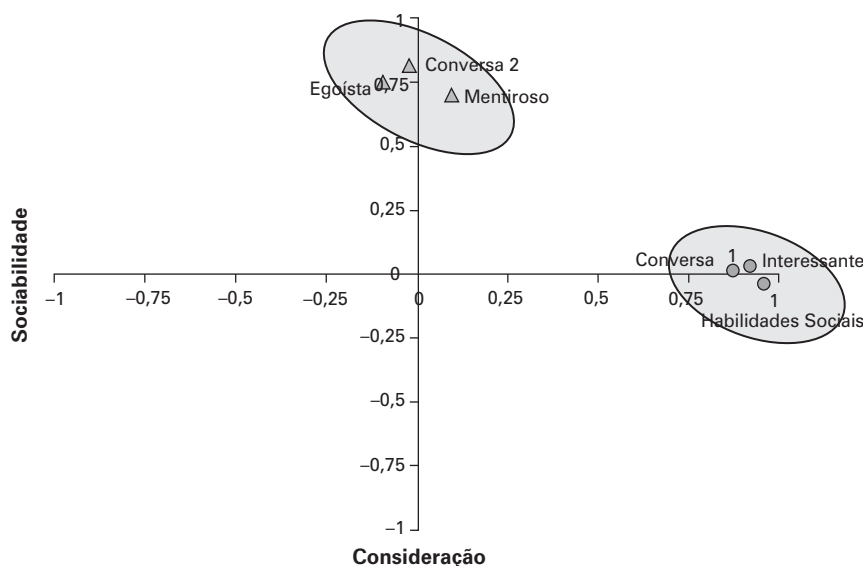


Figura 15.1 Exemplo de um diagrama de fatores.

quais existem somente dois fatores). Note que para ambos os fatores o eixo vai de -1 a 1 , que são os limites máximos dos coeficientes de correlação. Os círculos representam as três variáveis que se correlacionam altamente com o fator 1 (sociabilidade: eixo horizontal), mas tem uma baixa correlação com o fator 2 (consideração com os outros: eixo vertical). Contrariamente, os triângulos representam variáveis que se correlacionam altamente com consideração para com os outros, mas tem uma baixa correlação com sociabilidade. Desse gráfico, podemos afirmar que o egoísmo, a quantidade que a pessoa fala dela mesma, e a propensão à mentira contribuem para o fator de consideração para com os outros. Contrariamente, o quanto uma pessoa tem interesse por outras, quão interessante ela é e o seu grau de sociabilidade contribui para o segundo fator, sociabilidade. Esse diagrama, portanto, apoia a estrutura que estava aparente na matriz-R. É claro, se um terceiro fator existisse dentre esses dados, ele poderia ser representado por um terceiro eixo (criando um gráfico em 3-D). Também, se existirem mais do que três fatores em cada conjunto de dados, eles não podem ser representados por um desenho bidimensional.

Se cada eixo do gráfico representa um fator, as variáveis que irão compor um fator podem ser traçadas de acordo com a extensão à qual elas se relacionam com esse fator. As coordenadas de uma variável, portanto, representam seu relacionamento com os fatores. Num mundo ideal, uma variável deveria ter uma coordenada grande para um dos eixos e uma coordenada baixa para qualquer outro fator. Essa hipótese indicaria que essa variável específica se relacionou somente com um fator. Variáveis que tem grandes coordenadas no mesmo eixo supostamente medem diferentes aspectos de uma dimensão comum subjacente. **A coordenada de uma variável ao longo de um eixo de classificação é conhecida como carga do fator.** A carga de um fator pode ser pensada como a correlação de Paerson entre o fator e a variável (veja o Quadro 15.1). Do que já sabemos sobre a interpretação do coeficiente de correlação (veja a Seção 4.5.3) deve estar claro que

se elevarmos ao quadrado a carga do fator, obteremos uma medida da importância s de uma variável particular para um determinado fator.

15.2.2 Representação matemática dos fatores ②



Os eixos traçados na Figura 15.1 são linhas retas e, assim, podem ser descritas matematicamente por uma equação de uma linha reta. Portanto, fatores também podem ser descritos em termos dessa equação. Temos usado a equação de uma linha reta inúmeras vezes ao longo deste texto e, assim, você deve conhecê-la bem. A Equação (15.1) nos lembra da equação que descreve um modelo linear, e isso pode ser aplicado ao cenário da descrição de um fator. Você irá notar que não existe intercepto na equação: o motivo é que as linhas se interceptam no zero (portanto, o intercepto é, também, zero). Os b s na equação representam as cargas dos fatores:

$$Y_i = b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i$$

$$\text{Fator}_i = b_1\text{Variável}_1 + b_2\text{Variável}_2 + \dots + b_n\text{Variável}_n + \varepsilon_i \quad (15.1)$$

No nosso exemplo de popularidade, encontramos dois fatores subjacentes: sociabilidade e consideração. Podemos, portanto, construir uma equação que descreva cada fator em termos das variáveis que foram medidas. A equação é mostrada em (15.2):

$$Y_i = b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i$$

$$\begin{aligned} \text{Sociabilidade}_i = & b_1\text{Conversa1}_i + b_2\text{Habilidades} \\ & \text{Sociais}_i + b_3\text{Interesse}_i \\ & + b_4\text{Conversa2}_i + b_5\text{Egoísmo}_i \\ & + b_6\text{Mentiroso}_i + \varepsilon_i \end{aligned}$$

$$\begin{aligned} \text{Consideração}_i = & b_1\text{Conversa1}_i + b_2\text{Habilidades} \\ & \text{Sociais}_i + b_3\text{Interesse}_i \\ & + b_4\text{Conversa2}_i \\ & + b_5\text{Egoísmo}_i \\ & + b_6\text{Mentiroso}_i + \varepsilon_i \end{aligned} \quad (15.2)$$

Primeiro, note que as equações são idênticas na forma: ambas incluem todas as variá-

veis que medimos. Entretanto, os valores de b nas duas equações serão diferentes (dependendo da importância relativa de cada variável para um fator particular). Na verdade, podemos substituir cada valor de b com a coordenada daquela variável no gráfico da Figura 15.1 (isto é, substitua os valores de b pelas cargas dos fatores). As equações resultantes são mostradas em (15.3):

$$Y_i = b_1X_1 + b_2X_2 + \dots + b_nX_n + \varepsilon_i$$

Sociabilidade_{*i*} = 0,87 Conversa 1_{*i*}
+ 0,96 Habilidades Sociais_{*i*}
+ 0,92 Interesse_{*i*}
+ 0,00 Conversa2_{*i*}
- 0,10 Egoísmo_{*i*}
+ 0,09 Mentiroso_{*i*} + ε_i

Consideração_{*i*} = 0,01 Conversa 1_{*i*}
+ 0,03 Habilidades Sociais_{*i*}
+ 0,04 Interesse_{*i*}
+ 0,82 Conversa2_{*i*}
+ 0,75 Egoísmo_{*i*}
+ 0,70 Mentiroso_{*i*} + ε_i (15.3)

Note que, para o fator sociabilidade, os valores de b são altos para Conversa 1, Habilidades Sociais e Interesse. Para as variáveis restantes (Conversa 2, Egoísmo e Mentiroso) os valores de b são baixos (próximo a zero). Isso nos diz que três das variáveis são muito importantes para aquele fator (aquelas com valores altos de b) e que as outras três não são muito importantes (aquelas com valores de b baixos). Vimos que isso é verdadeiro pela maneira com

que as três variáveis estavam altamente aglomeradas no gráfico do fator. O que temos que levar em conta aqui é que o gráfico do fator e essas equações representam o mesmo: as cargas dos fatores no gráfico são simplesmente os valores de b nessas equações (veja o Quadro 15.1). Para o segundo fator, desconsiderando os demais, o padrão oposto pode ser visto, isto é, Conversa 2, Egoísmo e Mentiroso tem valores altos de b , enquanto as três variáveis restantes têm valores de b próximos a zero. Num mundo ideal, variáveis teriam valores de b bem altos para um fator e valores de b bem baixos para os demais fatores.

Essas cargas dos fatores podem ser colocadas numa matriz na qual as colunas representam os fatores e as linhas representam as cargas de cada variável em relação a cada fator. Para os dados de popularidade, essa matriz teria duas colunas (uma para cada fator) e seis linhas (uma para cada variável). Essa matriz, normalmente representada por A , pode ser vista abaixo. Para entender o que a matriz significa, tente relacionar os elementos da mesma com as cargas vistas na equação (15.3). Por exemplo, a linha de cima representa a primeira variável, Conversa 1, que tem uma carga de 0,87 para o primeiro fator (Sociabilidade) e uma carga de 0,01 para o segundo fator (Consideração). Essa matriz é chamada de *matriz dos fatores* ou *matriz dos componentes* (se estiver fazendo *análise de componentes principais*) – veja o Quadro 15.1 para descobrir as diferentes formas dessa matriz.

Quadro 15.1

Qual é a diferença entre uma matriz padrão e uma matriz estruturada? ③

Fui bastante vago na discussão sobre a carga dos fatores. Algumas vezes, afirmei que essas cargas podem ser pensadas como a correlação entre a variável e um dado fator e, em outras ocasiões, descrevi essas cargas em termos de coeficientes da regressão (b). Deveria ser óbvio a partir do que descobrimos nos Capítulos 4 e 5 que os coeficientes de correlação e de regressão são duas coisas bem diferentes, assim, do que, de fato, estou falando? Eu não deveria a essa altura saber o que essas cargas dos fatores são?

(Continua)

Quadro 15.1 (Continuação)

Bem, tanto o coeficiente de correlação quanto o de regressão representam o relacionamento entre uma variável e o modelo linear num sentido amplo, assim, a mensagem principal aqui é que as cargas dos fatores nos informam o quanto uma variável contribui para formar um fator. Se entender isso, você não terá problemas.

Entretanto, a carga de um fator numa determinada análise pode ser tanto um coeficiente de correlação quanto um de regressão. Após algumas seções iremos descobrir que a interpretação da análise de fatores melhora muito com uma técnica chamada de *rotação*. Sem entrar em detalhes, existem dois tipos de rotação: a ortogonal e a oblíqua (veja a Seção 15.3.6). Quando a rotação ortogonal é usada, qualquer fator subjacente é assumido como independentes e a carga do fator é a correlação entre o fator e a variável, mas é, também, o coeficiente de regressão. Colocando de outra maneira, os valores dos coeficientes de correlação são os mesmos que os valores dos coeficientes de regressão. Entretanto, existem situações em que os fatores subjacentes são supostamente relacionados ou correlacionados entre si. Nessas situações, a rotação oblíqua é usada e as correlações resultantes entre variáveis e fatores irão diferir dos correspondentes coeficientes regressão. Nesse caso, existem, de fato, dois conjuntos diferentes de cargas de fatores: os coeficientes de correlação entre cada variável e o fator (que estão na *matriz de estrutura dos fatores*) e os coeficientes de regressão para cada variável em cada fator (que estão na *matriz padrão dos fatores*). Esses coeficientes podem ter interpretações bem diferentes (veja Graham, Guthrie e Thompson, 2003).

$$A = \begin{pmatrix} 0,87 & 0,01 \\ 0,96 & -0,03 \\ 0,92 & 0,04 \\ 0,00 & 0,82 \\ -0,10 & 0,75 \\ 0,09 & 0,70 \end{pmatrix}$$

A hipótese principal da análise de fatores é que esses fatores algébricos representam dimensões reais, cuja natureza deve ser buscada inspecionando as variáveis que apresentam as maiores cargas no mesmo fator. Assim, psicólogos acreditam que fatores representam dimensões da psique, pesquisadores da educação acreditam que elas representam habilidades e sociólogos podem achar que elas representam raças ou classes sociais. Entretanto, essa hipótese é controversa e alguns acreditam que as dimensões derivadas da análise de fatores são reais somente no sentido estatístico – e uma ficção no mundo real.

15.2.3 Escores dos fatores ②

Um fator pode ser descrito em termos das variáveis mensuradas e da importância relativa delas para aquele fator (representado pelo

valor de b). Portanto, tendo descoberto quais fatores existem e estimado a equação que os descreve, deveria ser possível também estimar o escore de uma pessoa num fator baseado nos seus escores para as variáveis constituintes. Assim, se queremos derivar um escore de sociabilidade para uma determinada pessoa, podemos colocar seus escores nas várias medidas na equação (15.3). Esse método é conhecido como *média ponderada*. Na verdade, esse método é muito simplista e raramente usado, mas é provavelmente a maneira mais fácil de explicar o princípio. Por exemplo, imagine que temos seis escalas variando de 1 a 10 e que alguém obteve os seguintes valores: Conversa 1 (4), Habilidades Sociais (9), Interesse (8), Conversa 2 (6), Egoísmo (8), e Mentiroso (6). Podemos entrar com esses valores na equação (15.3) para obter um escore da sociabilidade da pessoa e sua consideração pelos outros (veja a Equação (15.4)). Os valores resultantes de 19,22 e 15,21 refletem o grau de sociabilidade da pessoa e sua consideração para com os outros, respectivamente. Essa pessoa obteve um escore maior em sociabilidade do que na consideração pelos outros. Entretanto, as escalas de medidas utilizadas irão influenciar

os resultados obtidos e se diferentes variáveis foram medidas com escalas diferentes, os escores para diferentes fatores não poderão ser comparados. Portanto, o método de calcular os escores dos fatores é pobre e métodos mais sofisticados são normalmente utilizados.

$$\begin{aligned}\text{Sociabilidade} &= 0,87 \text{ Conversa1} + 0,96 \\ &\quad \text{Habilidades Sociais} \\ &\quad + 0,92 \text{ Interesse}_i \\ &\quad + 0,00 \text{ Conversa2} \\ &\quad - 0,10 \text{ Egoísmo} \\ &\quad + 0,09 \text{ Mentiroso}_i\end{aligned}$$

$$\begin{aligned}\text{Sociabilidade} &= (0,87 \times 4) + (0,96 \times 9) \\ &\quad + (0,92 \times 8) + (0,00 \times 6) \\ &\quad - (0,10 \times 8) + (0,09 \times 6) \\ &= 19,22\end{aligned}$$

$$\begin{aligned}\text{Consideração} &= 0,01 \text{ Conversa1} \\ &\quad + 0,03 \text{ Habilidades Sociais} \\ &\quad + 0,04 \text{ Interesse} \\ &\quad + 0,82 \text{ Conversa2} \\ &\quad + 0,75 \text{ Egoísmo} \\ &\quad + 0,70 \text{ Mentiroso}_i\end{aligned}$$

$$\begin{aligned}\text{Consideração} &= (0,01 \times 4) - (0,03 \times 9) \\ &\quad + (0,04 \times 8) + (0,82 \times 6) \\ &\quad + (0,75 \times 8) + (0,70 \times 6) \\ &= 15,21\end{aligned}\quad (15.4)$$

15.2.3.1 O método da regressão ④

Existem várias técnicas sofisticadas para calcular os escores dos fatores que utilizam os coeficientes dos fatores como pesos ou ponderações na equação (15.1) em vez de utilizarem as cargas dos fatores. A forma da equação permanece a mesma, mas os *bs* da equação são substituídos pelos coeficientes dos fatores. Os coeficientes dos fatores podem ser calculados de várias formas. A mais simples é o método da regressão. Nesse método, as cargas dos fatores são ajustadas para levar em conta as correlações iniciais entre as variáveis e, ao fazer isso, as diferenças nas unidades de medidas e as variâncias das variáveis são estabilizadas.

Para obter a matriz dos coeficientes dos fatores (*B*), multiplicamos a matriz das cargas dos fatores pela inversa (R^{-1}) original das correlações ou matriz-*R*. No capítulo anterior,

vimos que matrizes não podem ser divididas (Seção 14.3.4.1). Dessa forma, não podemos dividir por uma matriz diretamente, precisamos multiplicar pela sua inversa. Portanto, ao multiplicarmos a matriz da carga dos fatores pela inversa da matriz das correlações, estamos dividindo as cargas dos fatores pelos coeficientes de correlação. A matriz dos escores dos fatores resultante, dessa forma, representa o relacionamento entre cada variável e cada fator levando em conta o relacionamento original entre os pares de variáveis. Desse modo, essa matriz representa uma medida mais pura da relação única entre variáveis e fatores.

As matrizes para os dados da popularidade são mostradas a seguir. A matriz dos coeficientes dos fatores, *B*, é fornecida pelo SPSS. As matrizes R^{-1} e *A* podem ser multiplicadas manualmente para a obtenção da matriz *B*; quem entende de álgebra matricial (ou quem consultou Namboodiri, 1984, ou Stevens, 1992) podem verificar o resultado (os cálculos estão no arquivo **FactorScores.pdf** disponível no site www.artmed.com.br). Para obter o mesmo resultado que o SPSS, será necessário trabalhar com pelo menos cinco casas decimais.

$$\begin{aligned}B &= R^{-1}A \\ B &= \begin{pmatrix} 4,76 & -7,46 & 3,91 & -2,35 & 2,42 & -0,49 \\ -7,46 & 18,49 & -12,42 & 5,45 & -5,54 & 1,22 \\ 3,91 & -12,42 & 10,07 & -3,65 & 3,79 & -0,96 \\ -2,35 & 5,45 & -3,65 & 2,97 & -2,16 & 0,02 \\ 2,42 & -5,54 & 3,79 & -2,16 & 2,98 & -0,56 \\ -0,49 & 1,22 & -0,96 & 0,02 & -0,56 & 1,27 \end{pmatrix} \\ &\quad \begin{pmatrix} 0,87 & 0,01 \\ 0,96 & -0,03 \\ 0,92 & 0,04 \\ 0,00 & 0,82 \\ -0,10 & 0,75 \\ 0,09 & 0,70 \end{pmatrix} \\ &= \begin{pmatrix} 0,343 & 0,006 \\ 0,376 & -0,020 \\ 0,362 & 0,020 \\ 0,000 & 0,473 \\ -0,037 & 0,437 \\ 0,039 & 0,405 \end{pmatrix}\end{aligned}$$

O padrão das cargas é o mesmo dos coeficientes dos fatores; isto é, as primeiras três variáveis têm altas cargas no primeiro fator e

baixas cargas no segundo, enquanto o padrão é inverso para as três últimas variáveis. A diferença está somente nos valores das ponderações, que são menores porque as correlações entre as variáveis estão, agora, sendo levadas em conta. Esses coeficientes dos fatores podem ser utilizados para substituir os valores b na equação (15.4):

$$\begin{aligned}
 \text{Sociabilidade} &= 0,343 \text{ Conversa1} \\
 &\quad + 0,376 \text{ Habilidades Sociais} \\
 &\quad + 0,362 \text{ Interesse} \\
 &\quad + 0,000 \text{ Conversa2} \\
 &\quad - 0,037 \text{ Egoísmo} \\
 &\quad + 0,039 \text{ Mentiroso} \\
 \text{Sociabilidade} &= (0,343 \times 4) + (0,376 \times 9) \\
 &\quad + (0,362 \times 8) + (0,000 \times 6) \\
 &\quad - (0,037 \times 8) + (0,039 \times 6) \\
 &= 7,59 \qquad (15.5) \\
 \text{Consideração} &= 0,006 \text{ Conversa1} \\
 &\quad - 0,020 \text{ Habilidades Sociais} \\
 &\quad + 0,020 \text{ Interesse} \\
 &\quad + 0,473 \text{ Conversa2} \\
 &\quad + 0,437 \text{ Egoísmo} \\
 &\quad + 0,405 \text{ Mentiroso} \\
 \text{Consideração} &= (0,006 \times 4) - (0,020 \times 9) \\
 &\quad + (0,020 \times 8) + (0,020 \times 6) \\
 &\quad + (0,473 \times 6) + (0,437 \times 8) \\
 &\quad + (0,405 \times 6) \\
 &= 8,768
 \end{aligned}$$

A equação (15.5) mostra como esses escores dos coeficientes são utilizados para produzir dois fatores para cada pessoa. Nesse caso, o participante apresenta os mesmos escores em cada variável que foram utilizados na equação (15.4). Os escores resultantes são mais parecidos do que quando as cargas dos fatores foram utilizadas como ponderações porque as diferentes variâncias entre as seis variáveis agora estão sendo controladas. O fato dos valores serem bem semelhantes reflete que essa pessoa não apenas apresenta um escore alto em relação às variáveis representando sociabilidade, mas ela apresenta também pouca consideração para com os outros (isto é, ela apresenta altos escores igualmen-

te nos dois fatores). Essa técnica de produzir escores de fatores assegura que os resultados tenham uma média de 0 e uma variância igual ao quadrado do coeficiente de correlação múltipla entre os escores dos fatores estimados e os verdadeiros valores dos fatores. Contudo, o problema com o método de regressão é que os escores podem estar correlacionados não apenas com outros fatores além daquele em que eles foram baseados, mas também com outros escores de diferentes fatores ortogonais.

15.2.3.2 Outros métodos ②



Para superar os problemas associados à técnica de regressão, dois ajustes foram propostos: o *método Bartlett* e o *método Anderson-Rubin*. O SPSS

pode produzir escores dos fatores baseados em qualquer um desses métodos. O método Bartlett produz escores imparciais e que se correlacionam apenas com seu próprio fator. A média e o desvio padrão dos escores são os mesmos do método de regressão. Entretanto, os escores dos fatores podem, ainda, se correlacionar entre si. O método final é o de Anderson-Rubin, uma modificação do de Bartlett que produz escores dos fatores padronizados e não-correlacionados (eles apresentam uma média de 0 e um desvio padrão de 1). Tabachnick e Fidell (2001) concluem que o método de Anderson-Rubin é melhor quando escores não-correlacionados são requeridos, mas o método de regressão é preferido em outras circunstâncias simplesmente porque é o de mais fácil compreensão. Embora não seja crucial que você entenda os cálculos por trás dos métodos, é importante que você saiba o que os escores dos fatores representam: a saber, um escore múltiplo para cada indivíduo em um fator particular.

15.2.3.3 Usos dos escores dos fatores ②

Existem vários usos para os escores dos fatores. Primeiro, se o propósito da análise de fatores é reduzir um grande conjunto de dados em um subconjunto menor de variáveis, os escores dos fatores nos darão um escore indivi-

dual nesse subconjunto de medidas. Portanto, qualquer análise adicional pode ser realizada nos escores dos fatores em vez de nos dados originais. **Por exemplo, podemos conduzir um teste t para verificar se as mulheres são significativamente mais sociáveis do que os homens usando o escore dos fatores para *sociabilidade*. Um segundo uso é superar problemas de colinearidade na regressão. Se, examinando uma análise de regressão múltipla, identificamos fontes de multicolinearidade, a interpretação da análise é questionável (veja a Seção 5.6.2.3). Nessa situação, podemos simplesmente conduzir uma análise de fatores nas variáveis previstas para reduzi-las a um subconjunto de fatores não-correlacionados. As variáveis que causam a multicolinearidade irão se agrupar formando um fator.** Se executarmos novamente a regressão, mas utilizando os escores dos fatores como variáveis previsoras, o problema da multicolinearidade deverá desaparecer (porque as variáveis agora estão agrupadas em um único fator). Existem meios nos quais podemos assegurar que os fatores não estão correlacionados (uma maneira é usar o método Anderson-Rubin – veja acima). Usando fatores não-correlacionados como previsores na regressão, podemos estar confiantes de que não haverá correlação entre os previsores e, conseqüentemente, também não haverá multicolinearidade.

15.3 DESCOBRINDO FATORES ②

Você já aprendeu o que é um fator, como ele é representado graficamente e algebricamente e como podemos calcular escores compostos representando um “desempenho” individual num único fator. Restringi a discussão a um nível conceitual sem investigar como realmente encontramos essas bestas míticas conhecidas como fatores. Esta seção irá apresentar como encontramos os fatores. Especificamente, examinaremos vários métodos, observaremos os cálculos por trás de um método (componentes principais), investigaremos o critério para determinar se os fatores são importantes e descobrir como melhorar a interpretação de uma dada solução.

15.3.1 Escolhendo um método ②

Há vários métodos para extrair fatores dos seus dados. O método que você irá escolher dependerá do que você espera fazer com a análise. Tinsley e Tinsley (1987) fornecem um excelente relato dos diferentes métodos disponíveis. Existem duas coisas a serem consideradas: se você quer generalizar os resultados da sua amostra para a população e se está explorando seus dados ou testando uma hipótese específica. Este capítulo descreve técnicas para a exploração de dados utilizando a análise de fatores. Testar hipóteses sobre as estruturas de variáveis latentes e seus relacionamentos é consideravelmente complexo e pode ser feito com programas de computador específicos, como o AMOS. Quem estiver interessado em técnicas de testar hipóteses (conhecida como análise de fatores confirmatória), consulte Pedhazur e Schmelkin (1991, Capítulo 23) para uma introdução. Supondo que queremos explorar nossos dados, devemos considerar se queremos aplicar nossas descobertas à amostra coletada (método descritivo) ou generalizar os resultados para a população (método inferencial). A análise dos fatores foi originalmente desenvolvida para ser utilizada na exploração dos dados gerando hipóteses futuras. Como tal, foi suposto que a técnica seria aplicada para toda a população de interesse. Portanto, algumas técnicas supõem que a amostra usada é a população e, assim, os dados não podem ser extrapolados além daquela amostra particular. A análise de componentes principais é um exemplo de uma dessas técnicas, assim como a análise dos fatores principais (eixo principal de fatoraçoão) e a análise da imagem da covariância (fatoraçoão de imagem). Dessas, a análise dos componentes principais e a análise dos fatores principais são os métodos preferidos e geralmente resultam em soluções similares (veja a Seção 15.3.3). Quando esses métodos são usados, as conclusões estão restritas às amostras coletadas e a generalização pode ser feita somente se uma análise usando diferentes amostras revelar a mesma estrutura de fatores.

Outra abordagem tem sido presumir que os participantes são selecionados aleatoria-

mente e que as variáveis medidas constituem a população de variáveis na qual estamos interessados. Presumindo isso, é possível desenvolver técnicas em que os resultados podem ser generalizados para uma população maior. Entretanto, uma restrição é que qualquer resultado é verdadeiro somente para o conjunto das variáveis medidas (porque assumimos que esse conjunto constitui toda a população de variáveis). Técnicas nessa categoria incluem o método de máxima verossimilhança (veja Harman, 1976) e a fatoração Alfa de Kaiser. A escolha do método depende de que tipo de generalização desejamos fazer.²

15.3.2 Comunalidades ②

Antes de continuar, é importante que você entenda alguns aspectos básicos sobre a variância dentro de uma matriz-R. É possível calcular a variabilidade dos escores (a variância) para qualquer medida dada (ou variável). Você já deve estar familiarizado com a ideia de variância e seu cálculo (veja o Capítulo 1). A variância total de uma variável apresenta dois componentes: uma parte compartilhada com outras variáveis ou medidas (variância comum) e outra parte específica daquela medida (variância única). Iremos utilizar o termo “variância única” para nos referir à variância que pode ser atribuída, com confiança, a uma única medida ou variável. Entretanto, existe, também, a variância que é específica de uma medida, mas não tão confiável; essa variância é chamada de *erro* ou *variância aleatória*. A proporção da variância comum presente numa variável é conhecida como *comunalidade*. Assim, a variável que não tem uma variância específica (ou variância aleatória) terá uma comunalidade de 1; a variável que não compartilha sua variância com qualquer outra variável terá uma comunalidade de 0.

Na análise dos fatores, estamos interessados em encontrar dimensões subjacentes

comuns dentro dos dados e, assim, estamos principalmente interessados na variância comum. Portanto, quando executamos uma análise dos fatores, é fundamental que saibamos o quanto da variância presente em nossos dados é a variância comum. Isso representa um impasse lógico: para fazer a análise dos fatores precisamos saber a proporção da variância comum presente nos dados, mas a única maneira de descobrir a extensão da variância comum é executar uma análise dos fatores! Existem duas maneiras de abordar esse problema. A primeira é assumir que toda a variância é variância comum. Assim, assumimos que a comunalidade de cada variável é 1. Fazendo essa suposição, simplesmente transpomos nossos dados originais para os componentes lineares constituintes (conhecidos como análise de componentes principais). A segunda abordagem é estimar o montante de variável comum estimando os valores da comunalidade para cada variável. Existem vários métodos para estimar a comunalidade, mas o mais usado (incluindo a fatoração alfa) é usar a correlação múltipla ao quadrado (CMQ) de cada variável com todas as outras. Assim, para os dados de popularidade, imagine que você executou uma regressão múltipla usando uma medida (Egoísmo) como a saída e as outras cinco medidas como previsores: o R^2 múltiplo resultante (veja a Seção 5.5.2) será usado como uma estimativa da comunalidade para a variável Egoísmo. Essa segunda abordagem é feita na análise de fatores. Essas estimativas permitem que a análise dos fatores seja executada. Uma vez que os fatores subjacentes tenham sido extraídos, novas comunalidades que representam a correlação múltipla entre cada variável e os fatores extraídos podem ser calculadas. Portanto, a comunalidade é uma medida de proporção da variância explicada pelos fatores extraídos.

15.3.3 Análise de fatores *versus* Análise de Componentes Principais ②

Acabei de explicar que existem duas abordagens para localizar dimensões subjacentes de um conjunto de dados: a análise dos fatores e a análise dos componentes principais. Essas

² Vale a pena notar que a análise de componentes principais não é, de fato, o mesmo que a análise dos fatores. Isso não impede que idiotas como eu as discuta como se fossem – voltaremos a esse assunto mais tarde.

técnicas diferem nas estimativas da comunalidade utilizada. De forma simplista, a análise dos fatores deriva de um modelo matemático em que os fatores são estimados, enquanto a análise dos componentes principais apenas decompõe os dados originais em um conjunto de variáveis lineares (veja Dunteman, 1989, Capítulo 8, para mais detalhes sobre as diferenças entre os dois procedimentos). Assim, somente a análise dos fatores pode estimar os fatores subjacentes e ela tem como base várias hipóteses para que essas estimativas sejam precisas. A análise dos componentes principais se preocupa somente em determinar que componentes lineares existem dentro dos dados e como uma variável particular pode contribuir com aquele componente. Em termos de teoria, este capítulo é dedicado mais à análise dos componentes principais do que propriamente à análise dos fatores. A razão é que a análise dos componentes principais é um procedimento totalmente psicométrico, ela é conceitualmente menos complexa do que a análise dos fatores e apresenta muitas similaridades com a análise discriminante (descrita no capítulo anterior).

Entretanto, devemos considerar se as técnicas fornecem soluções diferentes para o mesmo problema. Baseados em uma extensiva revisão da literatura, Guadagnoli e Velicer (1988) concluíram que a solução gerada pela análise dos componentes principais difere um pouco daquela derivada da técnica da análise dos fatores. Na realidade, existem algumas circunstâncias nas quais essa afirmação não é verdadeira. Stevens (1992) resume os exemplos e conclui que com 30 ou mais variáveis e comunalidades maiores do que 0,7 para todas as variáveis, diferentes soluções são improváveis; entretanto, com menos do que 20 variáveis e com baixas comunalidades ($< 0,4$) podem ocorrer diferenças.

O lado menos interessante desse argumento é eloquentemente descrito por Cliff (1987), que observou que defensores da análise de fatores “insistem que a análise de componentes é, no melhor caso, uma análise de fatores comuns com algum erro acrescido e, no pior

caso, uma bagunça irreconhecível de coisas das quais nada pode ser determinado” (p. 349). Os ânimos se exaltam nesse assunto, alguns argumentando que quando a análise dos componentes principais for utilizada ela não deveria ser descrita como uma análise dos fatores e que não deve atribuir um significado substantivo aos componentes resultantes. Entretanto, para os não estatísticos, a diferença entre um componente principal e um fator pode ser difícil de conceitualizar (ambos são modelos lineares) e as diferenças surgem principalmente nos cálculos.³

15.3.4 A teoria por trás da Análise de Componentes Principais ③



A análise dos componentes principais trabalha de uma maneira muito semelhante à MANOVA e à análise discriminante (veja o capítulo anterior). Embora não seja necessário entender os princípios matemáticos em detalhes, leitores do capítulo anterior podem ter se beneficiado das comparações entre as duas técnicas. Para aqueles que não leram o capítulo anterior, sugiro uma rápida consulta antes de continuar!

Na MANOVA, várias matrizes SSCP calculadas continham informações sobre os relacionamentos entre variáveis dependentes. Mencionei anteriormente que essas matrizes SSCP poderiam ser facilmente convertidas em matrizes variância-covariância, que apresentam a mesma informação, mas em forma de média (isto é, levando em conta o número de observações). Comentei, também, que dividindo cada elemento pelo desvio padrão pertinente, as matrizes variância-covariância se tornam padronizadas. O resultado é uma matriz de correlação. Na análise dos fatores, geralmente

³ Por esse motivo, usei os termos *componentes* e *fatores* da mesma forma ao longo do capítulo. Embora esse uso dos termos injurie alguns estatísticos (e psicólogos), aposto que essas pessoas não irão ler este livro! Reconheço as diferenças metodológicas, mas é mais fácil para os estudantes se eu enfatizar as semelhanças entre as técnicas e não as diferenças.

tratamos com matrizes de correlação (embora seja possível analisar a matriz de variância-covariância também) e que essa matriz apresenta a mesma informação que a matriz SSCP na MANOVA. A diferença é apenas que a matriz de correlação é uma versão média da SSCP que foi padronizada.

Na MANOVA, usamos várias matrizes SSCP que representam diferentes componentes da variação experimental (a variação do modelo e a variação residual). Na análise de componentes principais, a matriz de covariâncias (ou correlações) não pode ser desmembrada dessa maneira (porque todos os dados vêm do mesmo grupo de participantes). Na MANOVA, acabamos olhando para as combinações lineares (*variates*) ou componentes da matriz SSCP que representam a razão entre a variância do modelo e a variância do erro. Essas combinações lineares (*variates*) são dimensões lineares que separam os grupos testados, e vimos que as variáveis dependentes foram representadas por esses componentes subjacentes. Resumindo, verificamos se os grupos poderiam ser separados por uma combinação linear das variáveis dependentes. Essas combinações lineares foram encontradas calculando os autovetores da SSCP. O número de combinações lineares obtido foi menor do que p (o número de variáveis dependentes) ou $k - 1$ (onde k é o número dos grupos). Na análise de componentes, fazemos algo semelhante (estou simplificando um pouco as coisas, mas isso lhe dará uma ideia básica). Isto é, pegamos uma matriz de correlação e calculamos as combinações lineares. Não há grupos de observação, assim, o número de combinações lineares obtidas será sempre igual ao número de variáveis medidas (p). As combinações lineares são descritas, como na MANOVA pelos autovetores associados à matriz de correlação. Os elementos dos autovetores são os pesos de cada variável na combinação linear (veja a Equação 14.5). Esses valores são as cargas dos fatores descritas anteriormente. O maior autovalor associado com cada um dos autovetores fornece um indicador único da importância de cada combinação linear (ou componente). A

ideia básica é retermos fatores com autovalores relativamente grandes e ignoramos aqueles com autovalores relativamente pequenos.



Resumindo, a análise de componentes trabalha de uma maneira similar à MANOVA. Começamos com uma matriz representando os relaciona-

mentos entre variáveis. Os componentes lineares (também chamados de combinações lineares (*variates*) ou fatores) daquela matriz são, então, calculados pela determinação dos autovalores da matriz. Esses autovalores são usados para calcular os autovetores, os elementos que fornecem as cargas de uma variável ou um fator em particular (isto é, eles são os valores b na equação (15.1)). O autovalor é, também, uma medida da importância do autovetor com o qual ele está associado.

15.3.5 Extração de fatores: autovalores e o diagrama de inclinação ②

Nem todos os fatores são extraídos numa análise e existe um debate sobre os critérios usados para decidir se um fator é estatisticamente importante. Mencionei acima que os autovalores associa-

dos a uma combinação linear (*variate*) indicam a importância daquele fator. Portanto, parece lógico que devem ser retidos somente fatores com autovalores grandes. Como decidimos se um autovalor é suficientemente grande ou não para representar um fator significativo? Bem, uma técnica advogada por Cattell (1966b) é traçar um gráfico de cada autovalor (eixo-y) contra o fator com o qual ele está associado (eixo-x). Esse gráfico é conhecido como um *diagrama de declividade*. Mencionei anteriormente que é possível obter tantos fatores quanto as variáveis existentes, e que cada uma tem um autovalor associado. Representando num gráfico os autovalores, a importância relativa de cada fator se torna aparente. Tipicamente, existirão alguns fatores com autovalores bem



altos e muitos fatores com autovalores relativamente baixos, assim, esse gráfico tem uma forma característica: uma inclinação bem acentuada na curva seguida de uma cauda quase horizontal (veja a Figura 15.2). Cattell (1966b) argumentou que o ponto de corte para selecionar fatores deveria ser no ponto de inflexão dessa curva. Com uma amostra de mais de 200 participantes, o diagrama de declividade fornece um critério bastante confiável para a seleção dos fatores (Stevens, 1992).

Embora diagramas de declividade sejam úteis, a seleção de fatores não deve ser baseada apenas nesse critério. Kaiser (1960) recomendou reter todos os fatores com autovalores maiores do que 1. Esse critério é baseado na ideia de que autovalores representam o montante de variação explicada por um fator e que um autovalor de 1 representa um substancial montante de variação. Jolliffe (1972-1986) relata que o critério de Kaiser é muito rígido e sugere a terceira opção para reter todos os fatores com autovalores maiores do que 0,70. A diferença entre quantos fatores serão retidos utilizando os métodos de Kaiser comparado ao de Jolliffe pode ser enorme.

Você pode querer saber como os três métodos se comparam. De modo geral, pesquisas têm demonstrado que o critério de Kaiser é preciso quando o número de variáveis é me-

nor do que 30 e as comunalidades resultantes (após a extração) são todas maiores do que 0,7. O critério de Kaiser é, também, preciso quando o tamanho da amostra excede a 250 e a média da comunalidade é maior ou igual a 0,6. Em qualquer outra circunstância, é melhor você usar o diagrama de declividade desde que o tamanho da amostra seja maior do que 200 (veja Stevens, 1992, p. 378-380 para mais detalhes). Por omissão, o SPSS usa o critério de Kaiser para extrair fatores. Portanto, se você usar o diagrama de declividade para determinar quantos fatores serão retidos, deverá executar novamente a análise especificando que o SPSS extraia o número de fatores que você deseja.

Entretanto, como é geralmente o caso em estatística, os três critérios geralmente fornecem diferentes soluções! Nessas situações, as comunalidades dos fatores devem ser consideradas. Na análise dos componentes principais, começamos com comunalidades de 1 com todos os fatores retidos (porque assumimos que toda a variância é comum). Nesse estágio, encontramos as combinações lineares que existem nos dados – assim, transformamos os dados sem descartar informação alguma. Entretanto, para descobrir qual é a variância comum que *realmente* existe entre as variáveis, devemos decidir quais fatores são significativos e descartar os ir-

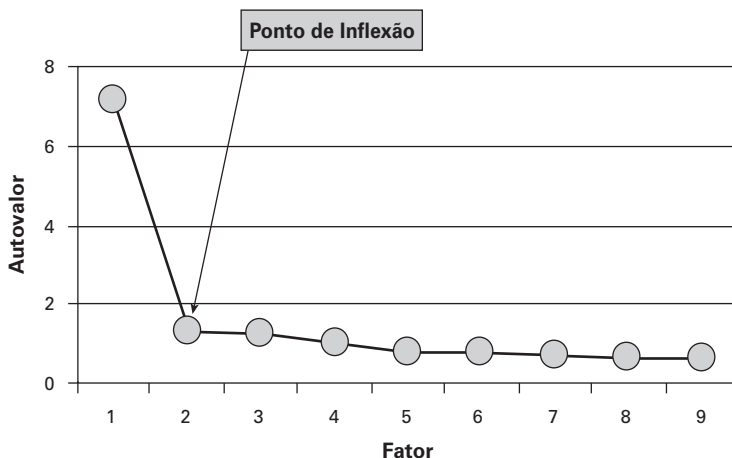


Figura 15.2 Exemplo de um diagrama de declividade para dados com apenas um fator subjacente.

relevantes. Portanto, descartamos alguma informação. Os fatores que retemos não irão explicar toda a variância dos dados (porque descartamos alguma informação), assim, as comunicações após a extração serão sempre menores do que 1. Os fatores retidos não correspondem perfeitamente às variáveis originais – eles apenas refletem a variância comum presente nos dados. Se as comunicações representam a perda de informação, elas são estatísticas importantes. Quanto mais próximas as comunicações estão de 1, melhor os fatores explicarão os dados originais. É lógico que quanto mais fatores retidos, maiores serão as comunicações (porque menos informação é descartada); portanto, as comunicações são bons índices para verificar se poucos fatores foram retidos. Na verdade, com a *análise dos fatores por mínimos quadrados generalizados* e a *análise dos fatores por máxima verossimilhança*, você consegue uma medida estatística da aderência da solução dos fatores (veja o próximo capítulo para mais testes de aderência). Isso basicamente mede a proporção da variância que a solução por fatores explica (assim, pode ser pensado como uma comparação de comunicações antes e após a extração).

Um conselho final: a sua decisão de quantos fatores extrair também dependerá de por que você está fazendo a análise em primeiro lugar (por exemplo, se você está tentando superar problemas de multicolinearidade na regressão, será melhor extrair mais fatores do que menos).

15.3.6 Melhorando a interpretação: a rotação de fatores ③

Depois de extrair os fatores, é possível calcular em que grau as variáveis se adaptam a esses fatores (isto é, calcular a carga da variável em cada fator). **Em geral, você descobrirá que a maioria das variáveis tem cargas altas no fator mais importante e cargas baixas nos demais fatores. Essa característica torna a interpretação difícil, assim, uma técnica chamada de rotação de fatores é usada para fazer a distinção entre fatores. Se um fator for um eixo classificatório ao longo do qual as variáveis podem ser traçadas, a rotação dos fatores gira esses eixos de modo que as variáveis estejam carregadas ao máximo somente em um fator. A Figura 15.3 demonstra como esse processo funciona usando um exemplo**

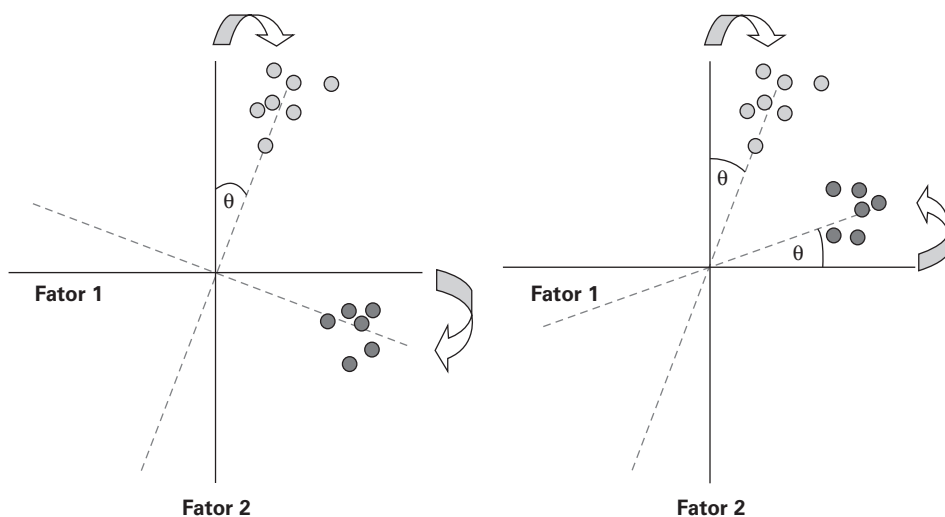


Figura 15.3 Representação esquemática da rotação de fatores. O gráfico da esquerda mostra a rotação ortogonal, enquanto o gráfico da direita mostra a rotação oblíqua (veja o texto para mais detalhes). θ é o ângulo pelo qual os eixos giram.

no qual há somente dois fatores. Imagine que uma socióloga está interessada em classificar professores universitários como um grupo demográfico. Ela descobriu que apenas duas dimensões subjacentes descrevem melhor esse grupo: alcoolismo e realização (vá a qualquer conferência e você verá que acadêmicos bebem muito!). O primeiro fator, alcoolismo, tem um aglomerado de variáveis associadas a ele (pontos pretos) e essas podem ser medidas, como o número de unidades bebidas em uma semana, dependência e personalidade obsessiva. O segundo fator, realização, também tem um aglomerado de variáveis associadas a ele (pontos claros) e essas podem ser medidas relacionadas a salário, *status* no emprego e número de pesquisas publicadas. Inicialmente, as linhas cheias representam os fatores, e olhando para as coordenadas deveria ficar claro que os pontos claros têm cargas altas para o fator 2 (eles estão bem alto no eixo) e cargas médias no fator 1 (elas não estão muito longe desse eixo). Contrariamente, os pontos escuros têm cargas altas para o fator 1 e cargas médias para o fator 2. Girando os eixos (linhas pontilhadas), asseguramos que ambos os aglomerados de variáveis tenham uma interseção pelo fator com o qual mais se relacionam. Assim, após a rotação, as cargas das variáveis são maximizadas sobre um fator (o fator que intersecta o aglomerado) e minimizadas no(s) fator(es) remanescente(s). Se um eixo passar através de um aglomerado de variáveis, essas variáveis terão uma carga de aproximadamente 0 no eixo oposto. Se essa ideia é confusa, veja a Figura 15.3 e pense sobre os valores das coordenadas antes e após a rotação (isso é melhor visualizado girando o livro quando você olha para os eixos que foram girados).

Existem dois tipos de rotação que podem ser feitos. O primeiro é a *rotação ortogonal*, e o lado esquerdo da Figura 15.3 representa esse método. No Capítulo 8, vimos que o termo “ortogonal” significa não-relacionado e, nesse contexto, ele significa que rotamos os fatores enquanto os mantemos independentes. Antes da rotação todos os fatores são independentes (isto é, eles não se relacionam de ma-

neira alguma) e a rotação ortogonal assegura que os fatores permaneçam não-relacionados. Isso ocorre porque na Figura 15.3 os eixos são girados enquanto permanecem perpendiculares⁴. A outra forma de rotação é a *rotação oblíqua*. A diferença da rotação oblíqua é que os fatores podem se relacionar (por essa razão, os eixos do lado direito do diagrama na Figura 15.3 não permanecem perpendiculares).

A escolha da rotação depende se existe uma boa razão teórica para supor que os fatores estejam relacionados ou que sejam independentes (veja meus próximos comentários sobre esse assunto) e, também, como as variáveis se aglomeram nos fatores antes da rotação. No primeiro ponto, não podemos esperar que alcoolismo seja completamente independente de realização (afinal, grandes realizações levam a um alto nível de estresse, o qual leva a bebida!). Portanto, teoricamente devemos escolher a rotação oblíqua. No segundo ponto, a Figura 15.3 demonstra como o posicionamento dos aglomerados é importante para determinar o sucesso da rotação (note a posição dos pontos escuros). Especificamente, se uma rotação ortogonal fosse executada no diagrama do lado direito, ela teria consideravelmente menos sucesso na maximização das cargas do que a rotação oblíqua que é mostrada. Uma abordagem é executar a análise usando ambos os tipos de rotação. Pedhazur e Schmelkin (1991) sugerem que se a rotação oblíqua demonstra uma correlação desprezível entre os fatores extraídos, é razoável usar a rotação ortogonal. Se a rotação oblíqua revela uma estrutura de fatores correlacionada, a solução ortogonal deve ser descartada. De qualquer modo, uma rotação oblíqua somente deve ser usada se existem boas razões para supor que os fatores subjacentes *poderiam* estar relacionados em termos teóricos.

A matemática por trás da rotação dos fatores é complexa (especificamente a rotação oblíqua). Entretanto, na rotação oblíqua, porque cada fator pode ser rodado por diferentes

⁴ Esse termo significa que os eixos formam um ângulo reto entre si.

valores uma *matriz de rotação dos fatores*, Λ , ela é necessária. A matriz de rotação dos fatores é quadrada e seu tamanho depende de quantos fatores foram extraídos dos dados. Se dois fatores foram extraídos, ela será uma matriz 2x2, mas se quatro fatores foram extraídos, ela se torna uma matriz 4x4. Os valores da matriz de rotação dos fatores consistem em senos e cossenos do ângulo de rotação (θ). Essa matriz é multiplicada pela matriz das cargas dos fatores não rotacionados A , para obter a matriz de cargas dos fatores rotacionados.

Para o caso de dois fatores, a matriz de rotação dos fatores seria:

$$\Lambda = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Portanto, você deve pensar nessa matriz como representando o ângulo no qual os eixos foram girados ou o grau no qual os fatores foram rotacionados. O ângulo da rotação necessário para otimizar a solução dos fatores é encontrado de forma iterativa, e métodos diferentes podem ser usados.

15.3.6.1 Escolhendo um método de rotação dos fatores ③

O SPSS tem três métodos de rotação ortogonal (*varimax*, *quartimax* e *equamax*) e dois métodos de rotação oblíqua (*direct oblimin* e *promax*). Esses métodos diferem na forma de rodar os fatores e, portanto, a saída resultante dependerá do método selecionado. A rotação *quartimax* tenta maximizar a dispersão da carga dos fatores de uma variável por todos os fatores. Portanto, se torna mais fácil interpretar as variáveis. Entretanto, isso geralmente resulta em muitas variáveis com altas cargas em um único fator. O *varimax* é o oposto disso, ele tenta maximizar a dispersão das cargas dentro dos fatores. Portanto, ele tenta agregar um menor número de variáveis sobre cada fator resultando em mais aglomerados de fatores interpretáveis. O *equamax* é um híbrido das outras duas abordagens e existem relatos de que ele apresenta um comportamento bastante errático (veja Tabachnick e Fidell, 2001). Para

uma primeira análise, você deve selecionar o *varimax*, pois ele é uma boa abordagem geral que simplifica a interpretação dos fatores.

A rotação oblíqua é mais complexa porque nela a correlação entre fatores é permitida. No caso do *oblimin* direto, o grau no qual é permitida a correlação dos fatores é determinado pelo valor de uma constante chamada delta. O valor por omissão no SPSS é zero e isso assegura que uma alta correlação entre os fatores não seja permitida (isso é conhecido como a rotação *quartimin* direta). Se você optar por um delta maior do que zero (até um máximo de 0,80), espere fatores altamente correlacionados; se você estabelecer um delta menor do que zero (até um mínimo de -0,8), espere fatores menos correlacionados. O valor padrão zero é sensível para a maioria das análises e eu não recomendo trocá-lo a não ser que você saiba o que está fazendo (veja Pedhazur e Schmelkin, 1991, p. 620). O *promax* é um procedimento mais rápido projetado para grandes conjuntos de dados.

Teoricamente, a escolha exata de rotação dependerá muito de se você acha ou não que os fatores subjacentes estão relacionados. Se você espera que os fatores sejam independentes, deve escolher uma das rotações ortogonais (recomendo a *varimax*). Se, entretanto, existem fundamentações teóricas para sustentar que seus fatores possam estar correlacionados, a *oblimin* direta deve ser selecionada. Na prática, existem fortes indícios de que as rotações ortogonais não fazem sentido para dados naturalistas e para qualquer dado envolvendo humanos (você pode pensar em algum construto psicológico que não esteja de maneira alguma correlacionado com qualquer outro construto?). Desse modo, alguns argumentam que as rotações ortogonais nunca devem ser usadas.

15.3.6.2 Importância da carga dos fatores ②

Uma vez que a estrutura dos fatores foi encontrada, é importante decidir quais variáveis compõem cada um dos fatores. Anteriormente, afirmei que as cargas dos fatores eram uma medida de substancial importância de uma determinada variável em relação a um dado fator.

Portanto, faz sentido que usemos esses valores para atribuir variáveis aos fatores. É possível acessar a significância estatística da carga de um fator (afinal, ela é simplesmente um coeficiente de correlação ou de regressão); entretanto, essa opção não é tão fácil quanto parece (veja Stevens, 1992, p. 382). **Tipicamente, os pesquisadores consideram uma carga de valor absoluto maior do que 0,3 como sendo importante.** Entretanto, a significância de uma carga do fator dependerá do tamanho da amostra. Stevens (1992) produziu uma tabela de valores críticos com os quais as cargas podem ser comparadas. Resumindo, ele recomenda que para um tamanho da amostra de 50, uma carga de 0,722 pode ser considerada significativa, para 100, a carga deve ser maior do que 0,512, para 200, ela deve ser maior do que 0,364, para 300, ela deve ser maior do que 0,298, para 600, ela deve ser maior do que 0,21, e para 1000, ela deve ser maior do que 0,162. Esses valores são baseados num nível alfa de 0,01 (bilateral) que permite o fato de que várias cargas precisarão ser testadas (ver Stevens, 1992, p. 382-84 para maiores detalhes). Portanto, em amostras muito grandes, cargas pequenas podem ser consideradas estatisticamente significativas. O SPSS não fornece testes de significância para a carga dos fatores, mas aplicando as diretrizes de Stevens você deve ter algum esclarecimento sobre a estrutura das variáveis e fatores.

A significância de uma carga não mostra a importância de uma variável para um fator. Esse valor pode ser encontrado elevando-se ao quadrado a carga do fator para uma estimativa do montante da variância num fator de responsabilidade daquela variável (como R^2). A esse respeito, Stevens (1992) recomenda interpretar somente cargas dos fatores com um valor absoluto maior do que 0,4 (o que explicaria por volta de 16% da variância da variável).

15.4 EXEMPLO DE PESQUISA ②

A análise dos fatores é usada para desenvolver questionários: afinal, se você quer medir uma habilidade ou característica, deve assegurar que as perguntas feitas estejam re-

lacionadas ao construto que deseja medir.⁵ Notei que muitos estudantes se interessam pelo SPSS. Portanto, projetei um questionário para medir uma característica que denominei “ansiedade com o SPSS”. Planejei um questionário para medir os vários aspectos da ansiedade dos alunos com o SPSS. Elaborei perguntas baseadas em entrevistas com estudantes com ansiedade e sem ansiedade e terminei com 23 perguntas possíveis. Cada pergunta era uma afirmação avaliada por uma escala de Likert de cinco pontos iniciando com “discordo totalmente” passando por “não concordo nem discordo” e finalizando em “concordo plenamente”. O questionário é apresentado na Figura 15.4.

O questionário foi projetado para prever quão ansioso um determinado indivíduo estaria em aprender a usar o SPSS. Também queria saber se a ansiedade em relação ao SPSS poderia ser desmembrada em formas específicas de ansiedade. Assim, em outras palavras, que outras variáveis latentes contribuem para a ansiedade com o SPSS? Com a ajuda de alguns amigos professores, coletei 2571 questionários completos (obviamente, esse exemplo é fictício!). Os dados estão armazenados no arquivo **SAQ.sav**. Carregue o conjunto de dados no editor de dados do SPSS e dê uma olhada nas variáveis e suas propriedades. Note que cada pergunta (variável) é representada por uma coluna diferente. Sabemos que os casos no SPSS (ou dados de pessoas) são armazenados em linhas e as variáveis são armazenadas em colunas, assim, essa disposição é consistente com os capítulos anteriores. Observe também que existem 23 variáveis rotuladas de **q₁** a **q₂₃**, e que cada uma tem um rótulo indicando a pergunta. Ao rotular minhas variáveis, posso ser bem claro sobre o que cada variável representa (essa é a vantagem de dar títulos às suas variáveis e não apenas utilizar os cabeçalhos restritivos das colunas).

⁵ Incluí um arquivo no material disponível no site chamado de **DesigningQuestionnaires.pdf** (projetando questionários), que fornece alguns conselhos sobre o processo de desenvolvimento de questionários.

DT = Discordo Totalmente, D = Discordo, N = Nem discordo, nem concordo, C = Concordo, CT = Concordo Totalmente

	DT	D	N	C	CT
01. A estatística me faz chorar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
02. Meus amigos irão pensar que sou estúpido por não saber mexer com o SPSS.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
03. Desvios padrão me excitam.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
04. Eu sonho com Pearson me atacando com coeficientes de correlação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
05. Eu não entendo estatística.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
06. Eu tenho pouca experiência com computadores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
07. Todos os computadores me odeiam.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
08. Eu nunca fui bom em matemática.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
09. Meus amigos são melhores em estatística do que eu.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Os computadores são úteis somente para jogos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Eu fui muito mal em matemática na escola.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Dizem que o SPSS torna a estatística mais fácil de entender, mas isso não é verdade.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Estou com medo de provocar danos irreparáveis por causa da minha incompetência com computadores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. Computadores têm suas próprias mentes e deliberadamente erram quando eu os uso.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Computadores existem para me sacanear.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Eu choro escancaradamente quando ouço falar em tendência central.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. Eu entro em coma toda vez que vejo uma equação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. O SPSS sempre trava quando eu tento usá-lo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. Todos olham para mim quando uso o SPSS.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20. Eu não consigo dormir quando penso em autovetores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. Eu tenho pesadelos com a distribuição normal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. Meus amigos são melhores com o SPSS do que eu.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. Eu sou bom em estatística e as pessoas acham que sou CDF.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 15.4 O questionário sobre ansiedade do SPSS (QAS).

15.4.1 Primeiras considerações ②

15.4.1.1 Tamanho da amostra ②

Os coeficientes de correlação flutuam de amostra para amostra, mais em pequenas amostras do que em grandes. Portanto, a confiabilidade da análise dos fatores depende também do tamanho da amostra. Muito já foi escrito sobre o tamanho da amostra necessário para a análise de fatores, resultando em diversas “regras práticas”. A regra comum é sugerir que o pesquisador tenha pelo menos entre 10 e 15 participantes por variável. No entanto, a base empírica dessa regra não está clara (Nunally, 1978, recomenda ter 10 vezes

mais participantes do que variáveis). Kass e Tinsley (1979) recomendam ter entre 5 e 10 participantes por variável até um total de 300 (além do que os testes paramétricos tendem a se estabilizar não importando a proporção entre participantes e variáveis). Na verdade, Tabachnick e Fidell (2001) concordam que “é confortável ter pelo menos 300 casos para cada análise dos fatores” (p. 640) e Comrey e Lee (1992) classificam 300 como um bom tamanho da amostra, 100 como pobre e 1000 como excelente.

Felizmente, nos últimos anos muitas pesquisas foram feitas na forma de experimentos usando dados simulados (chamados de

estudos de Monte Carlo). Arrindell e van der Ende (1985) usaram dados reais para investigar o efeito de diferentes participantes a fim de determinar a razão entre duas variáveis. Eles concluíram que mudanças dessa razão fizeram pouca diferença para a estabilidade das soluções dos fatores. Mais recentemente, Gudagnoli e Velicer (1998) descobriram que o mais importante para determinar soluções de fatores confiáveis era o tamanho da amostra e a magnitude absoluta das cargas dos fatores. Resumindo, eles argumentam que se um fator tem quatro ou mais cargas maiores do que 0,6 ele é confiável a despeito do tamanho da amostra. Além disso, fatores com 10 ou mais cargas maiores do que 0,40 são confiáveis se o tamanho da amostra é maior do que 150. Finalmente, fatores com poucas cargas baixas não devem ser interpretados a não ser que o tamanho da amostra seja de no mínimo 300. Mais recentemente, MacCallum, Widaman, Zhang e Hong (1999) mostraram que o tamanho da amostra mínimo ou a razão entre o tamanho da amostra e o número de variáveis depende entre outros aspectos do projeto do estudo. Seu estudo indicou que à medida que as comunalidades se tornam mais baixas, a importância do tamanho da amostra aumenta. Com todas as comunalidades acima de 0,6, tamanhos de amostras relativamente pequenos (menores do que 100) podem ser perfeitamente adequados. Com comunalidades na faixa de 0,5, amostras entre 100 e 200 podem ser boas o suficiente desde que cada uma tenha relativamente poucos fatores com somente um pequeno número de variáveis indicadoras. No pior cenário, de baixas comunalidades (bem abaixo de 0,5) e uma grande quantidade de fatores subjacentes, eles recomendam amostras acima de 500. Fica claro a partir deste trabalho que uma amostra de 300 ou mais provavelmente fornecerá uma solução de fatores estável, mas um pesquisador esperto medirá variáveis o suficiente para avaliar adequadamente todos os fatores que teoricamente ele esperaria encontrar.

Outra alternativa é usar a medida de adequação da amostra de Kaiser-Meyer-Olkin (KMO) (veja Kaiser, 1970). A KMO pode ser

calculada para variáveis individuais e múltiplas e representa a razão da correlação ao quadrado para a correlação parcial ao quadrado entre as variáveis. A estatística KMO varia entre 0 e 1. Um valor de 0 indica que a soma de correlações parciais é grande relativa à soma das correlações, indicando difusão no padrão das correlações (portanto, a análise de fatores provavelmente é inadequada). Um valor próximo de 1 indica que padrões de correlações são relativamente compactos, assim, a análise de fatores deveria dar preferência a valores distintos e confiáveis. Kaiser (1974) recomenda valores maiores do que 0,5 como sendo apenas aceitáveis (valores abaixo disso deveriam levar você a coletar mais dados ou repensar quais variáveis incluir). Além disso, valores entre 0,5 e 0,7 são medíocres, valores entre 0,7 e 0,8 são bons, valores entre 0,8 e 0,9 são ótimos e valores acima de 0,9 são excelentes (veja Hutcheson e Sofroniou, 1999, p. 224-225 para mais detalhes).

15.4.1.2 Vasculhando os dados ②

Neste livro, usei a expressão “se entrar lixo, sai lixo”. Essa expressão se aplica particularmente à análise dos fatores porque o SPSS irá sempre encontrar uma solução para um conjunto de variáveis. Entretanto, é improvável que a solução tenha qualquer significado real se as variáveis analisadas não são sensíveis. Quando conduzir uma análise de fatores, primeiro analise a intercorrelação entre as variáveis. Se as perguntas do seu teste medem a mesma dimensão subjacente (ou dimensões), esperamos que elas se correlacionem (porque elas estão medindo a mesma coisa). Mesmo se as perguntas medirem diferentes aspectos das mesmas coisas (por exemplo, poderíamos medir a ansiedade de modo geral em termos de subcomponentes, como preocupação, pensamentos intrusivos e excitação psicológica) deveria, ainda, haver altas intercorrelações entre as variáveis relacionadas a essas subcaracterísticas. Se você encontrar quaisquer variáveis que não se correlacionam com outras variáveis (ou com pouquíssimas), deve considerar a exclusão dessas variáveis antes da análise de fatores ser executada. Um extremo desse


problema é quando a matriz das correlações se assemelha a uma matriz identidade (veja a Seção 14.3.2). Nesse caso, as variáveis se correlacionam somente com elas mesmas e todos os outros coeficientes de correlação estão próximos de zero. O SPSS testa isso usando o teste de esfericidade de Bartlett (veja a próxima seção). As correlações entre variáveis podem ser checadas usando o procedimento *correlate* (correlação) (veja o Capítulo 4) para criar uma matriz de correlações de todas as variáveis. A matriz também pode ser criada como parte da análise dos fatores principal.

O problema oposto é quando as variáveis são altamente correlacionadas. Embora a multicolinearidade moderada não seja um problema para a análise de fatores, é importante evitar a multicolinearidade extrema (isto é, variáveis altamente correlacionadas) e *singularidade* (variáveis que estão perfeitamente correlacionadas). Assim como na regressão, a singularidade causa problemas na análise de fatores porque se torna impossível determinar a contribuição única para um fator se as variáveis estão altamente correlacionadas (como foi o caso para regressão múltipla). Portanto, nesse estágio inicial, queremos eliminar qualquer variável que não se correlacione com qualquer outra variável ou que seja altamente correlacionada com outras variáveis ($R > 0,9$). A multicolinearidade pode ser detectada ve-

rificando o determinante da matriz-R (veja a próxima seção).

Além de procurar por interrelações, você deve se assegurar que as variáveis tenham distribuições aproximadamente normais e que sejam medidas pelo menos em um nível de intervalo (supõe-se, erroneamente, que as escalas de Likert o sejam!). A hipótese de normalidade é mais importante se você deseja generalizar os resultados da sua análise além dos dados coletados.

15.5 EXECUTANDO A ANÁLISE ②

Acesse a caixa de diálogo principal (Figura 15.5) utilizando o menu **Analyze⇒Data Reduction⇒Factor...** (Analisar⇒Redução de Dados⇒Fator...), selecione as variáveis que você quer incluir na análise (lembre-se de excluir qualquer variável identificada como problemática durante a seleção dos dados) e transfira-as para a caixa **Variables** (Variáveis) clicando em .

Existem várias opções disponíveis, a primeira delas pode ser selecionada clicando em **Descriptives...** (Descritivas) para acessar a caixa de diálogo como a da Figura 15.6. A opção **Univariate descriptives** (Descritivas Univariadas) fornece médias e desvios padrão para cada variável. A maioria das outras opções se referem à matriz de correlação das variáveis

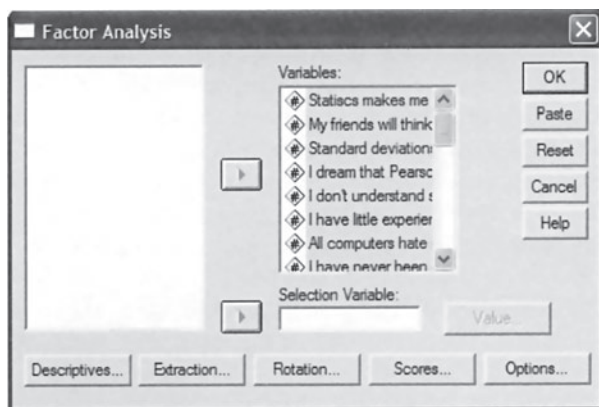


Figura 15.5 Caixa de diálogo principal para a análise dos fatores.

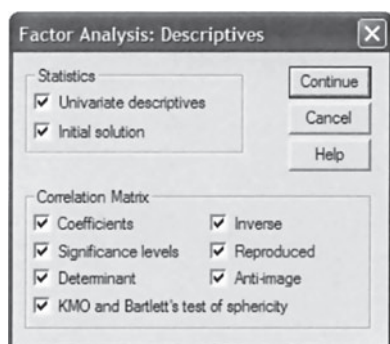


Figura 15.6 Descritivas em análise dos fatores.

(a matriz-R descrita anteriormente). A opção **Coefficients** (Coeficientes) produz a matriz-R e a opção **Significance levels** (Níveis de significância) irá produzir uma matriz indicando o valor da significância de cada correlação da matriz-R. Você pode, também, solicitar a **Determinant** (Determinante) dessa matriz e essa opção é vital para testar a multicolinearidade ou singularidade. O determinante da matriz-R deve ser maior do que 0,00001; se for menor do que esse valor, verifique na matriz de correlações as variáveis que se correlacionem altamente ($R > 0,8$) e considere eliminar uma dessas variáveis (ou mais, dependendo da extensão do problema) antes de continuar. Como foi mencionado na Seção 6.8, a escolha de quais das duas variáveis eliminar será arbitrária e encontrar multicolinearidade nos dados deve levantar questões sobre a escolha dos itens do seu questionário.

O **KMO and Bartlett's test of sphericity** produz o KMO e o teste de Bartlett. Com uma amostra de 2571 não devemos nos preocupar com o tamanho da amostra (veja a Seção 15.4.1.1). Já abordamos o KMO (veja a Seção 15.4.1.1) e vimos os vários critérios de adequação da amostra. Também vimos o teste de Bartlett no Capítulo 14: ele examina se a matriz de correlações da população se parece com uma matriz de identidade (isto é, se ela testa se os componentes fora da diagonal são zero). Se a matriz de correlação da população se parece com a matriz de identidade, isso significa que

cada variável se correlaciona pessimamente com todas as outras variáveis (isto é, todos os coeficientes de correlação estão próximos de zero). Se tivéssemos uma matriz identidade, isso significaria que todas as variáveis são perfeitamente independentes (todos os coeficientes de correlação seriam zero). Dado isso, procuramos por aglomerações de variáveis que meçam coisas similares, e deveria ser óbvio o porquê desse cenário ser problemático: se nenhuma das variáveis se correlaciona, não existem aglomerados para encontrar.

A opção **Reproduced** (Reproduzida) produz uma matriz de correlações baseada no modelo (em vez de dados reais). Diferenças entre a matriz baseada no modelo e a matriz baseada nos dados observados indicam os resíduos do modelo (isto é, diferenças). O SPSS produz esses resíduos na tabela inferior da matriz reproduzida, e queremos que relativamente poucos desses valores sejam maiores do 0,05. Felizmente, para nos salvar de vasculhar essa matriz, o SPSS produz um resumo de quantos resíduos estão acima de 0,05. A opção **Reproduced** (Reproduzida) deve estar selecionada para obter esse resumo. A opção **Anti-image** (Anti-imagem) produz uma matriz anti-imagem das covariâncias e correlações. Essas matrizes contêm medidas de adequação da amostra para cada variável ao longo da diagonal e os opostos da correlação/covariâncias parciais fora das diagonais. Os elementos diagonais, como o KMO, devem ser maiores do que 0,5, no mínimo, se a amostra for adequada para um determinado par de variáveis dado. Se qualquer par de variáveis apresenta um valor menor do que esse, considere a retirada de uma das variáveis da análise. Os elementos fora da diagonal devem ser todos bem pequenos (próximos a zero) num bom modelo. Quando você tiver terminado com essa caixa de diálogo, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

15.5.1 Extração de fatores no SPSS ②

Para acessar a janela de diálogo **Extraction** (Extração), clique em **Extraction...** (Figura 15.7) na caixa de diálogo principal. Existem várias

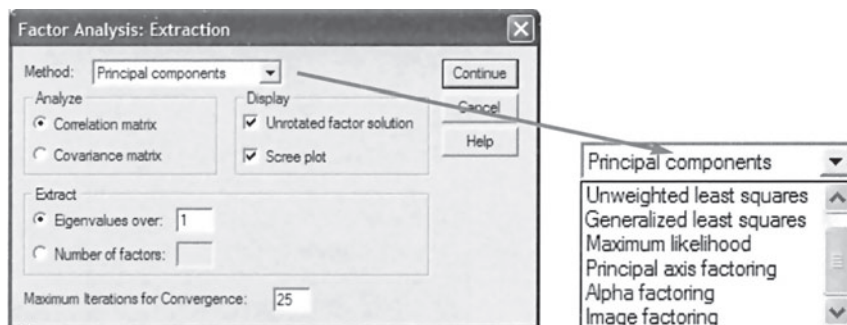


Figura 15.7 Caixa de diálogo para a extração de fatores.

maneiras de conduzir uma análise dos fatores (veja a Seção 15.3.1) e quando e onde você usa os vários métodos depende de vários fatores. Para nossos propósitos, iremos usar a análise dos *componentes principais*, que não é exatamente análise de fatores, mas os dois procedimentos, geralmente, conduzem a resultados semelhantes (veja a Seção 15.3.3).

Na caixa **Analyze** (Analisar), existem duas opções: analisar a matriz de correlações (**Correlation matrix**) ou analisar a matriz de covariância (**Covariance matrix**). Você deveria estar feliz com a ideia de que essas duas matrizes são realmente diferentes versões do mesmo: a matriz de correlação é uma versão padronizada da matriz de covariância. Analisar a matriz de correlação é um método padrão útil porque ele assume a forma padronizada da matriz; portanto, se as variáveis foram medidas usando diferentes escalas, isso não afetará a análise. Nesse exemplo, todas as variáveis foram medidas usando a mesma escala de medida (a escala de cinco pontos de Likert), mas, geralmente, você vai querer analisar variáveis que usam diferentes escalas de medidas. Analisar a matriz de correlação assegura que diferenças em escalas de medidas sejam consideradas. Além disso, mesmo variáveis medidas usando a mesma escala podem ter variâncias bem diferentes, e isso também cria problemas para a análise dos componentes principais. Usar a matriz de correlação elimina esse problema. Existem razões estatísticas para preferir ana-

lisar a matriz de covariâncias⁶, e geralmente os resultados irão diferir da análise na matriz de correlação. Entretanto, a matriz de covariância deve ser analisada somente quando as variáveis são comensuráveis.

A caixa **Display** (Apresentar) tem duas opções: solução dos fatores não-rotacionados (**Unrotated factor solution**) e o diagrama de declividade (**Scree plot**). O diagrama de declividade foi descrito anteriormente e é uma maneira útil de estabelecer quantos fatores devem ser retidos numa análise. A solução não-rotacionada dos fatores é útil para medir a melhoria na interpretação devido à rotação. Se a solução rotacionada é um pouco melhor do que a solução não-rotacionada, é possível que um método não apropriado (ou não ótimo) de rotação tenha sido utilizado.

A caixa de diálogo **Extract** (Extrair) fornece opções relativas à retenção dos fatores. Você tem a opção de selecionar tanto fatores com autovalores maiores do que o valor que você especificou quando reter um número fixo de fatores. Para a opção **Eigenvalues over** (Autovalores acima de), o padrão é a recomendação de Kaiser de autovalores acima de 1, mas você pode trocar isso para a recomendação de Jolliffe de autovalores acima de 0,7 ou qualquer outro valor que quiser. Execute uma

⁶ A razão é que o coeficiente de correlação é insensível à variação na dispersão dos dados, e a covariância não, produzindo, dessa maneira, uma estrutura de fatores mais bem definida (veja Tinsley e Tinsley, 1987).

análise preliminar com a opção **Eigenvalues over** (Autovalores acima de) 1 marcada, selecione também a opção do diagrama de declividade e compare os resultados. Se o exame do diagrama de declividade e dos autovalores acima de 1 levam você a reter o mesmo número de fatores, continue com a análise. Se os dois critérios dão resultados diferentes, examine as communalidades e decida você mesmo em qual dos dois critérios acreditar. Se decidir usar o diagrama de declividade, refaça a análise especificando o número de fatores a extrair. O número de fatores a serem extraídos é especificado selecionando **Number of factors** (Número de fatores) e digitando o número apropriado no espaço fornecido (por exemplo, 4).

15.5.2 Rotação ②

Já vimos que a interpretação dos fatores pode ser melhorada através da rotação. A rotação maximiza a carga de cada variável em um dos fatores extraídos enquanto minimiza a carga nos demais. Esse processo torna mais claro quais variáveis se relacionam com cada fator. A rotação é realizada através da mudança dos valores absolutos das variáveis enquanto mantêm constantes seus valores diferenciais. Clique em **Rotation...** (Rotação) para acessar a caixa de diálogo semelhante à da Figura 15.8. Discuti as várias opções de rotações na Seção 15.3.6.1, mas, para resumir, a escolha exata de rotação dependerá de se você acha ou não que

os fatores subjacentes devem se relacionar. **Se existem bases teóricas para achar que os fatores são independentes (não-relacionados), você deve escolher uma das rotações ortogonais (recomendo a *varimax*). Entretanto, se a teoria sugere que seus fatores podem se correlacionar, uma das rotações oblíquas (*oblimin* direta ou *promax*) deve ser selecionada. Nesse exemplo, selecionei a *varimax*.**

A caixa de diálogo também tem opções para mostrar uma solução rotacionada (**Rotated solution**) e um diagrama de carga (**Loading plot**). A solução rotacionada é mostrada por padrão e é essencial para interpretar a análise rotacionada final. O gráfico de carga fornecerá uma representação gráfica de cada variável traçada contra os fatores extraídos até um máximo de três fatores (infelizmente, o SPSS não pode produzir gráficos de quatro ou cinco dimensões!). Esse gráfico é semelhante ao da Figura 15.1 e usa a carga de fatores de cada variável para cada fator. Com dois fatores esses gráficos são de fácil interpretação, e você deve esperar ver um grupo de variáveis aglomeradas próximas do eixo *x* e um grupo diferente de variáveis aglomeradas em volta do eixo *y*. Se todas as variáveis estão aglomeradas entre os eixos, a rotação não teve sucesso na maximização da carga de uma variável sobre um único fator. Com três fatores esses gráficos podem se tornar bem confusos e forçar consideravelmente o sistema visual! Entretanto, eles podem ainda ser uma maneira útil para determinar as estruturas subjacentes dos dados.

Uma opção final é estabelecer a **Maximum Iterations for Convergence** (Iterações máximas para convergência), a qual especifica o número de vezes que o computador irá procurar por uma solução ótima. Em muitas circunstâncias, o padrão de 25 é mais do que adequado para o SPSS encontrar uma solução para um conjunto de dados. Entretanto, se você tiver um conjunto grande de dados (como temos aqui), o computador poderá ter dificuldade em encontrar uma solução (especialmente para a rotação oblíqua). Para permitir um grande conjunto de dados, mude o valor para 30.

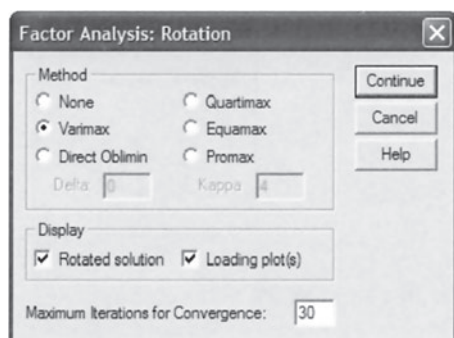


Figura 15.8 Análise dos fatores: caixa de diálogo da rotação.

15.5.3 Escores ②

A caixa de diálogo **factor scores** (escores dos fatores) (Figura 15.9) pode ser acessada clicando em **Scores...** (Escores) na caixa de diálogo principal. Essa opção permite que você salve os escores dos fatores (veja a Seção 15.2.3) para cada pessoa no editor de dados. O SPSS cria uma nova coluna para cada fator extraído e, então, coloca o escore do fator para cada pessoa dentro daquela coluna. Esses escores podem ser usados em análises posteriores ou simplesmente para identificar grupos de participantes que têm altos escores em fatores específicos. Existem três métodos para obter esses escores, todos eles foram descritos nas Seções 15.2.4 e 15.2.4.1. Se você quer assegurar que os escores dos fatores não são correlacionados, selecione o método de **Anderson-Rubin**; se as correlações entre os escores dos fatores são aceitáveis, escolha o método **Regression** (Regressão).

Como uma opção final, você pode solicitar ao SPSS que produza a matriz de coeficiente dos escores dos fatores. Essa é a matriz *B* descrita na Seção 15.2.4. Ela pode ser útil se, por qualquer motivo, você quiser construir equações dos fatores como as apresentadas em (15.5), porque ela fornece os valores de *b* para cada variável.

15.5.4 Opções ②

Esse conjunto de opções (Figura 15.10) pode ser obtido clicando em **Options...** (Opções) na caixa de diálogo principal. Dados perdidos são um problema para a análise dos fatores,

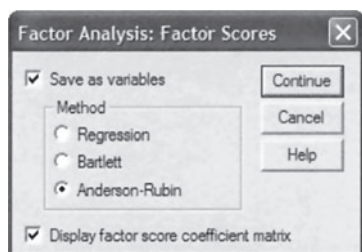


Figura 15.9 Análise dos fatores: caixa de diálogo dos escores dos fatores.

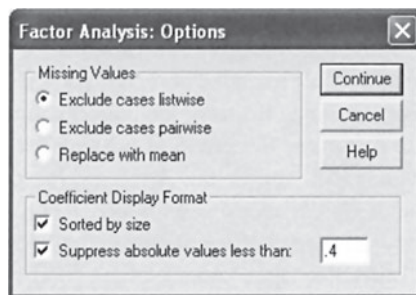


Figura 15.10 Análise dos fatores: caixa de diálogo opções.

assim como a maioria dos outros procedimentos, e o SPSS fornece uma escolha para excluir casos ou estimar um valor para um caso. Tabachnick e Fidell (2001) tem um capítulo excelente na análise dos dados (Capítulo 4 – veja também o não tão excelente Capítulo 3 desse livro). Baseado no conselho deles, você deve considerar a distribuição dos dados omissos. Se os dados omissos não são distribuídos normalmente ou o tamanho da amostra após a exclusão é muito pequena, a estimativa é necessária. O SPSS usa a média como uma estimativa **Replace with mean** (Substituir pela média). Esses procedimentos baixam o desvio padrão das variáveis e, desse modo, podem levar a resultados significativos que de outra forma poderiam ser não-significativos. Portanto, se os dados omissos são aleatórios, você deve considerar a exclusão dos casos. O SPSS permite tanto a exclusão de casos em lista **Exclude cases listwise**, em que um participante com dados omissos é excluído em todas as variáveis, quanto a exclusão por pares **Exclude cases pairwise**, em que um participante é excluído somente na variável com dados omissos. Se você excluir casos por pares, suas estimativas podem ir para qualquer lugar, portanto, talvez seja mais seguro excluir casos em lista, a não ser que isso resulte numa perda massiva de dados.

As duas opções finais dizem respeito a como os coeficientes são mostrados. Por omissão, o SPSS irá listar as variáveis na ordem que elas entraram no editor de dados. Usualmente,

esse formato é mais conveniente. Entretanto, na interpretação dos fatores, é algumas vezes útil listar as variáveis por tamanho. Selecionando **Sorted by size** (Ordenando por tamanho), o SPSS irá ordenar as variáveis pelas cargas dos fatores. Na verdade, ele faz essa ordenação de uma forma bastante inteligente, de modo que todas as variáveis que estão altamente carregadas no mesmo fator são apresentadas juntas. A segunda opção é **Supress absolute values less than** (Eliminar valores absolutos menores que) um valor especificado (por omissão, 0,1). Essa opção assegura que cargas dos fatores no intervalo de $\pm 0,1$ não serão mostradas na saída. Novamente, essa opção é útil para ajudar na interpretação. O valor por omissão provavelmente é adequado, mas para a sua primeira análise, recomendo trocá-lo por 0,4 (para fins de interpretação) ou por um valor que reflita o valor esperado de uma carga dos fatores significativa considerando o tamanho da amostra (veja a Seção 15.3.6.2). Isso tornará a interpretação mais simples. Se quiser, você pode executar novamente a análise e ajustar esse valor mais baixo para ver se não esqueceu nada (como uma carga de 0,39). Para esse exemplo, ajuste o valor em 0,4.

15.6 INTERPRETANDO SAÍDAS DO SPSS ②

Selecione as mesmas opções que selecionou e execute a análise dos fatores com a rotação ortogonal. Feito isso, selecione a opção **Direct Oblimin** (*Oblimin* direta) conforme Figura 15.8 e repita a análise. Você deve obter duas saídas idênticas em todos os aspectos com exceção daquela que usou a rotação ortogonal e a que usou a oblínua.

Para economizar espaço nesta seção, deixei as opções por omissão do SPSS, assim, cada variável é referida somente pelo seu nome no editor de dados (por exemplo, Q12). Na saída que você irá obter, você verá que o SPSS usa como rótulo de valor (a própria pergunta) em todas as saídas. Quando usar as saídas neste capítulo, lembre que Q1 representa a pergunta 1, Q2 representa a pergunta 2 e Q17 representa a

pergunta 17. Voltando à Figura 15.4 e combinando o número da pergunta com o nome da variável, você poderá identificar cada pergunta.

15.6.1 Análise preliminar ②

A primeira parte da saída se refere à análise dos dados, teste das hipóteses e adequação da amostra. Você encontrará muitas tabelas grandes (ou matrizes) que informam aspectos interessantes sobre nossos dados. Se você selecionou a opção **Univariate descriptives** (Descritivas univariadas) como na Figura 15.6, a primeira tabela conterá estatísticas descritivas para cada variável (a média, o desvio padrão e o número de casos). Essa tabela não está incluída aqui, mas você deve ser capaz de interpretá-la. A tabela também inclui o número de casos omisso; esse resumo é uma maneira útil para determinar o total de dados omisso.

A Saída 15.1 do SPSS mostra a matriz-R (ou matriz de correlação)⁷ produzida usando as opções **Coefficients and Significance levels** (Coeficientes e Níveis de significância) como na Figura 15.6. A parte superior da tabela apresenta o coeficiente correlação de Pearson entre todos os pares de perguntas, enquanto a parte inferior mostra a significância unilateral desses coeficientes. **Você deve estar confortável com a ideia de que para fazer uma análise de fatores, precisamos ter variáveis que se correlacionem muito bem, mas não perfeitamente. Também, quaisquer variáveis que não se correlacionem devem ser eliminadas.** Portanto, podemos usar essa matriz de correlações para checar o padrão dos relacionamentos. **A maneira mais fácil de fazer isso é examinar cuidadosamente os valores de significância e procurar por qualquer variável para qual a maioria dos valores seja maior do que 0,05. Depois, examine os coeficientes de correlação e procure por qualquer um maior do que 0,9. Se algum for encontrado, você deve ficar atento ao problema da singularidade dos dados: verifique o determinante da**

⁷ Para ganhar espaço, retirei várias colunas de dados das tabelas grandes; somente dados para as primeiras e as últimas cinco perguntas do questionário estão incluídos.

Saída 15.1 do SPSS

Correlation Matrix^a (Matriz de Correlações)

		Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Correlation (Correlação)	Q1	1.000	-0.099	-0.337	0.436	0.402	-0.189	0.214	0.329	-0.104	-0.004
	Q2	-0.099	1.000	0.318	-0.112	-0.119	0.203	-0.202	-0.205	0.231	0.100
	Q3	-0.337	0.318	1.000	-0.380	-0.310	0.342	-0.325	-0.417	0.204	0.150
	Q4	0.436	-0.112	-0.380	1.000	0.401	-0.186	0.243	0.410	-0.098	-0.034
	Q5	0.402	-0.119	-0.310	0.401	1.000	-0.165	0.200	0.335	-0.133	-0.042
	Q6	0.217	-0.074	-0.227	0.278	0.257	-0.167	0.101	0.272	-0.165	-0.069
	Q7	0.305	-0.159	-0.382	0.409	0.339	-0.269	0.221	0.483	-0.168	-0.070
	Q8	0.331	-0.050	-0.259	0.349	0.269	-0.159	0.175	0.296	-0.079	-0.050
	Q9	-0.92	0.315	0.300	-0.125	-0.096	0.249	-0.159	-0.136	0.257	0.171
	Q10	0.214	-0.084	-0.193	0.216	0.258	-0.127	0.084	0.193	-0.131	-0.062
	Q11	0.357	-0.144	-0.351	0.369	0.298	-0.200	0.255	0.346	-0.162	-0.086
	Q12	0.345	-0.195	-0.410	0.442	0.347	-0.267	0.298	0.441	-0.167	-0.046
	Q13	0.355	-0.143	-0.318	0.344	0.302	-0.227	0.204	0.374	-0.195	-0.053
	Q14	0.338	-0.165	-0.371	0.351	0.315	-0.254	0.226	0.399	-0.170	-0.048
	Q15	0.246	-0.165	-0.312	0.334	0.261	-0.210	0.206	0.300	-0.168	-0.062
	Q16	0.499	-0.168	-0.419	0.416	0.395	-0.267	0.265	0.421	-0.156	-0.082
	Q17	0.371	-0.087	-0.327	0.383	0.310	-0.163	0.205	0.363	-0.126	-0.092
	Q18	0.347	-0.164	-0.375	0.382	0.322	-0.257	0.235	0.430	-0.160	-0.080
	Q19	-0.189	0.203	0.342	-0.186	-0.165	1.000	-0.249	-0.275	0.234	0.122
	Q20	0.214	-0.202	-0.325	0.243	0.200	-0.249	1.000	0.468	-0.100	-0.035
	Q21	0.329	-0.205	-0.417	0.410	0.335	-0.275	0.468	1.000	-0.129	-0.068
	Q22	-0.104	0.231	0.204	-0.098	-0.133	0.234	-0.100	-0.129	1.000	0.230
	Q23	-0.004	0.100	0.150	-0.034	-0.042	0.122	-0.035	-0.068	0.230	1.000
Sig. (1-tailed) (Sig. Unilateral)	Q1		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.410
	Q2	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q3	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q4	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.043
	Q5	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.017
	Q6	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
	Q7	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
	Q8	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.005
	Q9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000
	Q10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	Q11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009
	Q13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
	Q14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
	Q15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
	Q16	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q17	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q18	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Q19	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000
	Q20	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.039
	Q21	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000
	Q22	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000
	Q23	0.410	0.000	0.000	0.043	0.017	0.000	0.039	0.000	0.000	

a Determinant = 5.271E-04 (Determinante = 0,0005271)

matriz de correlação e, se necessário, elimine uma das duas variáveis que podem estar causando o problema. O determinante está listado na parte de baixo da matriz (pisque e você irá perdê-lo). Para esses dados, ele vale 5,271E-04 (que é 0,0005271), que é maior do que o valor mínimo necessário de 0,00001

(veja a Seção 15.5)⁸ Portanto, podemos estar confiantes de que a multicolinearidade não

⁸ Rigorosamente falando, o determinante ou a matriz de correlação deve ser checada somente na análise dos fatores: na análise pura dos componentes principais ela não é relevante.

será um problema para esses dados. Resumindo, todas as questões no QAS se correlacionam muito bem com todas as outras (isso ocorre em parte devido ao grande tamanho da amostra) e nenhum dos coeficientes de correlação é particularmente grande; portanto, não há necessidade de considerar a eliminação de qualquer pergunta nesse estágio.

A Saída 15.2 do SPSS mostra o inverso da matriz de correlação (R^{-1}), que é usada em vários cálculos (incluindo os escores dos fatores – veja a Seção 15.2.4). Essa matriz é produzida usando a opção **Inverse** (Inverso), Figura 15.6, mas ela é somente útil se você quer ter algumas percepções sobre como os cálculos são feitos na análise de fatores. A maioria de nós tem coisas mais interessantes a fazer do que saber como realizar os cálculos na análise de fatores e o uso dessa matriz é mínimo – assim, ignore-a!

A Saída 15.3 do SPSS mostra várias partes importantes da saída: o KMO, o teste de esfericidade de Bartlett e as matrizes de covariâncias e anti-imagem das correlações (note que essas matrizes foram editadas para

conter somente a primeira e as últimas cinco variáveis). As matrizes de covariâncias e a anti-imagem das correlações fornecem informações semelhantes (lembre o relacionamento entre covariância e correlação); somente a matriz de anti-imagem das correlações precisa ser estudada em detalhes porque ela é a mais informativa. Essas tabelas são obtidas utilizando opções **KMO and Bartlett's test of sphericity** (KMO e o teste de esfericidade de Bartlett) e **Anti-image** (Anti-imagem), como na Figura 15.6.

Aprendemos sobre a estatística KMO na Seção 15.4.1.1 e vimos que Kaiser (1974) recomenda um mínimo de 0,5, e que os valores entre 0,5 e 0,7 são medíocres, valores entre 0,7 e 0,8 são bons, valores entre 0,8 e 0,9 são ótimos e valores acima de 0,9 são excelentes (veja Hutcheson e Sofroniou, 1999, p. 224-225, para mais detalhes). Para esses dados, o valor é 0,93, o qual está no limite de excelência; assim, devemos estar confiantes de que a análise dos fatores é apropriada para esses dados.

Mencionei que o KMO pode ser calculado para uma ou mais variáveis. Os valores do

Saída 15.2 do SPSS

Inverse of Correlation Matrix^a (Inversa da Matriz de Correlações)

	Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Q1	1.595	-0.028	0.087	-0.268	-0.233	0.017	-0.024	0.011	0.002	-0.078
Q2	-0.028	1.232	-0.224	-0.057	0.013	-0.037	0.076	0.062	-0.148	-0.003
Q3	0.087	-0.224	1.661	0.138	0.057	-0.175	0.118	0.122	-0.009	-0.103
Q4	-0.268	-0.057	0.138	1.626	-0.203	-0.049	-0.006	-0.149	-0.045	-0.023
Q5	-0.233	0.013	0.057	-0.203	1.410	-0.024	-0.016	-0.074	0.045	-0.006
Q6	0.034	-0.078	-0.072	-0.011	-0.055	-0.023	0.080	0.069	0.058	0.025
Q7	0.039	0.025	0.127	-0.152	-0.072	0.105	0.077	-0.386	0.019	-0.012
Q8	-0.087	-0.051	-0.013	-0.134	-0.045	0.074	0.034	-0.039	-0.035	0.003
Q9	-0.023	-0.242	-0.208	0.043	-0.027	-0.141	0.050	-0.047	-0.156	-0.110
Q10	-0.017	-0.015	-0.023	0.009	-0.124	-0.012	0.056	0.026	0.023	0.017
Q11	-0.075	0.061	0.121	-0.041	0.000	-0.010	-0.140	-0.009	0.055	0.015
Q12	-0.011	0.046	0.147	-0.259	-0.091	0.060	-0.100	-0.141	0.026	-0.038
Q13	-0.145	-0.011	-0.055	0.040	0.007	0.014	0.028	-0.061	0.077	-0.042
Q14	-0.064	0.033	0.115	-0.007	-0.040	0.063	0.002	-0.110	0.041	-0.034
Q15	0.138	0.050	0.013	-0.098	0.021	0.013	-0.054	0.058	0.034	-0.030
Q16	-0.454	-0.017	0.142	-0.063	-0.155	0.071	-0.008	-0.158	-0.005	0.033
Q17	-0.084	-0.045	0.063	-0.064	-0.030	-0.074	0.025	-0.077	0.015	0.080
Q18	-0.041	0.028	0.070	-0.044	0.004	0.047	-0.004	-0.136	-0.037	0.033
Q19	0.017	-0.037	-0.175	-0.049	-0.024	1.264	0.120	0.048	-0.141	-0.045
Q20	-0.024	0.076	0.118	-0.006	-0.016	0.120	1.370	-0.511	-0.014	-0.034
Q21	0.011	0.062	0.122	-0.149	-0.074	0.048	-0.511	1.830	-0.036	0.018
Q22	0.002	-0.148	-0.009	-0.045	0.045	-0.141	-0.014	-0.036	1.200	-0.202
Q23	-0.078	-0.003	-0.103	-0.023	-0.006	-0.045	-0.034	0.018	-0.202	1.094

Saída 15.3 do SPSS

KMO and Bartlett's Test (KMO e teste de Bartlett)

Kaiser-Meyer-Olkin Measure of Sampling Adequacy (Adequação da Amostra de Kaiser-Meyer-Olkin)	0.930
Bartlett's Test of Sphericity (Teste de Esfericidade de Bartlett)	Approx. Chi-Square (Qui-quadrado aproximado)
	19334.492
	Df (gl)
	253
	Sig. (Sig.)
	0.000

Anti-Image Matrices (Matrizes Anti-Imagem)

Anti-Image Correlation (Correlações Anti-Imagem)

	Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Q1	0.930	-2.00E-02	5.320E-02	-0.167	-0.156	1.231E-02	-1.61E-02	6.436E-03	1.459E-03	-5.92E-02
Q2	-2.00E-02	0.875	-0.157	-4.05E-02	1.019E-02	-2.93E-02	5.877E-02	4.139E-02	-0.121	-2.39E-03
Q3	5.320E-02	-0.157	0.951	8.381E-02	3.725E-02	-0.121	7.790E-02	6.983E-02	-6.57E-03	-7.63E-02
Q4	-0.167	-4.05E-02	8.381E-02	0.955	-0.134	-3.42E-03	-4.07E-03	-8.64E-02	-3.25E-02	-1.71E-02
Q5	-0.156	1.019E-02	3.725E-02	-0.134	0.960	-1.77E-02	-1.14E-02	-4.59E-02	3.455E-02	-4.75E-03
Q6	2.016E-02	-5.33E-02	-4.23E-02	-6.77E-03	-3.52E-02	1.53E-02	5.144E-02	3.867E-02	3.984E-02	1.846E-02
Q7	2.264E-02	1.648E-02	7.159E-02	-8.66E-02	-4.43E-02	6.802E-02	4.796E-02	-0.208	1.278E-02	-8.07E-03
Q8	-4.91E-02	-3.28E-02	-7.28E-03	-7.49E-02	-2.72E-02	4.696E-02	2.094E-02	-2.05E-02	-2.28E-02	2.202E-03
Q9	-1.64E-02	-0.193	-0.142	2.960E-02	-1.98E-02	-0.111	3.803E-02	-3.06E-02	-0.126	-9.25E-02
Q10	-1.22E-02	-1.21E-02	-1.62E-02	6.031E-03	-9.32E-02	-9.16E-03	4.272E-02	1.699E-02	1.867E-02	1.462E-02
Q11	-4.07E-02	3.757E-02	6.429E-02	-2.22E-02	-3.27E-05	-5.81E-03	-8.18E-03	-4.58E-03	3.425E-02	9.600E-03
Q12	-6.62E-03	3.129E-02	8.667E-02	-0.154	-5.83E-02	4.030E-02	-6.47E-02	-7.93E-02	1.802E-02	-2.79E-02
Q13	-8.51E-02	-7.60E-03	-3.16E-02	2.303E-02	4.238E-03	8.994E-03	1.785E-02	-3.33E-02	5.233E-02	-2.99E-02
Q14	-3.98E-02	2.313E-02	6.946E-02	-4.26E-03	-2.63E-02	4.399E-02	1.124E-03	-6.32E-02	2.896E-02	-2.56E-02
Q15	8.860E-02	3.680E-02	8.159E-03	-6.21E-02	1.413E-02	9.384E-03	-3.74E-02	3.483E-02	2.485E-02	-2.36E-02
Q16	-0.264	-1.14E-02	8.057E-02	-3.62E-02	-9.57E-02	4.659E-02	-5.30E-03	-8.54E-02	-3.26E-03	2.298E-02
Q17	-4.74E-02	-2.87E-02	3.467E-02	-3.55E-02	-1.80E-02	-4.68E-02	1.004E-02	-4.06E-02	1.004E-02	5.457E-02
Q18	-2.31E-02	1.825E-02	3.880E-02	-2.45E-02	2.489E-03	2.980E-02	-2.54E-03	-7.17E-02	-2.40E-02	2.257E-02
Q19	1.231E-02	-2.93E-02	-0.121	-3.42E-02	-1.77E-02	0.941	9.081E-02	3.145E-02	-0.115	-3.84E-02
Q20	-1.61E-02	5.877E-02	7.790E-02	-4.07E-03	-1.14E-02	9.081E-02	0.889	-0.323	-1.12E-02	-2.81E-02
Q21	6.436E-03	4.139E-02	6.983E-02	-8.64E-02	-4.59E-02	3.145E-02	-0.323	0.929	-2.41E-02	1.275E-02
Q22	1.459E-03	-0.121	-6.57E-03	-3.25E-02	3.455E-02	-0.115	-1.12E-02	-2.41E-02	0.078	-0.176
Q23	-5.92E-02	-2.39E-03	-7.63E-02	-1.71E-02	-4.75E-03	-3.84E-02	-2.81E-02	1.275E-02	-0.176	0.766

KMO para variáveis individuais são apresentados na diagonal da matriz anti-imagem e das correlações (marquei os valores em negrito). Esses valores fazem da matriz anti-imagem e das correlações uma parte extremamente importante da saída (embora a matriz anti-imagem das covariâncias possa ser ignorada). Além de verificar todas as estatísticas KMO, também é importante examinar os elementos diagonais da matriz anti-imagem das correlações: o valor deve estar acima de um mínimo de 0,5 para todas as variáveis (e preferencialmente mais alto). Para esses dados, todos os valores estão bem acima de 0,5, o que é uma ótima notícia! Se você encontrar qualquer variável com valor abaixo de 0,5, deve considerar excluí-la da análise (ou executar a análise com e sem aquela variável e ver a diferença). A remoção de uma

variável afeta as estatísticas do KMO, assim, se você realmente remover a variável, tenha certeza de re-examinar a nova matriz anti-imagem das correlações. Para o restante da matriz anti-imagem das correlações, os elementos fora da diagonal representam as correlações parciais entre as variáveis. Para uma boa análise dos fatores, queremos que essas correlações sejam bem pequenas (quanto menor, melhor). Assim, numa checagem final, você apenas verifica se os elementos fora da diagonal são pequenos (eles devem ser para esses dados).

As medidas de Bartlett testam a hipótese nula de que a matriz de correlações original é uma matriz identidade. Para a análise de fatores funcionar, precisamos de relacionamentos entre variáveis, e se a matriz-R é uma matriz identidade, todos os coeficientes de correla-

Dica da Samanta Ferrinho



- **Examine cuidadosamente a Matriz de Correlações:** procure variáveis que não se correlacionam com quaisquer outras variáveis ou se correlacionem altamente ($r = 0,9$) com uma ou mais variáveis. Na análise de fatores, cheque se o determinante dessa matriz é maior do que 0,00001; se for, a multicolinearidade não será um problema.
- **Na tabela rotulada de KMO e Teste de Bartlett, a estatística de KMO deve ser maior do que 0,5 no mínimo; se ela não for, colete mais dados. O teste de Bartlett para a esfericidade deve ser significativo (o valor de Sig. deve ser menor do que 0,05). Você também pode verificar a estatística de KMO para variáveis individuais olhando para a diagonal das Matrizes de Autoimagem: novamente, esses valores devem estar acima de 0,5 (isso é útil para identificar variáveis problemáticas se todo o KMO for insatisfatório).**

ção serão zero. Portanto, queremos que esse teste seja *significativo* (isto é, tenha um valor de significância menor do que 0,05). Um teste significativo nos informa que a matriz-R não é uma matriz identidade; portanto, existem relacionamentos entre variáveis que esperamos incluir na análise. Para esses dados, o teste de Bartlett é altamente significativo ($p < 0,001$) e, portanto, a análise dos fatores é apropriada.

15.6.2 Extração de fatores ②

A primeira parte do processo de extração do fator é determinar os componentes lineares dentro de um conjunto de dados (os autovetores) calculando os autovalores da matriz-R (veja a Seção 15.3.4). Sabemos que há tantos componentes (autovetores) na matriz-R quanto variáveis, mas a maioria não terá importância. Para determinar a importância de um vetor particular, olhamos para a magnitude do autovalor associado. Podemos, então, aplicar um critério para determinar quais fatores devemos reter e quais devemos descartar. Por omissão, o SPSS usa o critério de fatores retidos de Kaiser com autovalores maiores do que 1 (veja a Figura 15.7).

A Saída 15.4 do SPSS lista os autovalores associados com cada componente linear (fator) antes da extração, depois da extração e depois da rotação. Antes da extração, o SPSS identificou 23 componentes lineares dentro do conjunto de dados (sabemos que deveria

haver o mesmo número de autovetores que de variáveis, assim, haverá tantos fatores quanto variáveis – veja a Seção 15.3.4). Os autovalores associados com cada fator representam a variância explicada por aquele componente linear particular, e o SPSS também mostra o autovvalor em termos de percentagem da variância explicada (assim, o fator 1 explica 31,696% da variância total). Deve estar claro que os primeiros fatores explicam quantias relativamente grandes das variâncias (especialmente o fator 1), enquanto os fatores subsequentes explicam somente pequenas quantias das variâncias. O SPSS, então, extrai todos os fatores com autovalores maiores do que 1, o que nos deixa com quatro fatores. Os autovalores associados com esses fatores são, novamente, exibidos (e a percentagem da variância explicada) na coluna Somas Extraídas das Cargas ao Quadrado (**Extraction Sums Of Squared Loadings**). Os valores nessa parte da tabela são os mesmos de antes da extração, exceto que os valores para os fatores descartados são ignorados (por isso a tabela está em branco depois do quarto fator). Na parte final da tabela (denominada Rotação das Somas das Cargas ao Quadrado (**Rotation Sums of Squared Loadings**)), os autovalores dos fatores após rotação são exibidos. A rotação tem o efeito de otimizar a estrutura do fator e uma consequência para esses dados é que a relativa importância dos quatro fatores é equalizada. Antes da rotação, o fator 1 é responsável

Saída 15.4 do SPSS

Total Variance Explained (Variância Total Explicada)

<i>Component (Componente)</i>	<i>Initial Eigenvalues (Autovalores Iniciais)</i>			<i>Extraction Sums of Squared Loadings (Somos Extraídas das Cargas ao Quadrado)</i>			<i>Rotation Sums of Squared Loadings (Somos Rotacionadas das Cargas ao Quadrado)</i>		
	<i>Total (Total)</i>	<i>% of Variance (% da Variância)</i>	<i>Cumulative % (%) Acumulada)</i>	<i>Total (Total)</i>	<i>% of Variance (% da Variância)</i>	<i>Cumulative % (%) Acumulada)</i>	<i>Total (Total)</i>	<i>% of Variance (% da Variância)</i>	<i>Cumulative % (%) Acumulada)</i>
1	7.290	31.696	31.696	7.290	31.696	31.696	3.730	16.219	16.219
2	1.739	7.560	39.256	1.739	7.560	39.256	3.340	14.523	30.742
3	1.317	5.725	44.981	1.317	5.725	44.981	2.553	11.099	41.842
4	1.227	5.336	50.317	1.227	5.336	50.317	1.949	8.475	50.317
5	0.988	4.295	54.612						
6	0.895	3.893	58.504						
7	0.806	3.502	62.007						
8	0.783	3.404	65.410						
9	0.751	3.265	68.676						
10	0.717	3.117	71.793						
11	0.684	2.972	74.765						
12	0.670	2.911	77.676						
13	0.612	2.661	80.337						
14	0.578	2.512	82.849						
15	0.549	2.388	85.236						
16	0.523	2.275	87.511						
17	0.508	2.210	89.721						
18	0.456	1.982	91.704						
19	0.424	1.843	93.546						
20	0.408	1.773	95.319						
21	0.379	1.650	96.969						
22	0.364	1.583	98.552						
23	0.333	1.448	100.000						

Método de Extração: Análise dos Componentes Principais.

por consideravelmente mais variâncias do que as três remanescentes (31,696% comparado com 7,560%, 5,725% e 5,336%); entretanto, depois da extração é responsável por somente 16,219% da variância (comparada a 14,523%, 11,099% e 8,475%, respectivamente).

A Saída do SPSS 15.5 mostra uma tabela das comunalidades antes e depois da extração. Lembre que a comunalidade é a proporção da variância comum dentro da variável (veja a Seção 15.3.1). A análise dos componentes principais trabalha com a suposição inicial de que toda a variância é comum; portanto, antes da extração as comunalidades são todas 1 (veja a coluna Inicial (**Initial**)). De fato, todas as variâncias associadas com uma variável são supostamente uma variância comum. Depois que os fatores foram extraídos, temos uma ideia melhor de quanta variância é, na reali-

dade, comum. As comunalidades na coluna rotulada de Extração (*Extraction*) refletem essa variância comum. Assim, por exemplo, podemos dizer que 43,5% da variância associada à pergunta 1 é uma variância comum ou compartilhada. Outra maneira de analisar essas comunalidades é em termos da proporção da variância explicada pelos fatores adjacentes. Antes da extração existem tantos fatores quanto variáveis, assim, toda a variância é explicada pelos fatores e todas as comunalidades são 1. Entretanto, depois da extração alguns dos fatores são descartados e alguma informação é perdida. Os fatores retidos não podem explicar toda a variância presente nos dados, mas podem explicar uma parte. A quantidade de variância em cada variável que pode ser explicada pelos fatores retidos é representada pelas comunalidades depois da extração.

A Saída 15.5 do SPSS também mostra a matriz de componentes antes da rotação. Essa matriz contém as cargas de cada variável em cada fator. Por omissão, o SPSS exibe todas as cargas; entretanto, solicitamos que todas as cargas menores do que 0,4 sejam eliminadas da saída (veja a Figura 15.10), assim, existem espaços em branco para muitas das cargas. Essa matriz não é importante para interpretação, mas é interessante notar que antes da rotação a maioria das variáveis já estava com altas cargas no primeiro fator (isso porque esse fator é responsável pela maior parte da variância na Saída 15.4 do SPSS).

Nesse estágio, o SPSS extraiu quatro fatores. A análise de fatores é uma ferramenta de exploração, portanto, ela deveria guiar o pesquisador na tomada de várias decisões: você não deveria deixar o computador fazê-las. Uma decisão importante é o número de fatores a serem extraídos. Na Seção 15.3.5, vimos

vários critérios para determinar a importância dos fatores. Pelo critério de Kaiser, deveríamos extrair quatro fatores e isso é o que o SPSS fez. Entretanto, esse critério é preciso quando existem menos do que 30 variáveis e as communalidades após a extração são maiores do que 0,7, ou quando o tamanho da amostra excede a 250 e a média da communalidade é maior do que 0,6. As communalidades são exibidas na Saída 15.5 do SPSS e nenhuma excede a 0,7. A média das communalidades pode ser encontrada somando-as e dividindo-as pelo número de communalidades ($11,573/23 = 0,503$). Assim, em ambos os casos a regra de Kaiser pode não ser precisa. Entretanto, você deve considerar a amostra enorme que temos porque a pesquisa no critério de Kaiser recomenda amostras bem menores. Pelo critério de Jolliffe (reter fatores com autovalores maiores do que 0,7), devemos reter 10 fatores (veja a Saída 15.4 do SPSS), mas esse critério não é mais recomendável que

Saída 15.5 do SPSS

Communalities
(Comunalidades)

	Initial (Inicial)	Extraction (Extração)
Q1	1.000	0.435
Q2	1.000	0.414
Q3	1.000	0.530
Q4	1.000	0.469
Q5	1.000	0.343
Q6	1.000	0.654
Q7	1.000	0.545
Q8	1.000	0.739
Q9	1.000	0.484
Q10	1.000	0.335
Q11	1.000	0.690
Q12	1.000	0.513
Q13	1.000	0.536
Q14	1.000	0.488
Q15	1.000	0.378
Q16	1.000	0.487
Q17	1.000	0.683
Q18	1.000	0.597
Q19	1.000	0.343
Q20	1.000	0.484
Q21	1.000	0.550
Q22	1.000	0.464
Q23	1.000	0.412

Método de Extração: Componentes Principais.

Component Matrix^a
(Matriz dos componentes)

	Component (Componente)			
	1	2	3	4
Q18	0.701			
Q07	0.685			
Q16	0.679			
Q13	0.673			
Q12	0.669			
Q21	0.658			
Q14	0.656			
Q11	0.652			-0.400
Q17	0.643			
Q04	0.634			
Q03	-0.629			
Q15	0.593			
Q01	0.586			
Q05	0.556			
Q08	0.549	0.401		-0.417
Q10	0.437			
Q20	0.436		-0.404	
Q19	-0.427			
Q09		0.627		
Q02		0.548		
Q22		0.465		
Q06	0.562		0.571	
Q23				0.507

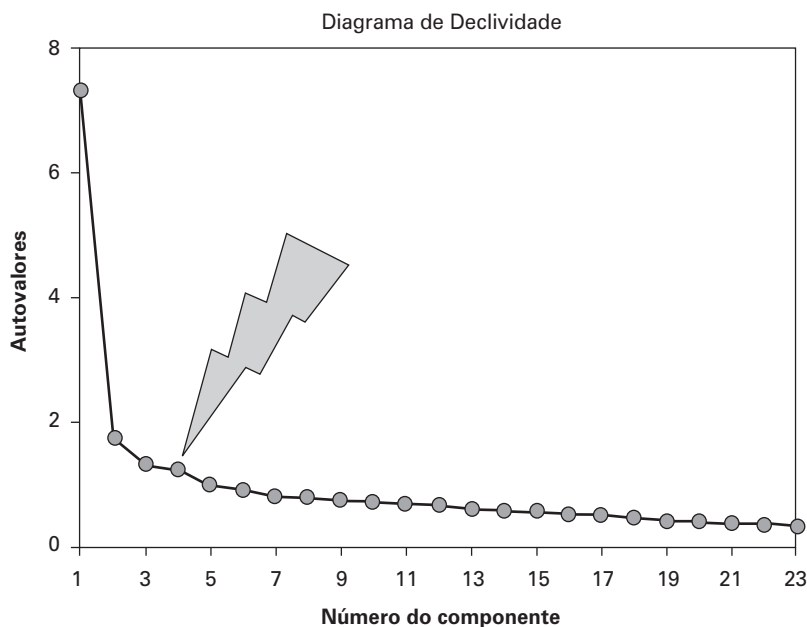
Método de Extração: Análise dos Componentes Principais.
a 4 componentes extraídos.

o de Kaiser. Como um guia final, podemos usar o diagrama de declividade que pedimos ao SPSS produzir usando a opção da Figura 15.7. O diagrama de declividade é mostrado na Saída 15.6 do SPSS com um raio indicando o ponto de inflexão na curva. Essa curva é difícil de interpretar porque ela começa a ter uma cauda após três fatores, mas há outra queda após o quarto fator antes que um platô estável seja

alcançado. Portanto, podemos, provavelmente, justificar reter dois ou quatro fatores. Dado a amostra grande, é provavelmente seguro aceitar o critério de Kaiser; entretanto, seria aconselhável executar novamente a análise especificando que o SPSS extraia somente dois fatores (veja a Figura 15.7) e comparar os resultados.

A Saída 15.7 do SPSS mostra uma versão editada da matriz de correlações reproduzida

Saída 15.6 do SPSS



Dicas da Samantha Ferrinho: Extração dos Fatores



- Para decidir quantos fatores extrair, olhe para a tabela *Communalities* (Comunalidades) e a coluna *Extraction* (Extração). Se todos esses valores são 0,7 ou acima e você tem menos do que 30 variáveis, a opção por omissão do SPSS para a extração dos fatores é boa (o critério de Kaiser da retenção dos fatores com autovalores maiores do que 1). Do mesmo modo, se o seu tamanho da amostra excede 250 e a média das comunalidades é 0,6 ou maior, a opção por omissão é ótima. Alternativamente, com 200 ou mais participantes o diagrama de declividade pode ser usado.
- Verifique a parte de baixo da tabela *Reproduced Correlations* (Correlações Reproduzidas) para a percentagem dos “resíduos não redundantes com valores absolutos $> 0,05$ ”. Essas percentagens devem ser menores do que 50%, e quanto menores ela forem, melhor.

Saída 15.7 do SPSS

Reproduced Correlations (Correlações reproduzidas)

	Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Reproduced Correlation (Correlação reproduzida)										
Q01	0.435^b	-0.112	-0.372	0.447	0.376	-0.204	0.342	0.449	-2.54E-02	4.463E-02
Q02	-0.112	0.414^b	0.380	-0.134	-0.122	0.357	-0.301	-0.254	3.333	0.246
Q03	-0.372	0.380	0.530^b	-0.399	-0.345	0.403	-0.440	-0.488	0.275	0.158
Q04	0.447	-0.134	-0.399	0.469^b	-0.399	-0.231	0.353	-0.480	-4.97E-02	4.179E-02
Q05	0.376	-0.122	-0.345	0.399	0.343^b	-0.207	0.292	0.412	-6.03E-02	2.794E-02
Q06	0.218	-3.34E-02	-0.200	0.278	0.312	-0.147	-2.06E-02	0.244	-0.299	-8.17E-02
Q07	0.366	-0.148	-0.373	0.419	0.380	-0.254	0.219	0.430	-0.179	-3.72E-02
Q08	0.412	2.192E-03	-0.270	0.390	0.312	-0.104	0.164	0.282	-989E-02	-0.136
Q09	-4.25E-02	0.430	0.352	-7.31E-02	-7.97E-02	0.363	-0.218	-0.191	0.417	0.323
Q10	0.172	-6.12E-02	-0.181	0.212	0.205	-0.137	5.763E-03	0.188	-0.197	-0.110
Q11	0.423	-9.74E-02	-0.357	0.419	0.348	-0.198	0.200	0.342	-0.209	-0.202
Q12	0.402	-0.219	-0.440	0.448	0.397	-0.302	0.354	0.503	-0.136	1.064E-02
Q13	0.347	-0.122	-0.342	0.395	0.360	-0.231	0.163	0.384	-0.203	-7.88E-02
Q14	0.362	-0.155	-0.373	0.411	0.370	-0.254	0.241	0.431	-0.159	-2.16E-02
Q15	0.311	-0.158	-0.337	0.343	0.306	-0.236	0.175	0.336	-0.230	-0.141
Q16	0.440	-0.217	-0.458	0.466	0.400	-0.299	0.373	0.494	-0.152	-4.93E-02
Q17	0.439	-4.80E-02	-0.331	0.434	0.359	-0.162	0.196	0.347	-0.145	-0.140
Q18	0.368	-0.149	0.376	0.424	0.388	-0.259	0.215	0.439	-0.183	-3.16E-02
Q19	-0.204	0.357	0.403	-0.231	-0.207	0.343^b	-0.308	-0.324	0.294	0.196
Q20	0.342	-0.301	-0.440	0.353	0.292	-0.308	0.484^b	0.457	-6.78E-02	2.125E-02
Q21	0.449	-0.254	-0.488	0.480	0.412	-0.324	0.457	0.550^b	-9.59E-02	3.196E-02
Q22	-2.54E-02	0.333	0.275	-4.97E-02	-6.03E-02	0.294	-6.78E-02	-9.59E-02	0.464^b	0.408
Q23	4.463E-02	0.246	0.158	4.179E-02	2.794E-02	0.196	2.125E-02	3.196E-02	0.408	0.412^b
Residual ^a (Resíduos)										
Q01	1.291E-02	1.291E-02	3.503E-02	-1.13E-02	2.694E-02	1.500E-02	-0.128	-0.120	-7.90E-02	-4.91E-02
Q02	3.503E-02	-6.15E-02	-6.15E-02	2.169E-02	2.930E-03	-0.153	9.905E-02	4.922E-02	-0.102	-0.147
Q03	-1.13E-02	2.169E-02	1.893E-02	1.893E-02	3.485E-02	-6.11E-02	0.115	7.125E-02	-7.09E-02	-8.01E-03
Q04	2.694E-02	2.930E-02	1.785E-03	1.785E-03	1.785E-03	4.510E-02	-0.110	-6.98E-02	-4.87E-02	-7.56E-02
Q05	-8.72E-04	-4.08E-02	-2.68E-02	1.197E-04	-1.55E-02	4.136E-02	-9.24E-02	-7.76E-02	-7.22E-02	-6.96E-02
Q06	-6.11E-02	-1.12E-02	-9.13E-03	-1.00E-02	-4.07E-02	-2.01E-02	0.122	2.875E-02	4.344E-02	1.304E-02
Q07	-8.11E	-5.18E-02	1.106E-02	-4.07E-02	-4.36E-02	-1.46E-02	1.628E-03	5.261E-02	1.040E-02	-3.31E-02
Q08						-5.59E-02	1.145E-02	1.407E-02	1.976E-02	8.616E-02

(Continua)

Saída 15.7 do SPSS Continuação

Correlations Reproduced (Correlações reproduzidas)

	Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Q09	-4.99E-02	-0.115	-5.19E-02	-5.15E-02	-1.60E-02	-0.114	5.965E-02	-0.161	-0.152	4.844E-02
Q10	4.194E-02	-2.28E-02	-1.29E-02	3.359E-03	5.299E-02	9.852E-03	7.830E-02	4.732E-02	6.599E-02	0.116
Q11	-6.60E-02	-4.64E-02	6.094E-03	-5.05E-02	-5.01E-02	-2.07E-02	5.579E-02	4.747E-03	4.741E-02	-5.71E-02
Q12	-5.67E-02	2.375E-02	3.036E-02	-6.06E-03	-5.01E-02	3.574E-02	-5.63E-02	-6.21E-02	-3.10E-02	2.583E-02
Q13	7.634E-03	-2.12E-02	2.367E-02	-5.11E-02	-5.81E-02	4.134E-03	-9.81E-03	7.402E-03	2.583E-02	-1.08E-02
Q14	-2.39E-02	-9.23E-03	1.909E-03	-5.97E-02	5.51E-02	-2.55E-04	-1.52E-02	-3.16E-02	-1.08E-02	7.91E-02
Q15	-6.49E-02	-7.26E-02	2.494E-02	-8.76E-03	2.658E-02	3.088E-02	-3.59E-02	6.223E-02	7.91E-02	4.880E-02
Q16	5.880E-02	4.966E-02	3.929E-02	-4.99E-02	-5.04E-02	3.159E-03	-3.22E-02	1.601E-02	1.900E-02	-4.88E-02
Q17	-6.86E-02	-3.90E-02	3.147E-03	-5.16E-02	-4.89E-02	-1.23E-03	9.384E-03	-8.65E-02	2.343E-02	-7.31E-02
Q18	-2.04E-02	1.49E-02	9.980E-04	-4.24E-02	-659E-02	2.821E-03	1.985E-02	4.882E-02	-6.01E-02	-5.59E-02
Q19	1.500E-02	-0.153	-6.11E-02	4.510E-02	4.136E-02	5.985E-02	5.985E-02	1.034E-02	-3.19E-02	-9.96E-02
Q20	-0.128	9.905E-02	0.115	-0.110	-924E-02	5.985E-02	1.034E-02	1.034E-02	-3.31E-02	-0.177
Q21	-0.120	4.922E-02	7.125E-02	-6.98E-02	-7.76E-02	4.882E-02	-3.19E-02	-3.31E-02	-3.31E-02	-0.177
Q22	-7.90E-02	-0.102	-7.09E-02	-4.87E-02	-7.22E-02	-6.01E-02	-5.59E-02	-9.96E-02	-0.177	
Q23	-4.91E	-0.147	-8.01E-03	-7.56E-02	-6.96E-02	-7.31E-02	-5.59E-02	-9.96E-02	-0.177	

Método de Extração: Análise de Componentes Principais.
a Os resíduos são calculados entre as correlações observadas e reproduzidas. Existem 91 (35,0%) resíduos não-redundantes com valores absolutos > 0,05.
a Comunalidades reproduzidas.

que foi solicitada usando a opção da Figura 15.6. A parte mais alta dessa matriz (rotulada de **Reproduced Correlations** (Correlações Reproduzidas)) contém os coeficientes de correlações entre todas as perguntas baseadas no modelo de fatores. A diagonal dessa matriz contém as communalidades após a extração de cada variável (coloquei esses valores em negrito; você pode checar os valores com a Saída 15.5 do SPSS).

As correlações na matriz reproduzida diferem daquelas da matriz-R porque elas provêm do modelo em vez de dos dados observados. Se o modelo fosse um ajuste perfeito dos dados, esperaríamos que os coeficientes de correlação reproduzidos fossem os mesmos que os coeficientes de correlação originais. Portanto, para verificar o ajuste do modelo, podemos olhar as diferenças entre as correlações observadas e as correlações baseadas no modelo. **Por exemplo, se pegarmos a correlação entre as perguntas 1 e 2, a correlação baseada nos dados observados é de -0,099 (tirada da Saída 15.1 do SPSS). A correlação baseada no modelo é de -0,112, um pouco mais alta. Podemos calcular a diferença desta forma:**

$$\begin{aligned}\text{resíduo} &= r_{\text{observado}} - r_{\text{do modelo}} \\ \text{resíduo}_{Q1Q2} &= (-0,99) - (-0,112) \\ &= 0,013 \text{ ou } 1,3\text{E-}02\end{aligned}$$

Você deve notar que essa diferença é o valor destacado na metade inferior da matriz reproduzida (rotulada de **Residual (Resíduos)) para as perguntas 1 e 2. Portanto, a parte inferior da matriz reproduzida contém as diferenças entre as correlações dos coeficientes observados e aquelas previstas pelo modelo. Para um bom modelo, todos esses valores serão pequenos. Na verdade, queremos que a maioria dos valores seja menor do que 0,05. O SPSS fornece um resumo no rodapé que declara quantos resíduos tem um valor absoluto maior do que 0,05. Para esses dados, existem 91 resíduos (35%) maiores do que 0,05. Não existem regras rígidas sobre o que a proporção dos resíduos deve ser abaixo de 0,05; entretanto, se mais do que 50% são maiores do que 0,05, você terá um motivo de preocupação.**

15.6.3 Rotação de fatores ②

A primeira análise que pedi para você executar foi usando a rotação ortogonal. Entretanto, foi solicitado que você executasse a análise novamente usando a rotação oblíqua também. Nesta seção, os resultados de ambas as análises serão relatados para destacar as diferenças entre as saídas. Essa comparação também será útil para mostrar as circunstâncias em que um tipo de rotação pode ser preferível à outra.

15.6.3.1 Rotação ortogonal (Varimax) ②

A Saída 15.8 do SPSS mostra a matriz dos componentes após a rotação (também chamada de matriz dos fatores rotacionada na análise de fatores), que é a matriz da carga dos fatores para cada variável em cada fator. Essa matriz contém a mesma informação que a matriz dos componentes na Saída 15.5 do SPSS, exceto que ela é calculada *após* a rotação. Existem vários pontos a considerar sobre o formato dessa matriz. Primeiro, as cargas de fatores menores do que 0,4 não foram exibidas porque pedimos que as cargas fossem suprimidas usando a opção da Figura 15.10. Se você não selecionou essa opção ou não ajustou o valor critério para 0,4, sua saída será diferente. Segundo, as variáveis estão listadas por ordem de tamanho das suas cargas dos fatores. Por omissão, o SPSS ordena as variáveis como elas estão no editor de dados; entretanto, solicitamos que a saída fosse Ordenada por tamanho (**Sorted by size**) usando a opção da Figura 15.10. Se essa opção não foi selecionada, sua saída será diferente. Finalmente, para todas as outras partes da saída, suprimi os rótulos das variáveis (por motivo de espaço), mas para essa matriz permiti a impressão dos rótulos da matriz para auxiliar na interpretação.

A lógica original por trás da supressão de cargas menores do que 0,4 foi baseada na sugestão de Stevens (1992) de que esse ponto de corte era apropriado para facilitar a interpretação (isto é, cargas maiores do que 0,4 representam valores substanciais). Porém, isso significa que suprimimos muitas cargas que são, sem dúvida, significativas (veja a Seção 15.3.6.2). Entretanto, aqui a própria significância não é importante.

Saída 15.8 do SPSS

Rotated Component Matrix^a (Matriz dos Componentes Rotacionada)

	Component (Componente)			
	1	2	3	4
Eu tenho pouca experiência com computadores.	0.800			
O SPSS sempre trava quando eu tento usá-lo.	0.684			
Eu tenho medo de provocar danos irreparáveis por causa da minha incompetência com computadores.	0.647			
Todos os computadores me odeiam.	0.638			
Computadores têm suas próprias mentes e deliberadamente erram quando eu os uso.	0.579			
Os computadores são úteis somente para jogos.	0.550			
Computadores existem para me sacanear.	0.459			
Eu não consigo dormir quando penso em autovetores.		0.677		
Eu tenho pesadelos com a distribuição normal.		0.661		
Desvios padrão me excitam.		-0.567		
Dizem que o SPSS torna a estatística mais fácil de entender, mas isso não é verdade.	0.473	0.523		
Eu sonho com Pearson me atacando com coeficientes de correlação.		0.516		
Eu choro escancaradamente quando ouço falar em tendência central.		0.514		
A estatística me faz chorar.		0.496		
Eu não entendo estatística.		0.429		
Eu nunca fui bom em matemática.			0.833	
Eu entro em coma toda vez que vejo uma equação.			0.747	
Eu fui muito mal em matemática na escola.			0.747	
Meus amigos são melhores em estatística do que eu.				0.648
Meus amigos são melhores com o SPSS do que eu.				0.645
Eu sou bom em estatística e as pessoas acham que sou um CDF.				0.586
Meus amigos pensarão que sou estúpido por não saber mexer com o SPSS				0.543
Todos olham para mim quando uso o SPSS.				0.427

(Método de Extração: Análise dos Componentes Principais.
Método de Rotação: Varimax com normalização de Kayser.
a A rotação convergiu em 9 iterações.

**Component Transformation Matrix
(Matriz de Transformação dos Componentes)**

Component (Componente)	1	2	3	4
1	0.635	0.585	0.443	-0.242
2	0.137	-0.168	0.488	0.846
3	0.758	-0.513	-0.403	0.008
4	0.067	0.605	-0.635	0.476

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: Varimax com normalização de Kayser.

Compare essa matriz com a solução não-rotacionada (Saída 15.5 do SPSS). Antes da rotação, a maioria das variáveis estava altamente carregada no primeiro fator e os fatores restantes não foram examinados. Entretanto, a rotação da estrutura do fator deixou tudo mais claro: existem quatro fatores e as variáveis estão altamente carregadas em somente um fator

(com a exceção de uma pergunta). A supressão de cargas menores do que 0,4 e a ordenação das variáveis pelo tamanho da carga também torna a interpretação mais fácil (porque você não precisa examinar cuidadosamente a matriz para identificar cargas substanciais).

O próximo passo é olhar o conteúdo das perguntas que estão carregadas no mesmo fa-

tor para tentar identificar temas comuns. Se o fator matemático produzido pela análise representa um construto real, temas comuns entre perguntas com cargas altas podem nos ajudar a identificar o que o construto pode ser. As perguntas com cargas altas no fator 1 parecem ter relação com o uso de computadores e o SPSS. Portanto, podemos rotular esse fator de medo de computador (*fear of computers*). As perguntas com cargas altas no fator 2 parecem ter relação a diferentes aspectos da estatística, portanto, podemos rotular esse fator de medo da estatística (*fear of statistics*). As três perguntas com altas cargas no fator 3 parecem estar relacionadas à matemática, assim, podemos rotular esse fator de medo de matemática (*fear of mathematics*). Finalmente, as perguntas com cargas altas no fator 4 contêm algum componente de avaliação social de amigos, desse modo, podemos rotular esse fator de avaliação dos pares (*peer evaluation*). Essa análise revela que o questionário inicial é composto por quatro subescalas: medo de computadores, medo de estatística, medo de matemática e medo da avaliação negativa dos amigos. Existem duas possibilidades. A primeira é que o QAS falhou em medir o tema projetado, a ansiedade em relação ao SPSS, mas mede alguns construtos relacionados. A segunda é que esses quatro construtos são subcomponentes da ansiedade em relação ao SPSS; entretanto, a análise dos fatores não indica qual dessas possibilidades é verdadeira.

A parte final da saída é a matriz de transformação dos fatores (veja a Seção 15.3.6). Essa matriz fornece informação sobre o grau que os fatores foram rotacionados para obter a solução. Se não houve necessidade de rotação, essa matriz será uma matriz de identidade. Se a rotação ortogonal foi completamente apropriada, podemos esperar uma matriz simétrica (mesmos valores acima e abaixo da diagonal). Entretanto, na realidade, a matriz não é de fácil interpretação, embora matrizes não-simétricas possam ser tomadas como um motivo para tentar a rotação oblíqua. Talvez seja melhor ignorar a matriz de transformação dos fatores.

15.6.3.2 Rotação oblíqua ②

Quando uma rotação oblíqua é conduzida, a matriz dos fatores é dividida em duas matrizes: a matriz padrão (*pattern matrix*) e a matriz de estrutura (*structure matrix*) (veja o Quadro 15.1). Para a rotação ortogonal, essas matrizes são as mesmas. A matriz padrão contém as cargas dos fatores e é comparável à matriz dos fatores que interpretamos para a rotação ortogonal. A matriz estrutural leva em conta o relacionamento entre fatores (na verdade, ela é um produto da matriz padrão e da matriz contendo os coeficientes de correlação entre fatores). A maioria dos pesquisadores interpreta a matriz padrão porque ela é geralmente mais simples; entretanto, existem situações nas quais valores na matriz padrão são suprimidos por causa dos relacionamentos entre valores. Portanto, a matriz estrutural é útil para uma revisão e Graham e colaboradores. (2003) recomendam relatar ambas (com alguns exemplos úteis de por que isso pode ser importante).

Para a matriz padrão desses dados (Saída 15.9 do SPSS), os mesmos quatro fatores parecem ter emergido (embora para algumas variáveis as cargas dos fatores sejam muito pequenas para serem exibidas). O fator 1 parece representar o medo de estatística, o fator 2 representa o medo da avaliação dos seus pares, o fator 3 representa o medo de computadores e o fator 4 representa o medo de matemática. A matriz estrutural (Saída 15.10 do SPSS) difere, pois a variância compartilhada não é ignorada. O quadro se torna mais complicado porque, com exceção do fator 2, muitas variáveis têm altas cargas em mais do que um fator. Isso ocorreu devido ao relacionamento entre os fatores 1 e 3 e os fatores 3 e 4. Esse exemplo destaca por que a matriz padrão é preferível na interpretação: porque ela contém informação sobre a contribuição *única* de uma variável a um fator.

A parte final da saída é a matriz de correlações entre os fatores (Saída 15.11 do SPSS). Essa matriz contém os coeficientes de correlação entre os fatores. Como foi previsto da matriz estrutural, o fator 2 tem pouco ou nenhum relacionamento com qualquer outro fator (os coeficientes de correlação são baixos), mas to-

Saída 15.9 do SPSS

Pattern Matrix^a (Matriz Padrão)

	Component (Componente)			
	1	2	3	4
Eu não consigo dormir quando penso em autovetores.	0.706			
Eu tenho pesadelos com a distribuição normal.	0.591			
Desvios padrão me excitam.				
Eu sonho com Pearson me atacando com coeficientes de correlação.	-0.511			
Eu choro escancaradamente quando ouço falar em tendência central.	0.405			
A estatística me faz chorar.	0.400			
Eu não entendo estatística.		0.643		
Meus amigos são melhores com o SPSS do que eu.		0.621		
Meus amigos são melhores em estatística do que eu.		0.615		
Sou bom em estatística e as pessoas acham que sou CDF.		0.507		
Meus amigos irão pensar que eu sou estúpido por não saber mexer com o SPSS.			0.885	
Todos olham para mim quando uso o SPSS.			0.713	
Eu tenho pouca experiência com computadores.			0.653	
O SPSS sempre trava quando eu tento usá-lo.			0.650	
Todos os computadores me odeiam.			0.588	
Eu tenho medo de provocar danos irreparáveis por causa da minha incompetência com computadores.			0.585	
Computadores têm suas próprias mentes e deliberadamente erram quando eu os uso.	0.412		0.462	
Os computadores são úteis somente para jogos.			0.411	
Dizem que o SPSS torna a estatística mais fácil de entender, mas isso não é verdade.				
Computadores existem para me sacanear.				-0.902
Eu nunca fui bom em matemática.				-0.774
Eu entro em coma toda vez que vejo uma equação.				-0.774
Eu fui muito mal em matemática na escola.				

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: *Oblimin* com normalização de Kayser.
a A rotação convergiu em 29 iterações.

Dicas da Samantha Ferrinho: Interpretação



- Se você conduziu uma rotação ortogonal, olhe para a tabela Matriz dos Componentes Rotados (*Rotated Component Matrix*). Para cada variável, observe o componente para o qual a variável tem a carga maior. Também, para cada componente observe as variáveis que tem altas cargas nele (alta significa que as cargas devem estar acima de 0,4 quando você ignorar o sinal de mais ou menos). Tente entender o que os fatores representam observando os temas comuns nos itens que tem cargas.
- Se você conduziu a rotação oblíqua, olhe para a tabela Matriz Padrão (*Pattern Matrix*). Para cada variável, observe o componente para o qual a variável tem a carga mais alta. Também, para cada componente, observe as variáveis que tem a carga mais alta nele (alta significa que as cargas devem estar acima de 0,4 ignorando o sinal de mais ou menos). Reexamine o que você encontrar fazendo o mesmo para a Matriz de Estrutura (*Structure Matrix*). Tente entender o que os fatores representam procurando por temas comuns nos itens que tem cargas.

Saída 15.10 do SPSS

Structure Matrix (Matriz de Estrutura)

	Component (Componente)			
	1	2	3	4
Eu tenho pesadelos com a distribuição normal.	0.695		0.477	
Eu não consigo dormir quando penso em autovetores.	0.685			
Desvios padrão me excitam.	-0.632		-0.407	
Eu choro escancaradamente quando ouço falar em tendência central.	.567		0.516	-0.491
Sonho com Pearson me atacando com os coeficientes de correlação.	0.548		0.487	-0.485
A estatística me faz chorar.	0.520		0.413	-0.501
Eu não entendo estatística.	0.462		0.453	
Meus amigos são melhores com o SPSS do que eu.		0.660		
Meus amigos são melhores em estatística do que eu.		0.653		
Eu sou bom em estatística e as pessoas acham que sou um CDF.		0.588		
Meus amigos irão pensar que eu sou estúpido por não saber mexer com o SPSS		0.546		
Todos olham para mim quando uso o SPSS.	-0.435	0.446		
Eu tenho pouca experiência com computadores.			0.777	
O SPSS sempre trava quando eu tento usá-lo.	0.404		0.761	
Todos os computadores me odeiam.	0.401		0.723	
Eu tenho medo de provocar danos irreparáveis por causa da minha incompetência com computadores.			0.723	-0.429
Computadores têm suas próprias mentes e deliberadamente erram quando eu os uso.	0.426		0.671	
Dizem que o SPSS torna a estatística mais fácil de entender, mas isso não é verdade.	0.576		0.606	
Computadores existem para me sacanear.			0.561	-0.441
Os computadores são úteis somente para jogos.			0.556	
Eu nunca fui bom em matemática.				-0.855
Eu entro em coma toda vez que vejo uma equação.			0.453	-0.822
Eu fui muito mal em matemática na escola.			0.451	-0.818

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: *Oblimin* com normalização de Kayser.

Saída 15.11 do SPSS

Component Correlation Matrix (Matriz das Correlações dos Componentes)

Component (Componente)	1	2	3	4
1	1.000	-0.154	0.364	-0.279
2	-0.154	1.000	-0.185	8.155E-02
3	0.364	-0.185	1.000	-0.464
4	-0.279	8.155E-02	-0.464	1.000

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: *Oblimin* com normalização de Kayser.

dos os outros fatores estão inter-relacionados a um determinado grau (notavelmente os fatores 1 e 3 e os fatores 3 e 4). O fato de que essas correlações existem indica que os construtos medidos podem estar inter-relacionados. Se os construtos fossem independentes, esperaríamos que a rotação oblíqua fornecesse uma

solução idêntica para uma rotação ortogonal e a matriz de correlação do componente deveria ser uma matriz de identidade (isto é, todos os fatores têm coeficientes de correlação igual a 0). Portanto, essa matriz final indica se é razoável assumir a independência entre os fatores: para esses dados parece que não podemos assumir independência. Portanto, os resultados da rotação ortogonal podem não ser confiáveis: a solução da rotação oblíqua é, provavelmente, mais apropriada.

Num nível teórico, a dependência entre nossos fatores não causa preocupação; podemos esperar um relacionamento bem forte entre medo de matemática, medo de estatística e medo de computadores. Geralmente pessoas com menos inclinação à matemática lutam com a estatística. Entretanto, não esperamos que esses construtos se correlacionem com o medo de avaliação dos

seus pares (porque esses construtos têm base mais social). De fato, esse fator é o que menos se correlaciona com os outros – assim, num nível teórico, as coisas acabaram muito bem!

15.6.4 Escores do fator ②

Depois de encontrar uma solução adequada e fazer a rotação daquela solução, podemos olhar os escores do fator. A Saída 15.12 do SPSS mostra a matriz dos escores do componente *B* (veja a Seção 15.2.4), da qual os escores do fator são calculados e a matriz das covariâncias dos escores do fator. A matriz dos escores do componente não é particularmente útil. Ela pode ser útil na compreensão de como os escores do fator foram calculados, mas com grandes conjuntos de dados como esse, não é provável que você queira pesquisar a matemática por trás dos escores do fator. Entretanto, a matriz das covariâncias dos escores é útil. Essa matriz nos mostra os relacionamentos entre os escores do fator (ela é uma matriz de correlação não-padronizada). Se os escores do fator não são correlacionados, essa matriz deve ser uma matriz de identidade (isto é, elementos diagonais serão 1, mas todos os outros elementos são 0). Para esses dados, todas as covariâncias são zero (lembrando que 4,37E-16 é realmente 0,000000000000000437), indicando que os escores resultantes não são correlacionados.

Na análise original, pedimos que os escores fossem calculados baseados no método de Anderson-Rubin (por isso eles não são correlacionados). Você encontrará esses escores no editor de dados. Deve haver quatro novas colunas de dados (uma para cada fator) rotuladas de FAC1_1, FAC2_1, FAC3_1 e FAC4_1, respectivamente. Se você pedir por escores dos fatores na rotação oblíqua, esses escores aparecerão no editor de dados em outras quatro colunas rotuladas de FAC2_1 e assim por diante. Esses escores do fator podem ser listados no visualizador de saída usando o comando **Analyze⇒Reports⇒Case Summaries...** (Analisar⇒Relatar⇒Resumir Casos...) (veja a Seção 5.8.6). Como existem mais do que 1500 casos, você pode querer restringir a saída dos primeiros 10 ou 20. A

Saída 15.12 do SPSS

Component Score Coefficient Matrix (Matriz dos Coeficientes dos Escores dos Componentes)

	Component (Componente)			
	1	2	3	4
Q01	-0.053	0.173	0.089	0.110
Q02	0.102	-0.129	0.086	0.281
Q03	0.087	-0.195	0.013	0.137
Q04	-0.011	0.170	0.045	0.107
Q05	0.021	0.131	0.014	0.083
Q06	0.383	-0.211	-0.088	0.014
Q07	0.213	0.004	-0.078	0.038
Q08	-0.129	-0.074	0.460	0.013
Q09	0.025	-0.029	0.108	0.354
Q10	0.244	-0.161	-0.021	-0.036
Q11	-0.066	-0.087	0.379	-0.059
Q12	0.097	0.161	-0.116	0.051
Q13	0.224	-0.065	-0.019	0.013
Q14	0.180	0.040	-0.084	0.043
Q15	0.114	-0.055	0.061	-0.058
Q16	-0.015	0.146	0.046	0.014
Q17	-0.057	-0.067	0.372	0.005
Q18	0.242	-0.001	-0.104	0.043
Q19	0.048	-0.115	0.061	0.199
Q20	-0.195	0.359	-0.061	-0.002
Q21	-0.039	0.270	-0.064	0.059
Q22	-0.036	0.162	-0.048	0.382
Q23	0.032	0.211	-0.162	0.379

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: *Varimax* com normalização de Kayser.

Component Score Covariance Matrix (Matriz das Covariâncias dos Escores dos Componentes)

Component (Componente)	1	2	3	4
1	1.000	1.093E-16	0.000	0.000
2	1.093E-16	1.000	4.373E-16	0.000
3	0.000	4.373E-16	1.000	0.000
4	0.000	0.000	0.000	1.000

Método de Extração: Análise dos Componentes Principais.
Método de Rotação: *Varimax* com normalização de Kayser.

Saída 15.13 do SPSS mostra os fatores do escore para os primeiros 10 participantes. Deve estar bem claro que o participante 9 teve um escore alto em todos os quatro fatores e, assim, essa pessoa tem muita ansiedade quanto a estatística, informática e matemática, mas menos quanto a avaliação dos seus pares (fator 4). Os escores do fator podem ser usados dessa maneira para avaliar o medo relativo de

Saída 15.13 do SPSS

Case Summaries^a (Resumo de casos)

	Case Number (Caso Número)	FAC1_1	FAC2_2	FAC3_3	FAC4_4
1	1	0.10584	-0.92797	-1.82768	-0.45958
2	2	-0.58279	-0.18934	-0.04137	0.29330
3	3	-0.54761	0.02968	0.19913	-0.97142
4	4	0.74664	0.72272	-0.68650	-0.18533
5	5	0.25167	-0.51497	-0.63353	0.68296
6	6	1.91613	-0.27351	-0.68205	-0.52710
7	7	-0.26055	-1.40960	-0.00435	0.90986
8	8	-0.28574	-0.91775	-0.08948	1.03732
9	9	1.71753	1.15063	3.15671	0.81083
10	10	-0.69153	-0.73340	0.18379	1.49529

a Limitado aos primeiros 10 casos.

uma pessoa comparado com outra, ou podemos adicionar os escores para obter um único escore para cada participante (que podemos assumir que representa a ansiedade do SPSS como um todo). Podemos, também, usar os escores do fator em regressão quando grupos de previsores se correlacionam de tal forma que há multicolinearidade.

15.6.5 Resumo ②

A análise revela quatro escalas subjacentes em nosso questionário que podem ou não se relacionar a subcomponentes genuínos da ansiedade em relação ao SPSS. Uma solução utilizando uma rotação oblíqua foi preferida devido às inter-relações entre os fatores. O uso da análise dos fatores é exploratório; ela deve ser usada somente para guiar futuras hipóteses ou informar aos pesquisadores sobre os padrões dentro dos conjuntos de dados. Muitas decisões são deixadas aos pesquisadores usando a análise dos fatores e eu recomendo que você tome decisões bem-informadas em vez de basear suas decisões nas saídas que você gostaria de conseguir. A próxima pergunta é se a escala é ou não confiável.

15.7 ANÁLISE DE CONFIABILIDADE ②

15.7.1 Medidas de confiabilidade ③

Se você estiver usando a análise de fatores para validar um questionário, é aconselhável checar a confiabilidade da escala.

Confiabilidade significa apenas que a escala deve, consistentemente, refletir o construto que está medindo. Uma maneira de definir isso é que, mantendo todo o resto igual, uma pessoa

deve ter um mesmo escore num questionário se ela completá-lo em dois pontos diferentes no tempo (isso é chamado de *confiabilidade de teste-reteste*). Assim, alguém que tenha pavor de estatística e que teve altos escores no nosso QAS deve ter escores similarmente altos se o testarmos um mês depois (assumindo que ele não tenha feito algum tipo de terapia para tratar a ansiedade em relação à estatística no último mês). Outra maneira de analisar a confiabilidade é afirmar que duas pessoas sendo medidas, que são as mesmas em termos de construto, devem ter o mesmo escore. Assim, se pegarmos duas pessoas que tinham fobia a estatística, elas devem ter, mais ou menos, escores idênticos no QAS. Da mesma forma, se pegarmos duas pessoas que adoram estatística, ambas devem ter escores igualmente baixos. Portanto, se pegarmos alguém que adora estatística e alguém que odeia e eles alcançarem o mesmo escore no nosso questionário, não seria uma medida precisa da ansiedade em relação à estatística! Em termos estatísticos, a maneira padrão para analisar a confiabilidade é baseada na ideia de que itens individuais (ou um conjunto de itens) devem produzir resultados

Como sei se o meu questionário é confiável?



consistentes em todo o questionário. Assim, se tomarmos alguém com medo de estatística, seu escore geral no QAS será alto; se o QAS é confiável e se nós aleatoriamente selecionarmos alguns itens do QAS, o escore da pessoa naqueles itens também deverá ser alto.

A maneira mais simples de fazer isso na prática é usar a confiabilidade meio-a-meio (*split-half reliability*). Esse método divide, aleatoriamente, o conjunto de dados em dois. Um escore para cada participante é calculado em cada metade da escala. Se a escala for confiável, o escore da pessoa numa metade da escala deve ser o mesmo (ou similar) ao seu escore na outra metade: portanto, através de vários participantes os escores das duas metades do questionário devem ter uma correlação bastante alta. A correlação entre as duas metades é a estatística calculada nesse método, com grandes correlações indicando confiabilidade. O problema desse método é que existem várias maneiras de dividir um conjunto de dados em dois e, assim, os resultados podem resultar da forma como os dados são divididos. Para solucionar esse problema, Cronbach (1951) apresentou uma medida vagamente equivalente à separação dos dados em dois de todas as maneiras possíveis e com coeficiente de correlação calculado para cada parte. A média desses valores é o alfa de Cronbach (*Cronbach alpha*), a medida mais comum de confiabilidade.⁹

⁹ Embora essa seja a maneira mais fácil de conceitualizar o alfa de Cronbach, α , ser ou não exatamente igual à média de todas as possíveis confiabilidades meio-a-meio depende exatamente de como você calculou a confiabilidade meio-a-meio (veja o glossário para detalhes computacionais). Se você utilizar a fórmula de Spearman-Brown, a qual não leva em consideração os desvios padrão dos itens, o α de Cronbach será igual à média da confiabilidade meio-a-meio somente quando os desvios padrão dos itens são iguais; caso contrário, α , será menor do que a média. Entretanto, se você usar uma fórmula para a confiabilidade meio-a-meio que leva em conta os desvios padrão para os itens (tal como Flanagan, 1937; Rulon, 1939), α será sempre igual à média da confiabilidade meio-a-meio (veja Cortina, 1993).

O α de Cronbach é dado por:

$$\alpha = \frac{N^2 \overline{\text{Cov}}}{\sum s_{\text{item}}^2 + \sum \text{Cov}_{\text{item}}} \quad (15.6)$$

o qual pode parecer complicado, mas na verdade não é. Observe que para cada item na nossa escala podemos calcular duas coisas: a variância dentro do item e a covariância entre um item particular e qualquer outro item na escala. Colocando de outra forma, podemos construir uma matriz de variâncias-covariâncias de todos os itens. Nessa matriz, os elementos diagonais serão as variâncias dentro de um item particular e os elementos fora da diagonal serão covariâncias entre pares de itens. O numerador da equação é simplesmente o número dos itens (N) ao quadrado multiplicado pela média da covariância entre itens (a média dos elementos fora da diagonal na matriz de variâncias-covariâncias acima mencionada). O denominador é apenas a soma de todas as variâncias e covariâncias do item (isto é, a soma de tudo na matriz de variâncias-covariâncias).

Também existe uma versão padronizada do coeficiente, a qual essencialmente usa a mesma equação, exceto que as correlações são usadas em vez das covariâncias, e no denominador da equação usamos a soma dos elementos da matriz de correlações dos itens (incluindo aqueles que aparecem na diagonal daquela matriz). O α normal é apropriado quando os itens numa escala são somados para produzir um escore simples para aquela escala (o α padronizado não é apropriado nesses casos). O α padronizado é útil quando os itens de uma escala são padronizados antes de serem somados.

15.7.2 Interpretando o α de Cronbach (advertências) ②

Geralmente afirma-se que um valor de 0,7-0,8 é aceitável para o α de Cronbach e valores substancialmente mais baixos indicam uma escala não confiável. Kline (1999) registra que embora o valor comumente aceito de 0,8 seja apropriado para testes cognitivos como o teste

de inteligência, para testes de habilidade um ponto de corte de 0,7 é mais adequado. Ele também afirma que quando se tratar de construtos psicológicos, valores abaixo de 0,7 podem ser esperados, por causa da diversidade dos construtos que estão sendo medidos.

Entretanto, Cortina (1993) registra que tais procedimentos gerais precisam ser usados com cautela porque o valor de α depende do número de itens na escala. Você pode verificar que o numerador da equação para o α inclui o número de itens elevado ao quadrado. Portanto, à medida que o número de itens da escala aumenta, o valor de α aumentará também. Assim, é possível conseguir um valor alto para o α porque você tem muitos itens na escala e não porque ela seja confiável! Por exemplo, Cortina registra dados das duas escalas, ambos tendo $\alpha = 0,8$. A primeira escala tem somente três itens e a média da correlação entre os itens foi um respeitável 0,57; entretanto, a segunda escala tinha 10 itens com uma correlação média entre esses itens de um menos respeitável 0,28. Claramente, a consistência interna dessas escalas difere muito, mas de acordo com o α de Cronbach ambas são confiáveis!

A segunda interpretação comum do α é que ele mede a “unidimensionalidade” ou a extensão em que a escala mede um fator subjacente ou construto. Essa interpretação se origina pelo fato de que quando existe um fator subjacente aos dados, o α é uma medida da força daquele fator (veja Cortina, 1993). Entretanto, Grayson (2004) demonstra que conjuntos de dados com o mesmo α podem ter estruturas bem diferentes. Ele mostrou que um α de 0,8 pode ser alcançado numa escala com um fator subjacente, com dois fatores moderadamente correlacionados e com dois fatores não-correlacionados. Cortina (1993) também mostrou que com mais de 12 itens e correlações bem altas entre itens ($r > 0,5$), α pode alcançar valores por volta ou acima de 0,7 (0,65 a 0,84). Esses resultados mostram que o α não deve ser usado como uma medida da “unidimensionalidade”. Na verdade, Cronbach (1951) sugeriu que se vários fatores existem, a fórmula deve ser aplicada separa-

damente a itens relacionados a diferentes fatores. Em outras palavras, se o seu questionário tem subescalas, o α deve ser aplicado separadamente a essas subescalas.

Um aviso final é sobre itens que tem frases invertidas. Por exemplo, no QAS que usamos na parte da análise de fatores neste capítulo, tínhamos um item (pergunta 3) que foi formulada de maneira contrária a todos os demais itens. O item era “os desvios padrão me excitam”. Compare-o com qualquer outro item e você verá que

ele requer uma resposta contrária. Por exemplo, o item 1 é “a estatística me faz chorar”. Se você não gosta de estatística, você concorda plenamente com essa

afirmação e, assim, terá um escore de 5 na nossa escala. Para o item 3, se você odeia estatística, desvios padrão provavelmente não o excitam, assim, você irá discordar totalmente e terá um escore de 1 na escala. Esses itens de frases invertidas são importantes para reduzir respostas tendenciosas; os participantes precisarão ler os itens com atenção. Para a análise dos fatores, essas frases invertidas não importam, o que acontece é que você consegue uma carga do fator negativa por qualquer item reverso (de fato, olhe a Saída 15.8 do SPSS e você irá notar que o item 3 tem a carga do fator negativa). Entretanto, em análise de confiabilidade esses itens de escores reversos fazem realmente a diferença. Para entender o porquê, pense na equação do α de Cronbach. Nessa equação, o numerador incorpora a covariância média entre os itens. Se um item é uma frase invertida, ele terá um relacionamento negativo com os demais; portanto, as covariâncias entre esse item e os outros itens serão negativas. A covariância média é, obviamente, a soma das covariâncias dividida pelo número de covariâncias, e incluindo um grupo de valores negativos, reduzimos a soma das covariâncias e, portanto, também reduzimos o valor do α de Cronbach, porque o numerador da equação fica menor. Em casos extremos, é até

Meu alfa deu negativo! O que eu faço?



mesmo possível conseguir um valor negativo para o α de Cronbach simplesmente porque a magnitude das covariâncias negativas é maior do que a magnitude das positivas! Um α de Cronbach negativo não faz muito sentido, mas realmente acontece, e se acontecer, pergunte a você mesmo se você não incluiu alguns itens com frases invertidas!

Se você tiver itens com frases invertidas, deve, também, reverter a maneira de obter os escores antes de realizar a análise de confiabilidade. Isso é muito fácil. Veja os dados do seu QAS, onde temos um item que está com o escore de 1 = discorda totalmente, 2 = discorda, 3 = nem concorda, nem discorda, 4 = concorda e 5 = concorda totalmente. Isso é bom para itens em que a concordância indica ansiedade com a estatística, mas para o item 3 (desvios padrão me excitam), a discordância indica ansiedade em relação à estatística. Para refletir isso numericamente, precisamos reverter a escala tal que 1 = concorda totalmente, 2 = concorda, 3 = nem concorda, nem discorda, 4 = discorda e 5 = discorda totalmente. Dessa maneira, uma pessoa ansiosa ainda tem 5 nesse item (porque

ela discordaria totalmente), ou seja, nesse caso é preciso inverter a escala utilizada.

Para fazer isso, encontre o valor máximo da sua escala de resposta (nesse caso, 5) e adicione 1 a ela (assim, você tem 6 nesse caso). Para cada pessoa, pegue esse valor e subtraia dele o escore que a pessoa realmente obteve. Assim, alguém com um escore 5 originalmente terá, agora, $6 - 5 = 1$ e alguém com escore 1 original terá $6 - 1 = 5$. Alguém no meio da escala com um escore de 3 terá $6 - 3 = 3$! Obviamente, levaria muito tempo para fazer isso para cada pessoa, mas podemos usar o comando **Transform⇒Compute...** (Transformar⇒Calcular...) do SPSS para fazer isso automaticamente (veja a Seção 3.3.4). Vimos esse comando no Capítulo 3, e devemos entrar com o nome da variável que queremos trocar no espaço **Target Variable** (Variável Alvo) (veja a Figura 15.11) (nesse caso, a variável é chamada de **q03**). Você pode usar um nome diferente, se quiser, mas o SPSS criará uma nova variável e você precisará lembrar ela na análise de confiabilidade. Portanto, em **Numerical Expression** (Expressão Numérica), você precisa dizer ao SPSS como calcular a nova variável.

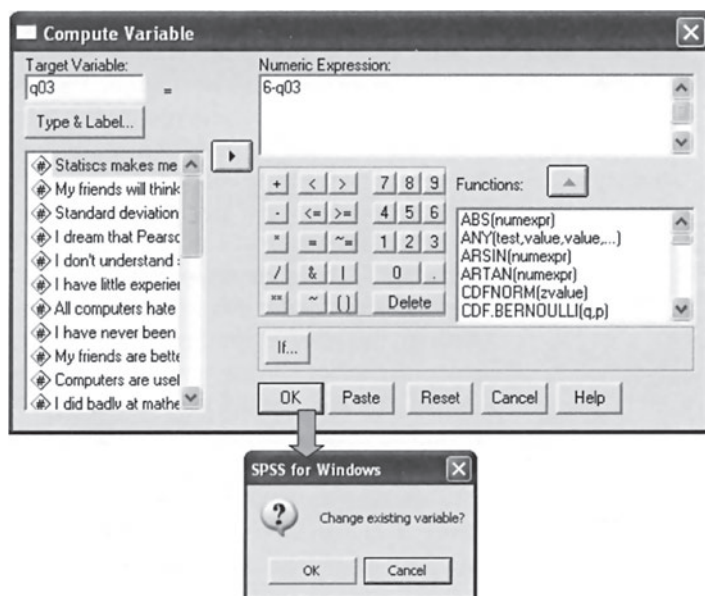


Figura 15.11 Escore reverso do item 3 (q03) no SPSS.

Nesse caso, queremos pegar o escore original de cada pessoa no item 3 e subtrair de 6. Portanto, simplesmente digitamos $6 - q03$ (o que significa 6 menos o valor encontrado na coluna rotulada de $q03$). Se você usou o mesmo nome, quando clicar em **OK** obterá uma caixa de diálogo pedindo se você quer trocar a variável existente; apenas clique em **OK** se quiser que os novos valores sejam trocados pelos antigos. A Figura 15.11 mostra esse processo.

15.7.3 Análise de confiabilidade no SPSS ②

Vamos testar a confiabilidade do QAS usando os dados que estão em **SAQ.sav**. Agora, você deve ter os escores invertidos no item 3 (veja acima), mas se você não quer ser incomodado, carregue o arquivo **SAQ(Item 3 Reversed).sav** em vez do arquivo **SAQ.sav**. Lembre, também, que eu disse que você deveria realizar a análise de confiabilidade nas subescalas individualmente. Se usarmos os resultados da rotação ortogonal (veja a Saída 15.8 do SPSS), temos quatro subescalas:

1. Subescala 1 (Medo de computadores): itens 6, 7, 10, 13, 14, 15, 18
2. Subescala 2 (Medo de estatística): itens 1, 3, 4, 12, 14, 10, 21
3. Subescala 3 (Medo de matemática): itens 8, 11, 17
4. Subescala 4 (Medo da avaliação dos pares): itens 2, 9, 19, 22, 23

Para realizar cada análise de confiabilidade nesses dados, você precisa seguir o caminho do menu **Analyze⇒Scale⇒Reliability Analysis...** (Analisar⇒Escala⇒Análise de Confiabilidade...) para exibir a caixa de diálogo semelhante à da Figura 15.12. Selecione qualquer item da lista que você quer analisar (para começar, vamos analisar os itens da subescala medo de computadores) no lado esquerdo da caixa de diálogo e transfira-os para a caixa **Items** (Itens) clicando em **▶**.

Selecionando o item de marcar **List item labels** (Listar os rótulos dos itens), você obterá uma listagem dos rótulos de todas as variáveis (o que pode ser útil para checar com quais itens suas variáveis se relacionam). Existem várias análises de confiabilidade que você pode executar, mas a que será executada por omissão é o α de Cronbach. Você pode trocar essa análise (para, por exemplo, a do método meio-a-meio) usando o quadro **Model** (Modelo), mas um bom método a ser utilizado é o método por omissão.

Se você clicar em **Statistics...** (Estatísticas), irá acessar uma caixa de diálogo como a da Figura 15.13. Na caixa de diálogo **Statistics** (Estatísticas), você pode selecionar vários fatores, mas o mais importante para a confiabilidade do questionário é: **Scale if item deleted** (Escala se o item for deletado). **Essa opção fornece um valor do α de Cronbach para cada item na sua escala. Ele nos diz qual seria o valor de α se aquele item fosse deletado. Se o seu questionário**

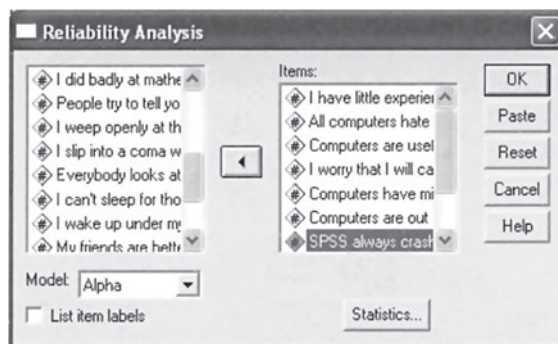


Figura 15.12 Caixa de diálogo principal para a análise de confiabilidade.

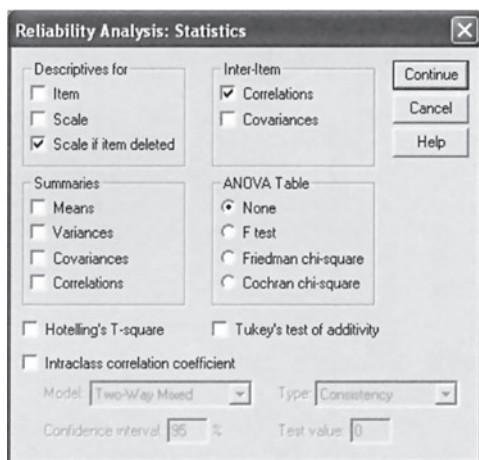


Figura 15.13 Estatísticas para a análise de confiabilidade.

nário é confiável, não vamos esperar que algum item afete grandemente a confiabilidade. Em outras palavras, nenhum item deve causar um substancial decréscimo em α . Se ele causar, isso é preocupante e você deve considerar tirar aquele item do questionário. Como 0,8 é visto como um bom valor de α , esperamos que todos os valores de **alpha if item deleted** (alfa se o item for deletado) estejam por volta de 0,8 ou acima.

As intercorrelações e covariâncias entre itens nos fornecem coeficientes de correlações e médias para os itens na nossa escala. Já vimos esses valores na nossa análise de fatores, portanto, não vale a pena selecionar essas opções. Entretanto, se você ainda não fez a análise dos fatores, pode ser útil pedir as correlações interitens porque o α global é afetado pelo número de itens sendo analisado e, assim, você deveria reexaminar para ver se os itens têm uma boa inter-relação. Opções como a Tabela da ANOVA irão simplesmente comparar tendências centrais de itens diferentes do questionário usando um teste F. Ela aplica uma ANOVA de medidas repetidas de um fator nos itens do questionário, o qui-quadrado de Friedman (se os seus dados estão ordenados) ou o qui-quadrado de Cochran (se os seus dados são dicotômicos, por exemplo, se os itens do questionário são de respostas sim/não).

O T quadrado de Hotelling faz praticamente o mesmo, mas produz uma estatística multivariada equivalente ao teste F. Esses testes são úteis se você quiser verificar se os itens têm distribuições similares (isto é, o mesmo valor médio), mas dado o alto valor do tamanho das amostras que você deve estar usando, eles irão, inevitavelmente, produzir resultados significativos mesmo quando somente existam pequenas diferenças entre as médias dos itens do questionário.

Use o conjunto de opções como o da Figura 15.13 para executar uma análise de confiabilidade básica. Clique em **Continue** para retornar à caixa de diálogo principal e clique em **OK** para executar a análise.

15.7.4 Interpretando a saída ②

A Saída 15.14 do SPSS mostra os resultados dessa análise de confiabilidade básica para a subescala medo de computadores. Os valores na coluna **Corrected Item-Total Correlation** (Correlação total dos itens corrigidos) são as correlações entre cada item e o escore total do questionário. Numa escala confiável, todos os itens devem se correlacionar com o total. Assim, estamos procurando por itens que não se correlacionam com o escore total da escala; se quaisquer desses valores são menores do que aproximadamente 0,3 (depende levemente do tamanho da amostra – com amostras maiores, coeficientes de correlações menores são aceitáveis), estamos com problemas porque isso significa que um item específico não se correlaciona muito bem com toda a escala. Itens com baixa correlação talvez tenham que ser descartados. Para esses dados, todos os itens apresentam boas correlações com o total de itens, o que é animador.

Os valores na coluna **Alpha if Item Deleted** (Alfa se item eliminado) são os valores do α global se aquele item não está incluído nos cálculos. Assim, eles refletem a mudança no α de Cronbach que seria observada se um item particular fosse retirado. O α global é 0,823 e, assim, todos os valores nessa coluna devem estar por volta desse mesmo valor. O que realmente estamos procurando são valores do α

Saída 15.14 do SPSS

RELIABILITY ANALYSIS - SCALE (ALPHA) (Análise de Confiabilidade - Escala (Alfa))							
Correlation Matrix (Matriz de Correlação)							
	Q06	Q07	Q10	Q13	Q14	Q15	Q18
Q06	1.0000						
Q07	0.5136	1.0000					
Q10	0.3222	0.2837	1.0000				
Q13	0.4664	0.4421	0.3020	1.0000			
Q14	0.4022	0.4407	0.2547	0.4498	1.0000		
Q15	0.3599	0.3914	0.2952	0.3422	0.3801	1.0000	
Q18	0.5133	0.5009	0.2925	0.5329	0.4983	0.3429	1.0000
N of Cases = 2571.0 (Número de casos = 2571,0)							
Item-Total Statistics (Estatísticas Item-Total)							
	Scale Mean If Item Deleted (Média da escala se o item for retirado)	Scale Variance if Item Deleted (Variância da escala se o item for retirado)	Corrected Item-total Correlation (Correlação Item-total corrigida)	Squared Multiple Correlation (Correlação múltipla ao quadrado)	Alpha if Item Deleted (Alfa se o item for retirado)		
Q06	15.8650	17.6141	0.6187	0.3981	0.7906		
Q07	15.1684	17.7370	0.6190	.3949	0.7905		
Q10	15.8114	20.7360	0.3999	0.1665	0.8239		
Q13	15.6429	18.8086	0.6067	0.3844	0.7973		
Q14	15.2159	18.7188	0.5768	0.3504	0.7980		
Q15	15.3259	19.3217	0.4913	0.2497	0.8119		
Q18	15.5235	17.8324	0.6474	0.4475	0.7855		
Reliability Coefficients (Coeficientes de confiabilidade)		7 items (7 itens)					
Alpha = .8324 (Alfa = 0.8324)		Standardized item alpha = .8214 (Alfa Padronizado = 0.8214)					

maiores do que o α global. Se a eliminação de um item aumenta o α de Cronbach, isso significa que a eliminação daquele item aumenta a confiabilidade. Portanto, qualquer item que resulte em valores substancialmente maiores do α do que o α global talvez precisem ser eliminados da escala para aumentar sua confiabilidade. Nenhum dos itens aqui irá afetar substancialmente a confiabilidade se forem retirados. O pior transgressor é a pergunta 10: eliminando essa pergunta, o α irá aumentar de 0,823 para 0,824. Contudo, esse aumento não é significativo e os dois valores refletem um grau razoável de confiabilidade.

Finalmente, e talvez mais importante, o valor de **Alfa** na parte bem abaixo é o α de Cronbach: a confiabilidade geral da escala. Para reiterar, estamos procurando por valores na magnitude de 0,7 para 0,8 (ou em torno desses valores), tendo em mente que acabamos de observar sobre os efeitos do número de itens. Nesse caso, o α está levemente acima de 0,8 e está na região indicada por Kline, assim, provavelmente ele indica confiabilidade. **Também é válido observar que se um item realmente precisa ser retirado nesse estágio, você deve executar novamente sua análise dos fatores para ter certeza de que a retirada do item não afetou a estrutura do fator!**

Vamos continuar com a subescala medo de estatística (itens 1, 3, 4, 5, 12, 16, 20 e 21). Não voltarei ao assunto do SPSS novamente, mas a Saída 15.15 do SPSS mostra os resultados da análise (para economizar espaço, omiti as correlações interitens). Os valores na coluna Correlação Total Corrigida dos Itens estão, novamente, todos acima de 0,3, o que é bom, e os valores na coluna Alfa se o item for deletado indica que nenhum dos itens aqui aumentaria a confiabilidade se fossem eliminados porque todos os valores nessa coluna são menores do que a confiabilidade geral de 0,819. Assim como as duas subescalas anteriores, o α global está em volta de 0,8, o que indica boa confiabilidade.

Para entender a importância dos escores reversos dos itens antes de executar a análise de confiabilidade, observe que a Saída 15.16 do SPSS mostra a análise de confiabilidade para a subescala medo da estatística, mas feita com os dados originais (isto é, sem reverter o item 3). Note que o α global é consideravelmente menor (0,606 em vez de 0,821). Observe, também, que esse item

tem uma correlação item-total negativa (que é uma boa maneira de mostrar que temos um potencial item reverso nos dados). Finalmente, note que para o item 3, o **Alpha if Item Deleted** (alfa se o item for deletado) é aproximadamente 0,80. Quer dizer, se o item for eliminado, a confiabilidade irá aumentar de 0,6 para aproximadamente 0,8! Isso, eu espero, ilustra que é preciso tomar cuidado com os itens que possam estar com as escalas invertidas, pois eles podem prejudicar bastante a confiabilidade de uma escala!

Indo agora para a subescala medo da matemática (itens 8, 11 e 17), a Saída 15.17 do SPSS mostra a saída dessa análise. Os valores na coluna **Corrected Item-Total Correlation** (Correlação Item-total Corrigida) estão todos acima de 0,3, o que é bom, e os valores na coluna **Alpha if Item Deleted** (alfa se o item for deletado) indicam que nenhum dos itens aumentará a confiabilidade caso sejam retirados porque todos os valores dessa coluna são menores que o a confiabilidade global, 0,819. Assim como nas duas subescalas anteriores,

Saída 15.15 do SPSS

RELIABILITY ANALYSIS - SCALE (ALPHA) (Análise de Confiabilidade - Escala (Alfa))					
Item-Total Statistics (Estatísticas Item-Total)					
	Scale Mean If Item Deleted (Média da escala se o item for retirado)	Scale Variance if Item Deleted (Variância da escala se o item for retirado)	Corrected Item-total Correlation (Correlação Item-total corrigida)	Squared Multiple Correlation (Correlação múltipla ao quadrado)	Alpha if Item Deleted (Alfa se o item for retirado)
Q01	21.7569	21.4417	0.5361	0.3435	0.8017
Q03	20.7165	19.8250	0.5492	3093	0.7996
Q04	21.3450	20.4105	0.5750	0.3553	0.7955
Q05	21.4088	20.9422	0.4944	0.2724	0.0866
Q12	20.9716	20.6393	0.5715	0.3370	0.7962
Q16	21.2517	20.4507	0.5973	0.3886	0.7928
Q20	20.5068	21.1761	0.4185	0.2440	0.8185
Q21	20.9603	19.9385	0.6061	0.3988	0.7908
Reliability 8 items (8 itens)					
Coefficients					
(Coeficientes de					
confiabilidade)					
Alpha = .8208		Standardized item alpha = .8234			
(Alfa = 0,8208)		(Alfa Padronizado = 0,8234)			

Saída 15.16 do SPSS

RELIABILITY ANALYSIS - SCALE (ALPHA) (Análise de Confiabilidade - Escala (Alfa))					
Item-Total Statistics (Estatísticas Item-Total)					
	Scale Mean If Item Deleted (Média da escala se o item for retirado)	Scale Variance if Item Deleted (Variância da escala se o item for retirado)	Corrected Item-total Correlation (Correlação Item-total corrigida)	Squared Multiple Correlation (Correlação múltipla ao quadrado)	Alpha if Item Deleted (Alfa se o item for retirado)
Q01	20.9277	12.1247	0.5051	0.3435	0.5213
Q03	20.7165	19.8250	-0.5492	0.3093	0.7996
Q04	20.5158	11.4467	0.5260	0.3553	0.5048
Q05	20.5795	11.7138	0.4662	0.2724	0.5230
Q12	20.1424	11.7393	0.5006	0.3370	0.5154
Q16	20.4224	11.5842	0.5291	0.3886	0.5066
Q20	19.6776	12.1073	0.3528	0.2440	0.5576
Q21	20.1311	11.1894	0.5410	0.3988	0.4969
-					
Reliability Coefficients (Coeficientes de confiabilidade)	8 items (8 itens)				
Alpha = .6055 (Alfa = 0.6055)	Standardized item alpha = .6413 (Alfa Padronizado = 0.6413)				

Saída 15.17 do SPSS

RELIABILITY ANALYSIS - SCALE (ALPHA) (Análise de Confiabilidade - Escala (Alfa))					
Item-Total Statistics (Estatísticas Item-Total)					
	Scale Mean If Item Deleted (Média da escala se o item for retirado)	Scale Variance if Item Deleted (Variância da escala se o item for retirado)	Corrected Item-total Correlation (Correlação Item-total corrigida)	Squared Multiple Correlation (Correlação múltipla ao quadrado)	Alpha if Item Deleted (Alfa se o item for retirado)
Q08	4.7219	2.4701	0.6845	0.4704	0.7396
Q11	4.7036	2.4530	0.6818	0.4672	0.7422
Q17	4.4920	2.5037	0.6520	0.4251	0.7725
Reliability Coefficients (Coeficientes de confiabilidade)	3 items (3 itens)				
Alpha = .8194 (Alfa = 0,8194)	Standardized item alpha = .8195 (Alfa Padronizado = 0,8195)				

o α global está em torno de 0,80, o que indica uma confiabilidade boa.

Finalmente, se você executar a análise para a última subescala, a avaliação pelos pares, você obterá a Saída 15.18 do SPSS. Os valores na coluna **Corrected Item-total Correlation** (Correlação Item-Total Corrigida) estão todos em volta de 0,3, exceto o item 23, que está abaixo. Isso indica uma consistência interna ruim e identifica o item 23 como o problema potencial. Os valores na coluna **Alpha if Item Deleted** (Alfa se item for deletado) indicam que nenhum dos itens da coluna aumentará a confiabilidade se forem eliminados porque todos esses valores são menores do que a confiabilidade geral de 0,5699 (ou 0,570, arredondando). Diferente das subescalas anteriores, o α global é baixo (0,57) e, embora isso esteja de acordo com a afirmação de Kline que devemos esperar esse tipo de resultado nas ciências sociais, ele está bem abaixo das demais subescalas. Esta subescala tem cinco itens, comparada com sete, oito e três das outras subescalas, assim, sua confiabilidade reduzida não será significativamente afetada pelo número dos itens (na verdade, ela tem mais itens do que a subescala medo de matemática). Se você analisar os itens na

subescala, verá que eles abordam temas diversos de avaliação dos seus pares, e isso pode explicar a falta de consistência. Isso pode nos levar a repensar essa subescala.

15.8 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ②

Este capítulo abordou a análise de fatores, uma técnica para identificar aglomerações de variáveis que se relacionam. Uma das dificuldades da estatística é que ela pode ser subjetiva: muitos livros (inclusive este) dão a impressão de que a estatística é como um bom livro de receitas e, se você seguir as instruções, fará um bolo de chocolate delicioso (hum!). A análise de fatores, talvez mais do que qualquer outro teste neste livro, mostra que isso nem sempre é assim. O mundo da estatística está repleto de regras arbitrárias (a do 0,05 é um exemplo clássico) e, quase sempre, você se dando conta ou não, deve agir usando o seu próprio entendimento. Assim, eu espero que você tenha descoberto o suficiente para trabalhar com a análise de fatores! Vimos que o primeiro estágio da análise de fatores é examinar as suas variáveis para verificar se elas se relacionam de algum modo, mas não muito



Dicas da Samantha Ferrinho: Confiabilidade

- A confiabilidade é a consistência de uma medida.
- A análise de confiabilidade pode ser usada para medir a consistência de um questionário.
- Lembre-se de reverter o escore de qualquer item que teve a frase invertida no questionário original antes de executar a análise.
- Execute análises de confiabilidades separadas para todas as subescalas do seu questionário.
- O α de Cronbach indica a confiabilidade geral do questionário, e valores por volta de 0,8 são bons (ou 0,7 para testes de habilidade e similares).
- A coluna alfa se o item for deletado (*Alpha if Item Deleted*) informa se a remoção de um item irá aumentar a confiabilidade geral: valores maiores do que a confiabilidade geral indicam que se removermos aquele item teremos aumento da confiabilidade geral da escala. Procure por itens que aumentem muito o valor de α .
- Se você realmente remover algum item, execute novamente a análise de fatores para verificar se a estrutura dos fatores se mantém.

Saída 15.18 do SPSS

RELIABILITY ANALYSIS - SCALE (ALPHA) (Análise de Confiabilidade - Escala (Alfa))					
Item-Total Statistics (Estatísticas Item-Total)					
	Scale Mean If Item Deleted (Média da escala se o item for retirado)	Scale Variance if Item Deleted (Variância da escala se o item for retirado)	Corrected Item-total Correlation (Correlação Item-total corrigida)	Squared Multiple Correlation (Correlação múltipla ao quadrado)	Alpha if Item Deleted (Alfa se o item for retirado)
Q02	11.4609	8.1186	0.3389	0.1337	0.5153
Q09	10.2380	6.3955	0.3907	0.1674	0.4765
Q19	10.7923	7.3810	0.3162	0.1060	0.5218
Q22	10.1964	7.2824	0.3776	0.1441	0.4870
Q23	9.6499	7.9879	0.2389	0.0689	0.5628
Reliability Coefficients (Coeficientes de confiabilidade)					
Alpha = .5699 (Alfa = 0.5699)		Standardized item alpha = .5724 (Alfa Padronizado = 0.5724)			

fortemente! A análise de fatores, propriamente dita, tem muitos estágios: verificar alguns assuntos (por exemplo, adequação do tamanho da amostra), decidir quantos fatores reter e, finalmente, decidir quantos itens carregar nos fatores (e tentar entender o significado dos fatores). Tendo feito tudo isso, você deve verificar se os itens que você tem são medidas confiáveis do que você está tentando medir.

15.9 TERMOS-CHAVE QUE DESCOBRIMOS

- Análise de fatores
- Carga dos fatores
- Variância aleatória
- Confiabilidade
- Rotação
- Diagrama de declividade
- Matriz do fator
- Escores dos Fatores
- Matriz de transformação do fator Λ
- Medida da adequação da amostra de Kaiser-Meyer-Olkin (KMO)
- O critério de Kaiser
- Variável latente
- Rotação oblíqua
- Rotação ortogonal
- Matriz padrão
- Análise dos componentes principais
- *Promax*
- *Quartimax*
- Matriz estrutural
- Confiabilidade teste-reteste
- Variância única
- *Varimax*
- Confiabilidade meio-a-meio
- Fatoração alfa
- Método de Anderson-Rubin
- Variância comum
- Comunalidade
- Matriz do componente
- Análise de fatores confirmatória
- α de Cronbach
- *Oblimin* direto
- *Equamax*
- Extração
- Fator

15.10 TAREFAS DO ALEX ESPERTO



- A Universidade de Sussex procura empregar as melhores pessoas possíveis como professores (sério?). Ela queria revisar um questionário base-

ado na teoria de Bland sobre professores de métodos de pesquisa. Essa teoria prevê que bons professores de métodos de pesquisa devem ter quatro características: (1) amor profundo por estatística; (2) entusiasmo por projeto experimental; (3) amor por ensinar e (4) completa ausência de habilidades interpessoais normais. Essas características devem estar relacionadas (isto é, correlacionadas). O TOSSE (*Teaching Of Statistics for Scientific Experiments**) ou já existia, mas a universidade revisou esse questionário e ele se tornou o TOSSE-R (*Teaching Of Statistics for Scientific Experiments-Revised*). A universidade deu esse questionário a 239 professores de métodos de pesquisa em todo o mundo para ver se ele sustentava a teoria de Bland. O questionário está apresentado na Figura 15.14 e os dados estão no arquivo **TOSSE-R.sav**. Realize

uma análise de fatores (com a rotação apropriada) para verificar a estrutura de fatores desses dados. ②

15.11 LEITURAS COMPLEMENTARES

CORTINA, J. M. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, p. 98-104, 1993. Um artigo de boa leitura sobre o α de Cronbach.

DUNTEMAN, G. E. Principal Components analysis. Sage university paper series on quantitative applications in the social sciences, 07-69. Newbury Park (CA): Sage, 1989. Essa monografia é de alto nível, mas abrangente.

FIELD, A. P. Designing a questionnaire. 2000. Essa apostila fornece conselhos sobre como projetar questionários e está no material disponível no *site* no arquivo **DesigningQuestionnaires.pdf**. Vale a pena lê-lo se você quiser elaborar questionários.

PEDHAZUR, E., SCHMELKIN, L. *Measurement, design and analysis*. Hillsdale (NJ): Erlbaum, 1991. O Capítulo 22 é uma excelente introdução à teoria da análise de fatores.

TABACHNICK, B. G., FIELD, L. S. *Using multivariate statistics* (4th edition). Boston: Allyn e Bacon, 2001. O Capítulo 13 tem uma visão técnica, mas maravilhosa, da análise de fatores.

* N. de T.: Em português, EEEEC, Ensino de Estatística para Experimentos Científicos.

DT = Discorda Totalmente, D = Discorda, N = Nem discorda, nem concorda, C = Concorde, CT = Concorde Totalmente						
		DT	D	N	C	CT
1	Uma vez acordei em um canteiro de vegetais abraçado num nabo que eu havia arrancado por engano achando que fosse a maior raiz de Roy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Se eu tivesse uma arma, atiraria nos meus alunos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Eu memorizo valores de probabilidade para a distribuição F.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Eu adoro o santuário de Pearson.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Eu ainda moro com minha mãe e tenho poucos hábitos de higiene.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Ensinar os outros dá vontade de engolir uma garrafa de alvejante – a dor do meu esôfago queimando seria um alívio em comparação.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Ajudar os outros entender a soma dos quadrados é maravilhoso.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	Eu gosto de condições de controle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	Eu calculo três ANOVAs de cabeça antes de levantar da cama.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	Eu poderia passar o dia todo explicando estatística para as pessoas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11	Gosto quando as pessoas me dizem que eu as ajudei a entender a rotação dos fatores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12	As pessoas caem no sono no momento em que eu abro minha boca para falar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13	Projetos de experimentos são divertidos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14	Eu prefiro pensar sobre variáveis dependentes apropriadas a ir ao bar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15	Eu me borro de medo com a simples menção da análise dos fatores.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16	Pensar sobre o uso ou não de medidas repetidas ou independentes me excita.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17	Adoro sentar no parque e meditar sobre o uso da observação do participante no meu próximo experimento.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18	Ficar na frente de 300 pessoas não me causa ansiedade.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19	Eu gosto de ajudar os alunos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
20	Passar conhecimento adiante é o maior presente que você pode dar a alguém.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21	Pensar sobre as correções de Bonferroni me dá um formigamento na virilha.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22	Eu tremo de excitação quando penso no meu próximo experimento.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23	Eu geralmente passo o meu tempo livre falando com os pombos... e eles morrem de tédio.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24	Eu tentei construir uma máquina do tempo para que pudesse voltar aos anos 30 para seguir Fisher de joelhos e lamber o chão que ele pisasse.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25	Eu adoro dar aula.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26	Eu passo muito tempo ajudando os alunos.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27	Eu adoro dar aula porque os alunos precisam fingir que gostam de mim para não ganharem péssimas notas.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
28	Meu gato é meu único amigo.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 15.14 O TOSSE-R (*Teaching Of Statistics for Scientific Experiments-Revised*).

16

DADOS CATEGÓRICOS

16.1 O QUE VOCÊ VAI APRENDER NESTE CAPÍTULO? ①

Finalmente o último capítulo! Do meu ponto de vista, guardei o pior para o fim porque raramente coleciono dados categóricos, desse modo, sei pouco sobre como analisá-los! Mas, a completa ausência de conhecimento sobre um tópico nunca me impediu de escrever sobre ele no passado. Algumas vezes, não estamos interessados em testar escores ou medidas contínuas, mas sim *variáveis categóricas*. Essas são o que até agora denominamos variáveis de grupo. Elas são variáveis que descrevem categorias de entidades. Vimos esse tipo de variável praticamente em todos os capítulos deste livro. Existem tipos diferentes de variáveis categóricas (veja a Seção 4.5.6), mas teoricamente uma pessoa ou caso deve estar em apenas uma categoria. Bons exemplos de variáveis categóricas são gênero (com poucas exceções, as pessoas podem ser biologicamente somente homens ou mulheres)¹, gravidez (uma mulher pode estar grávida ou

não) e votar numa eleição (como regra geral, é permitido votar em um único candidato). Em todos os casos (exceto na regressão logística), até agora usamos tais variáveis para prever algum tipo de resultado contínuo, mas às vezes queremos verificar os relacionamentos entre diversas variáveis categóricas. Este capítulo trata de duas técnicas para fazer isso. Começamos com o caso simples de duas variáveis categóricas e descobrimos a estatística qui-quadrado (a qual não estamos, de fato, descobrindo porque inconscientemente vimos ela várias vezes antes). Depois, estendemos o modelo para verificar os relacionamentos entre diversas variáveis categóricas.

16.2 TEORIA DA ANÁLISE DE DADOS CATEGÓRICOS ①

Vamos começar este capítulo olhando para a situação mais simples que podemos encontrar: isto é, analisar duas variáveis categóricas. Se quisermos olhar o relacionamento entre duas variáveis categóricas, não podemos usar a média ou qualquer estatística similar porque não temos qualquer variável que tenha sido medida de modo contínuo. Não faz sentido tentar calcular a média de uma variável

¹ Antes que alguém proteste, sei que existem cromossomos e condições hormonais que complicam a questão. Também, as pessoas podem ter diferentes identidades de gênero.

categórica porque os valores numéricos que anexamos para categorias diferentes são arbitrários e a média daqueles valores numéricos dependerá de quantos membros cada categoria tem. Portanto, quando medimos somente variáveis categóricas, analisamos frequências. Isto é, analisamos o número de fatores que se encaixam em cada categoria. Como exemplo, um pesquisador estava interessado se animais poderiam ser treinados para dançar em linha*. Ele pegou 200 gatos e tentou treiná-los para dançar dando a eles comida ou afeto como recompensa para o comportamento desejado. No final da semana, ele contou quantos animais podiam dançar e quantos não. Existem duas variáveis categóricas: **training** (treinamento) (o animal foi treinado usando ou comida ou afeição, não ambos) e **dance** (dança) (ele aprendeu a dançar ou não). Combinando as categorias, ficamos com quatro categorias diferentes. Tudo o que precisamos fazer é contar quantos gatos caem dentro de cada categoria. Podemos tabular essas frequências como na Tabela 16.1 (que mostra os dados para esse exemplo), conhecida como *tabela de contingência*.

16.2.1 Teste qui-quadrado de Pearson ①

Se queremos ver se existe um relacionamento entre duas variáveis categóricas (isto é, a quantidade de gatos que dançam está relacionada ao tipo de treinamento usado?) podemos usar o *teste qui-quadrado* de Pearson. Ele é uma estatística extremamente elegante baseada na ideia simples de comparar frequên-

cias que você observou em certas categorias com as frequências que você espera conseguir nessas categorias por acaso. Voltando ao Capítulo 1, 5 e 8, vimos que se ajustarmos um modelo a qualquer conjunto de dados, podemos avaliar esse modelo usando uma equação muito simples (ou alguma variante dela):

$$\text{desvio} = \sum (\text{modelo} - \text{observado})^2$$

Essa equação foi a base para nossas somas de quadrados em regressão e ANOVA. Podemos usar a mesma equação quando temos dados categóricos. Existe uma pequena variação que consiste em dividir os desvios ao quadrado pelos escores do modelo, o que é o mesmo processo que dividir a soma dos quadrados pelos graus de liberdade na ANOVA. A estatística resultante é o qui-quadrado de Pearson (χ^2) e é dada a equação (16.1):

$$\chi^2 = \sum \frac{(\text{Observado}_{ij} - \text{Modelo}_{ij})^2}{\text{Modelo}_{ij}} \quad (16.1)$$

Os dados observados são, obviamente, as frequências na Tabela 16.1, mas precisamos saber em que consiste o modelo nesse caso. Na ANOVA, o modelo que usamos é a média dos grupos, mas mencionei que não podemos trabalhar com médias quando temos variáveis categóricas, portanto, usamos “valores esperados”. Uma maneira de estimar os valores esperados seria dizer “bem, se temos 200 gatos no total e quatro categorias, os valores esperados são simplesmente $200/4 = 50$ ”. Isso seria correto se, por exemplo, tivéssemos o mesmo número de gatos que receberam afeição como recompensa e comida como recompensa, mas isso não ocorreu: 38 ganha-

* N. de T.: Dança típica em que os dançarinos ficam posicionados em uma ou mais linhas paralelas.

Tabela 16.1 Tabela de contingência mostrando quantos gatos irão dançar depois de serem treinados com tipos diferentes de prêmios

		Treinamento		Total
		Comida como recompensa	Afeição como recompensa	
Poderão dançar?	Sim	28	48	76
	Não	10	114	124
Total		38	162	200

ram comida e 162 ganharam afeto como recompensa. Da mesma forma, os números dos que podem ou não dançar não são os mesmos. Para levar isso em consideração, calculamos os valores esperados para cada uma das células da tabela (nesse caso, quatro células) utilizando os totais das linhas e colunas para uma célula em particular a fim de calcular o valor esperado. Assim:

$$\text{Modelo}_{ij} = E_{ij} = \frac{\text{Total da Linha}_i \times \text{Total da Coluna}_j}{n}$$

n é simplesmente o número total de observações (nesse caso, 200). Podemos calcular esses valores esperados para as quatro células da nossa tabela (o total das linhas e o total das colunas são abreviados por TL e TC, respectivamente):

$$\begin{aligned}\text{Modelo}_{\text{Comida, Sim}} &= \frac{\text{TL}_{\text{Sim}} \times \text{TC}_{\text{Comida}}}{n} \\ &= \frac{76 \times 38}{200} = 14,44\end{aligned}$$

$$\begin{aligned}\text{Modelo}_{\text{Comida, Não}} &= \frac{\text{TL}_{\text{Não}} \times \text{TC}_{\text{Comida}}}{n} \\ &= \frac{124 \times 38}{200} = 23,56\end{aligned}$$

$$\begin{aligned}\text{Modelo}_{\text{Afeição, Sim}} &= \frac{\text{TL}_{\text{Sim}} \times \text{TC}_{\text{Afeição}}}{n} \\ &= \frac{76 \times 162}{200} = 61,56\end{aligned}$$

$$\begin{aligned}\text{Modelo}_{\text{Afeição, Não}} &= \frac{\text{TL}_{\text{Não}} \times \text{TC}_{\text{Afeição}}}{n} \\ &= \frac{124 \times 162}{200} = 100,44\end{aligned}$$

Dado que agora temos esses valores do modelo, tudo o que precisamos fazer é pegar cada valor em cada célula dos dados da nossa tabela, subtrair dele o valor correspondente do modelo, elevar o resultado ao quadrado e, então, dividir pelo valor correspondente do modelo. Uma vez que você tenha feito isso para cada célula da tabela, some os resultados.

$$\begin{aligned}\chi^2 &= \frac{(28 - 14,44)^2}{14,44} + \frac{(10 - 23,56)^2}{23,56} \\ &\quad + \frac{(48 - 61,56)^2}{61,56} + \frac{(114 - 100,44)^2}{100,44} \\ &= \frac{(13,56)^2}{14,44} + \frac{(-13,56)^2}{23,56} + \frac{(-13,568)^2}{61,56} \\ &\quad + \frac{(13,56)^2}{100,44} \\ &= 12,73 + 7,80 + 2,99 + 1,83 \\ &= 25,35\end{aligned}$$

Esta estatística pode ser avaliada utilizando uma distribuição χ^2 com propriedades conhecidas. Tudo o que precisamos saber é o grau de liberdade e ele é calculado como $(r-1)(c-1)$, onde r é o número de linhas e c é o número de colunas. Outra maneira de calcular o grau de liberdade é pensar no número de níveis de cada variável menos 1 multiplicados entre si. Nesse caso, temos $(2-1)(2-1) = 1$ gl. Se você estivesse fazendo o teste manualmente, encontraria um valor crítico para a distribuição qui-quadrado com 1 gl e se o valor observado fosse maior do que esse valor crítico, você diria que houve um relacionamento significativo entre as duas variáveis. Esses valores críticos são apresentados no Apêndice A4; para 1 gl os valores críticos são 3,84 (para $p = 0,05$) e 6,63 (para $p = 0,01$) e, assim, como o qui-quadrado calculado é maior do que esses valores, ele é significativo ao nível $p < 0,01$. Entretanto, se você usar o SPSS, ele simplesmente produzirá uma estimativa da probabilidade exata de obter um valor qui-quadrado tão grande quanto esse (no caso, 25,35) apenas por acaso.

16.2.2 A razão de verossimilhança ②

Uma alternativa para o qui-quadrado de Pearson é a estatística da razão de verossimilhança, que é baseada na teoria de máxima verossimilhança. A ideia geral por trás dessa teoria é que você coleta alguns dados e cria um modelo para o qual a probabilidade de obter o conjunto de dados observados é maximizada e, então, compara esse modelo à probabilidade

de obter aqueles dados sob a hipótese nula. A estatística resultante é, portanto, baseada na comparação das frequências observadas com as previstas pelo modelo:

$$L\chi^2 = 2 \sum \text{Observado}_{ij} \ln \left(\frac{\text{Observado}_{ij}}{\text{Modelo}_{ij}} \right) \quad (16.2)$$

onde \ln é o logaritmo natural (essa é uma função padrão matemática que encontramos no Capítulo 6; você pode encontrá-la na sua calculadora como \ln ou \log_e). Usando o mesmo modelo e os valores observados como na seção anterior, isso nos daria:

$$\begin{aligned} L\chi^2 &= 2 \left[28 \times \ln \left(\frac{28}{14,44} \right) + 10 \times \ln \left(\frac{10}{23,56} \right) \right. \\ &\quad \left. + 48 \times \ln \left(\frac{48}{61,56} \right) + 114 \right. \\ &\quad \left. \times \ln \left(\frac{114}{100,44} \right) \right] \\ &= 2[(28 \times 0,662) + (10 \times -0,857) \\ &\quad + (48 \times -0,249) + (114 \times 0,127)] \\ &= 2[18,54 - 8,57 - 11,94 + 14,44] \\ &= 24,94 \end{aligned}$$

Como o qui-quadrado de Pearson, essa estatística tem uma distribuição qui-quadrada com os mesmos graus de liberdade (nesse caso, 1). Assim, ele é testado da mesma forma: isto é, podemos olhar o valor crítico do qui-quadrado para o número de graus de liberdade que temos. Como antes, o valor que temos aqui será significativo porque ele é maior do que os valores críticos de 3,84 (para $p = 0,05$) e 6,63 (para $p = 0,01$). Para amostras maiores, essa estatística será grosseiramente a mesma que o qui-quadrado de Pearson, mas ela é preferível quando as amostras são pequenas.

16.2.3 Correção de Yates ②

Quando você tem uma tabela de continência 2×2 (isto é, duas variáveis categóricas onde cada uma tem duas categorias), o qui-quadrado de Pearson tende a produzir valores significativos que são muito peque-

nos (em outras palavras, ela tende a cometer o erro Tipo I). Portanto, Yates sugeriu uma correção para a fórmula de Pearson (geralmente chamada de *correção de continuidade de Yates*). A ideia básica é que quando você calcula os desvios do modelo (os valores $\text{Observado}_j - \text{Modelo}_{ij}$ na equação (16.1)), subtrai 0,5 do valor absoluto desse desvio antes de elevá-lo ao quadrado. Simplificando, isso significa que você calcula o desvio, ignora se ele é positivo ou negativo, subtrai 0,5 desse desvio e então eleva-o ao quadrado. A equação de Pearson, nesse caso, fica:

$$\chi^2 = \sum \frac{(|\text{Observado}_{ij} - \text{Modelo}_{ij}| - 0,5)^2}{\text{Modelo}_{ij}}$$

Para os dados do nosso exemplo, isso se traduz em:

$$\begin{aligned} \chi^2 &= \frac{(13,56 - 0,5)^2}{14,44} + \frac{(13,56 - 0,5)^2}{23,56} \\ &\quad + \frac{(13,56 - 0,5)^2}{61,56} + \frac{(13,56 - 0,5)^2}{100,44} \\ &= 11,81 + 7,24 + 2,77 + 1,70 \\ &= 23,52 \end{aligned}$$

Uma coisa importante a ser notada é que ele baixa o valor da estatística do qui-quadrado e, portanto, a torna menos significativa. Embora pareça uma boa solução para o problema, existem evidências que isso corrige demais e produz valores do qui-quadrados muito menores! Howell (2002) fornece uma excelente discussão sobre o problema com a correção de continuidade de Yates, se você estiver interessado; tudo o que direi é que embora valha a pena saber sobre isso, provavelmente é melhor ignorar!

16.3 HIPÓTESES DO TESTE QUI-QUADRADO ①

Deveria ser óbvio que o teste do qui-quadrado não se baseia em hipóteses como ter dados contínuos normalmente distribuídos como a maioria dos outros testes neste livro (dados

categóricos não podem ser normalmente distribuídos porque eles não são contínuos). Entretanto, o teste do qui-quadrado tem duas hipóteses importantes:

1. Para o teste ser significativo é imperativo que cada pessoa, item ou entidade contribuam para somente uma célula da tabela de contingência. Portanto, você não pode usar o teste do qui-quadrado num modelo de medidas repetidas (por exemplo, se tivéssemos treinado alguns gatos com comida para ver se dançariam e, então, treinado os mesmos gatos com afeto para ver se dançariam, não poderíamos analisar os dados resultantes com o teste do qui-quadrado de Pearson).
2. O segundo ponto importante é que as frequências esperadas devem ser maiores do que 5 em cada célula; embora seja aceitável que grandes tabelas de contingências tenham até 20% de frequências esperadas abaixo de 5%, o resultado é uma perda de poder estatístico (assim, o teste pode falhar em detectar um efeito genuíno). Mesmo em grandes tabelas de contingência, frequências esperadas não devem ficar abaixo de 1. Howell (2002) fornece uma boa explicação de por que a violação dessa hipótese cria problemas.

Finalmente, embora não seja uma hipótese, é adequado mencionar que proporcionalmente pequenas diferenças em frequências nas células podem resultar em associações estatisticamente significativas entre duas variáveis. Portanto, precisamos olhar as percentagens das linhas e colunas para interpretar qualquer efeito obtido. Essas percentagens refletirão os padrões dos dados bem melhor do que as próprias frequências (porque as frequências dependerão do tamanho da amostra nas diferentes categorias).

16.4 REALIZANDO O TESTE DO QUI-QUADRADO NO SPSS ①

Existem duas maneiras nas quais dados categóricos podem ser fornecidos ao *software*: entrar com os escores brutos ou entrar com casos ponderados. Veremos as duas formas.

16.4.1 Entrando com os dados: escores brutos ①

Entrar com os dados brutos significa que cada linha do editor de dados representa cada entidade das quais temos dados (nesse exemplo, cada linha representa um gato). Assim, você deveria criar duas variáveis codificadas **training** e **dance** (treino e dança) e especificar códigos numéricos apropriados para cada uma. O **training** poderia ser codificado com 0 para representar a recompensa com comida e 1 para representar afeto, e **dance** poderia ser codificado com 0 para representar um animal que dançou e 1 para um que não dançou. Para cada animal, você coloca o código numérico apropriado em cada coluna. Assim, um gato que foi treinado com comida e não dançou teria 0 na coluna do treinamento e 1 na coluna da dança. Os dados no arquivo **Cats.sav** estão apresentados dessa maneira e você deverá ser capaz de identificar as variáveis descritas. Havia 200 gatos no total, assim, há 200 linhas de dados.

16.4.2 Entrando com os dados: ponderando casos ①

Um método alternativo para a entrada dos dados é criar as mesmas variáveis codificadas de antes, mas ter uma terceira variável que representa o número de animais que estão em cada combinação de categorias. Em outras palavras, entramos os dados das frequências (o número de casos de cada categoria). Podemos chamar essa variável de **frequent**. A Figura 16.1 mostra o editor de dados com essa terceira variável adicionada. Agora, em vez de ter 200 linhas, cada uma representa um animal diferente, temos uma linha representando cada combinação de categorias e uma variável mostrando quantos animais estão nessa combinação de categorias. Assim, a primeira linha representa gatos que tiveram comida como recompensa e dançaram. A variável **frequent** mostra 28 gatos nessa situação. A informação foi previamente representada por 28 linhas diferentes no arquivo **Cats.sav** e, assim, você pode ver que esse método de entrada de dados poupa muito tempo! Estendendo esse princí-

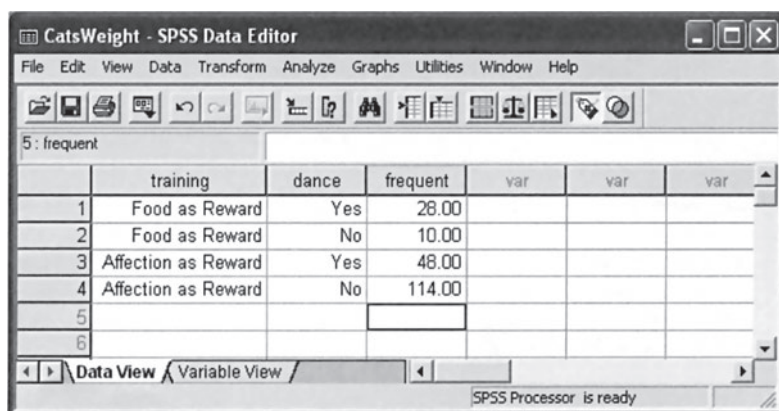


Figura 16.1 Entrada de dados usando casos ponderados.

pio, podemos ver que quando o afeto foi usado como recompensa, 114 gatos não dançaram.

Entrar dados usando uma variável representando o número de casos que estão numa combinação de categorias pode economizar muito tempo. Entretanto, para analisar dados dessa maneira precisamos dizer ao computador que a variável **frequente** representa o número de casos que estão numa combinação particular de categorias. Para fazer isso, acesse a função *casos ponderados* usando o menu **Data⇒Weight Cases...** (Dados⇒Ponderar Casos...) para acessar uma caixa de diálogo semelhante à da Figura 16.2. Selecione a opção **Weight cases** (Ponderar Casos) e depois selecione a variável que representa o número casos (nesse caso, **frequent**) e transfira-a para o quadro **Frequency Variable** (Variável

Frequência) clicando em . Esse processo informa ao computador que ele deve ponderar cada combinação de categorias pelo número na coluna **frequent**. Portanto, o computador irá fingir, por exemplo, que há 28 linhas de dados que tem a combinação de categoria 0, 0 (representando gatos treinados com comida que dançaram). Os dados que entraram dessa maneira estão no arquivo **CatsWeight.sav**, e se você usar esse arquivo lembre-se de ponderar os casos como foi descrito.

16.4.3 Executando a análise ①

Resumir dados em categorias é feito usando o comando **crosstabs** (tabulação cruzada) (que também produz o teste do qui-quadrado). O **crosstabs** está no menu **Descriptive Statis-**

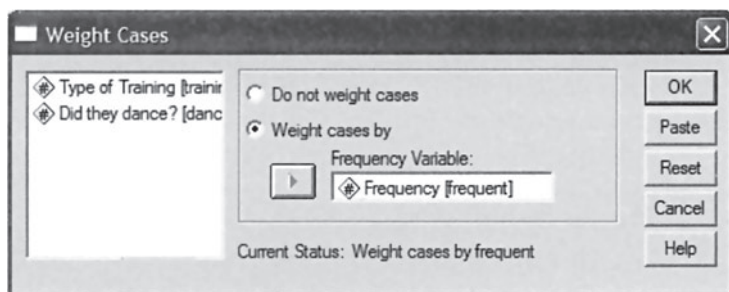









Figura 16.2 A caixa de diálogo para o comando *weight cases* (ponderar casos).

tics (Estatística Descritiva) (**Analyze⇒Descriptive Statistics⇒Crosstabs...**) (Analisar⇒Estatística Descritiva⇒Tabulação Cruzada). A Figura 16.3 mostra as várias caixas de diálogo para o comando **crosstabs** (tabulação cruzada) (a variável **frequent** está no diagrama porque executei a análise nos dados do arquivo **CatsWeight.sav**). Primeiro, entre com uma das variáveis de interesse no quadro **Row(s)** (linha(s)) marcando-a no lado esquerdo e transfira-a clicando em . Para esse exemplo, selecionei **dance** (dança) para ser representada como linhas na tabela. A seguir, selecione a outra variável de interesse, **training** (treinamento), e transfira-a para o quadro **Column(s)** (Colunas) clicando em . Além disso, é possível selecionar uma variável camada (isto é, você pode dividir as linhas de uma tabela em categorias adicionais). Se você tiver uma terceira variável categórica (como teremos mais tarde neste capítulo), poderá dividir a tabela de contingência por essa variável (assim, a camada da tabela representa diferentes categorias dessa terceira variável).

Se você clicar em  (Estatísticas), uma caixa de diálogo aparece em que você pode especificar vários testes estatísticos. As

opções mais importantes do menu estatísticas para dados categóricos estão descritas no Quadro 16.1.

Selecione o teste do qui-quadrado, a correção de continuidade, fi e lambda e, depois, clique em  (Continuar). Se você clicar em  (Células), uma caixa de diálogo aparece na qual você pode especificar o tipo de dados exibidos na tabela de tabulação cruzada. É importante você solicitar as frequências esperadas porque, para que o qui-quadrado seja preciso, os valores esperados devem ser maiores do que certos limites. A regra básica é que em tabelas de contingência 2×2 nenhum valor esperado deve ficar abaixo de 5. Em tabelas maiores, a regra é que todos os valores esperados devem ser maiores do que 1, e não mais do que 20% das frequências esperadas podem ser menores do que 5. É útil, também, olhar a linha, coluna e o total percentual porque esses valores são normalmente mais fáceis de interpretar do que as frequências reais e eles fornecem uma ideia da origem de qualquer efeito significativo. Depois de selecionar essas opções, clique em  (Continuar) para retornar à caixa de diálogo principal. Clique em  (Exato) (se você tem os testes exa-

Quadro 16.1

Opções estatísticas para tabulações cruzadas

- **Qui-quadrado (*Chi-Square*):** Ele executa o teste básico do qui-quadrado de Pearson (Seção 16.2.1). O teste do qui-quadrado detecta se existe uma associação significativa entre duas variáveis categóricas. Entretanto, ele não diz nada sobre quão forte essa associação pode ser.
- **Fi e o V de Cramer (*Phi and Cramer's V*):** Essas são medidas do grau de associação entre duas variáveis categóricas. O Fi é usado com tabelas de contingência 2×2 (tabelas em que você tem duas variáveis categóricas e cada variável tem somente duas categorias). O Fi é calculado dividindo o valor do qui-quadrado pelo tamanho da amostra e, então, tirando a raiz quadrada desse resultado. Se uma das duas variáveis categóricas apresenta mais do que duas categorias, o V de Cramer é mais adequado do que o Fi, porque o Fi não alcança o valor mínimo de zero (indicando nenhuma associação) nesses casos.
- **Lambda (*Lambda*):** O λ de Goodman e Kruskal mede a redução proporcional no erro obtida quando uma variável é utilizada para prever outra variável. Um valor de 1 significa que uma variável prevê perfeitamente a outra, e um valor de 0 indica que uma variável de maneira nenhuma prevê a outra.
- **A estatística de Kendall (*Kendall's statistic*):** Essa estatística é discutida na Seção 4.5.5.

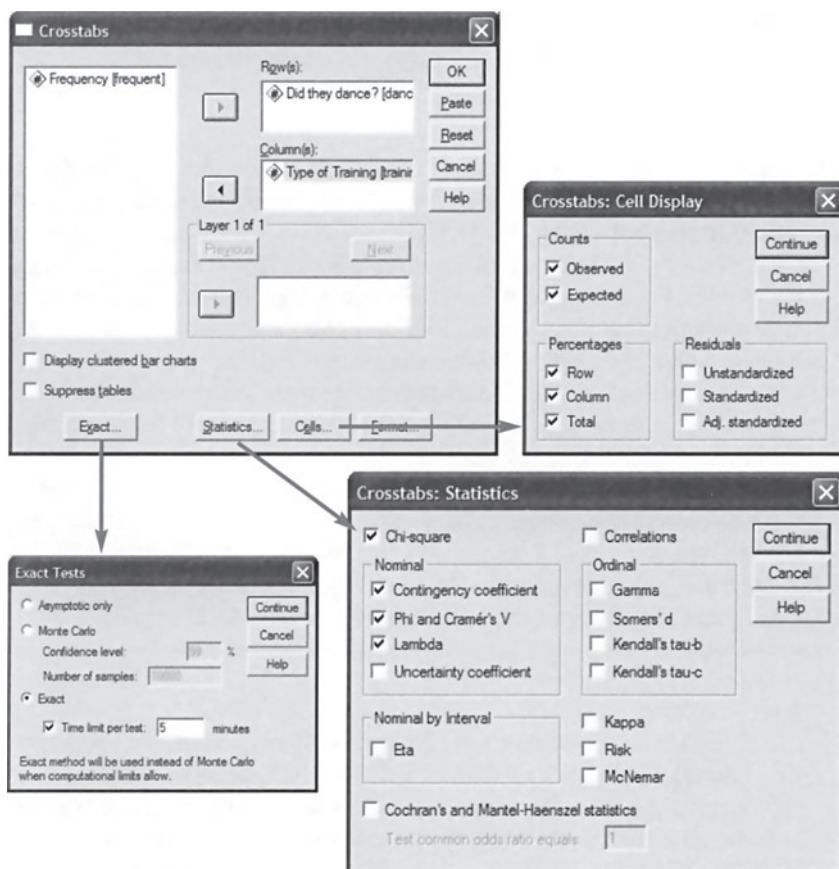


Figura 16.3 Caixa de diálogo para o comando *crosstabs* (tabulações cruzadas).

tos instalados). Expliquei essa opção na Seção 13.2.3, assim, não vou insistir nela aqui, mas selecione a opção teste *Exato*. Clique em **Continue** (Continue) para retornar à caixa de diálogo principal e, então, clique em **OK** para efetuar a análise.

16.4.4 Saída para o teste qui-quadrado ①

A tabela de tabulação cruzada produzida pelo SPSS (Saída 16.1 do SPSS) apresenta o número de casos que se encaixam em cada combinação de categorias e é muito parecida com nossa tabela de contingência original. Podemos ver que no total 76 gatos dançaram (38% do total) e, desses, 28 foram treinados

usando comida (36,8% dos que dançaram) e 48 foram treinados com afeto (63,2% dos que dançaram). Cento e vinte e quatro gatos não dançaram (62% do total) e daqueles que não dançaram 10 foram treinados usando comida como recompensa (8,1% dos que não dançaram) e 114 foram treinados usando afeto (91,9% dos que não dançaram). O número de gatos pode ser lido nas linhas denominadas *Count* (Contagem) e a percentagem pode ser lida nas linhas denominadas *% within Did they dance?* (% dentro de Eles dançaram?). Podemos, também, verificar as percentagens dentro das categorias de treinamento olhando para as linhas chamadas de *% within Type of Training* (% dentro do Tipo de Treinamento). Isso nos informa, por exemplo, que daqueles treinados

Saída 16.1 do SPSS

**Did they dance? Type of Training Crosstabulation (Eles dançaram?
Tabulação cruzada do tipo de Treinamento)**

			Type of Training (Tipo de Treinamento)		Total
			Food as Reward (Comida como recompensa)	Affection as Reward (Afeto como recompensa)	
Did they dance? (Eles dan- çaram?)	Yes (Sim)	Count (Contagem)	28	48	76
		Expected Count (Contagem Esperada)	14.4	61.6	76.0
		% within Did they dance? (% dentro de Eles dançaram?)	36.8%	63.2%	100.0%
		% within Type of Training (% dentro do Tipo de Treinamento)	73.7%	29.6%	38.0%
		% of total (% do total)	14.0%	24.0%	38.0%
	No (Não)	Count (Contagem)	10	114	124
		Expected Count (Contagem Esperada)	23.6	100.4	124.0
		% within Did they dance? (% dentro de Eles dançaram?)	8.1%	91.9%	100.0%
		% within Type of Training (% dentro do Tipo de Treinamento)	26.3%	70.4%	62.0%
		% of total (% do total)	5.0%	57.0%	62.0%
	Total	Count (Contagem)	38	162	200
		Expected Count (Contagem Esperada)	38.0	162.0	200.0
		% within Did they dance? (% dentro de Eles dançaram?)	19.0%	81.0%	100.0%
		% within Type of Training (% dentro do Tipo de Treinamento)	100.0%	100.0%	100.0%
		% of total (% do total)	19.0%	81.0%	100.0%

com comida como recompensa, 73,7% dançaram e 26,3% não dançaram. Similarmente, para aqueles treinados com afeto, somente 29,6% dançaram comparado com 70,4% que não dançaram. Resumindo, quando a comida foi usada como recompensa, a maioria dos gatos dançou, mas quando o afeto foi usado, a maioria dos gatos se recusou a dançar.

Antes de vermos o teste estatístico propriamente dito, é vital checarmos se a hipótese do qui-quadrado foi satisfeita. A hipótese é que nas tabelas 2×2 (que temos aqui) todas as frequências esperadas devem ser maiores do que 5. Se você olhar para as contagens esperadas na tabela de tabulação cruzada (que casualmente é a mesma que calculamos antes), deve ficar claro que a menor contagem esperada é 14,4 (para gatos que foram treinados com comida e dançaram). Esse valor é bem maior do que 5 e, assim, a hipótese está satisfeita. Se você encontrar uma contagem esperada mais baixa do que 5, o melhor seria coletar mais

dados e tentar aumentar a proporção de casos em cada categoria.

Como vimos antes, o teste qui-quadrado de Pearson examina se existe uma associação entre variáveis categóricas (nesse caso, o tipo de treinamento e se o animal dançou ou não). Como parte do procedimento das tabulações cruzadas, o SPSS produz uma tabela (Saída 16.2 do SPSS) que inclui a estatística qui-quadrado e sua significância. A estatística qui-quadrado testa se as duas variáveis são dependentes. Se o valor de significância é pequeno o suficiente (por convenção, a significância deve ser menor do que 0,05), rejeitamos a hipótese de que as variáveis são independentes e aceitamos a hipótese de que elas estão de alguma maneira relacionadas. O valor da estatística qui-quadrado (e os graus de liberdade) é dado na tabela assim como o valor de significância. O valor da estatística qui-quadrado é 25,356, a qual, a menos que haja um erro de arredondamento, é o valor que calculamos

Saída 16.2 do SPSS

Qui-Square Tests (Testes Qui-quadrado)

	Value (Valor)	df (gl)	Asymp. Sig. (Sig. Assintótica)	Exact Sig. (2-sided) (Sig. Exata – Bilateral)	Exact Sig. (1-sided) (Sig. Exata Unilateral)
Pearson Chi-Square (Qui-quadrado de Pearson)	25.356 ^b	1	0.000	0.000	0.000
Continuity Correction ^a (Correção de continuidade)	23.520	1	0.000		
Likelihood Ratio (Razão de verossimilhança)	24.932	1	0.000		
Fisher's Exact Test (Teste Exato de Fisher)					
Linear-by-Linear Association (Associação Linear-por-Linear)	25.229	1	0.000		
N of Valid Cases (Casos Válidos)	200				

a Computed only for 2 × 2 table (Calculado somente para tabelas 2 × 2)
b. 0 cells (.0%) have expected count less than 5. (0 células (0%) tem valor esperado menor do que 5)

na Seção 16.2.1. Esse valor é altamente significativo ($p < 0,001$), indicando que o tipo de treinamento usado teve um efeito significativo em se o animal dançaria ou não.

Uma série de outras estatísticas também está incluída na tabela (muitas delas foram especialmente solicitadas usando as opções de uma caixa de diálogo semelhante à da Figura 16.3). *Continuity Correction* (Correção de Continuidade) é o qui-quadrado com a correção de Yates (veja a Seção 16.2.3) e o seu valor é o mesmo que o valor calculado anteriormente (23,52). Como mencionei antes, é melhor ignorar esse teste, mas ele confirma o resultado do teste principal do qui-quadrado. A *Likelihood Ratio* (razão de verossimilhança) é a estatística que encontramos na Seção 16.2.2 (e, a menos que haja um erro de arredondamento, é o valor que já calculamos: 24,93). Novamente, isso confirma o resultado do qui-quadrado principal, mas essa estatística é preferível em amostras menores.

Embaixo da tabela do qui-quadrado há algumas notas de rodapé relacionadas à hipótese de que as frequências esperadas devem ser maiores do que 5. Se você esqueceu de verificar essas suposições, o SPSS fornece um resumo das contagens esperadas abaixo de 5. Nesse caso, não há frequências esperadas menores do que 5, assim, sabemos que a estatística qui-quadrado é precisa.

O resultado altamente significativo indica que existe uma associação entre o tipo de treinamento e se o gato dançou ou não. O que queremos dizer por associação é que o padrão das respostas (isto é, a proporção de gatos que dançaram para a proporção de gatos que não dançaram) nas duas condições de treinamento é significativamente diferente. O achado significativo reflete no fato de que quando a comida é usada como recompensa, em torno de 74% dos gatos aprenderam a dançar e 26% não, enquanto o afeto é usado, o contrário é verdadeiro (em torno de 70% se recusou a dançar e 30% dançaram). Portanto, podemos concluir que o tipo de treinamento usado influenciou significativamente o comportamento dos gatos: eles dançarão por comida, mas não por amor! Tendo vivido com um adorável gato por muitos anos, isso sustenta minha visão cética de que os gatos nada farão a não ser que tenham um pote de comida esperando por eles!

Se solicitado, o SPSS produzirá outra tabela de saída contendo alguns testes estatísticos adicionais. Muitos desses testes são medidas do grau da associação. Essas medidas são



baseadas na estatística qui-quadrado modificada levando em consideração o tamanho da amostra e os graus de liberdade, e elas tentam restringir a escala da estatística teste ao intervalo de 0 a 1 (para torná-las similares ao coeficiente de correlação descrito no Capítulo 4). Elas são mostradas na Saída 16.3 do SPSS.

- **Fi (Phi):** Essa estatística é precisa para tabelas de contingência 2×2 . Entretanto, para tabelas com mais do que duas dimensões, o valor de Fi pode não estar entre 0 e 1 porque o valor do qui-quadrado pode exceder ao tamanho da amostra. Por esse motivo, Pearson sugeriu o uso do coeficiente de contingência.
- **O V de Cramer (Cramer's V):** Quando ambas as variáveis têm somente duas categorias, o Fi e o V de Cramer são idênticos. Entretanto, quando as variáveis têm mais do que duas categorias, a estatística de Cramer pode alcançar seu máximo de 1 – diferente das outras duas – e, desse modo, ela é a mais útil.
- **Coefficiente de contingência (Contingency Coefficient):** Esse coeficiente produz um valor entre 0 e 1, mas, infelizmente, ele raramente alcança seu limite superior que é 1; por esse motivo, Cramer elaborou o coeficiente V.

Para esses dados, a estatística de Cramer é 0,36 de um máximo possível de 1. Isso representa uma associação média entre o tipo de treinamento e se os gatos dançaram ou não (se você pensar em termos de coeficiente de correlação, isso representa um tamanho de

efeito médio). Esse valor é altamente significativo ($p < 0,001$), indicando que um valor da estatística teste desse tamanho dificilmente ocorreu por acaso e, portanto, a força do relacionamento é significativa. Esses resultados confirmam o que o teste qui-quadrado já havia demonstrado, mas também nos fornece uma ideia do tamanho de efeito.

16.4.5 Calculando o tamanho de efeito ②

Embora o V de Cramer seja um tamanho de efeito adequado (no sentido de que ele está restrito entre 0 e 1 e é, portanto, de fácil interpretação), a mais comum e, possivelmente, a mais útil medida do tamanho de efeito para dados categóricos é o risco relativo. Os riscos relativos são particularmente úteis em tabelas de contingência 2×2 porque para essas tabelas a interpretação do risco relativo é muito clara. Entretanto, isso não é tão restritivo quanto você poderia pensar porque, como relatei nos capítulos MLG, os tamanhos de efeito são somente úteis quando eles resumem uma comparação específica. Uma tabela de contingência 2×2 são dados categóricos equivalentes a uma comparação específica!

O risco relativo é simples o suficiente para poder ser calculado. Se pegarmos nosso exemplo, podemos calcular primeiro a chance de um gato dançar dado que ele teve comida como recompensa. Isso é simplesmente o número de gatos para os quais comida foi fornecida e que dançaram, dividido pelo número

Saída 16.3 do SPSS

Symmetric Measures (Medidas simétricas)

		Value (Valor)	Asymp. Sig. (Sig. Assintótica)
Nominal by Nominal (Nominal versus Nominal)	Phi (Fi)	0.356	0.000
	Cramer's V (V de Cramer)	0.356	0.000
	Contingency Coefficient (Coeficiente de Contingência)	0.335	0.000
N of Valid Cases (Casos Válidos)		200	

a Não assumindo a hipótese nula.
b. Utilizando o erro padrão assintótico assumindo a hipótese nula.

de gatos para os quais comida foi dada e não dançaram:

$$\begin{aligned}\text{Chances}_{\text{dançar após comida}} &= \frac{\text{Número dos que comeram e dançaram}}{\text{Número dos que comeram e não dançaram}} \\ &= \frac{28}{10} \\ &= 2,8\end{aligned}$$

Em seguida, calculamos a chance de um gato dançar dado que ele teve afeto como recompensa. Isso é simplesmente o número de gatos que ganharam afeto e dançaram dividido pelo número de gatos que ganharam afeto e não dançaram:

$$\begin{aligned}\text{Chances}_{\text{dançar após afeto}} &= \frac{\text{Número dos que ganharam afeto e dançaram}}{\text{Número dos que ganharam afeto e não dançaram}} \\ &= \frac{48}{114} \\ &= 0,421\end{aligned}$$

O risco relativo é simplesmente a chance de dançar após comida dividida pela chance de dançar após afeto:

$$\begin{aligned}\text{Risco relativo} &= \frac{\text{Chance}_{\text{Dançar após comer}}}{\text{Chance}_{\text{Dançar após afeto}}} \\ &= \frac{2,8}{0,421} \\ &= 6,65\end{aligned}$$

Isso nos informa que se um gato foi treinado com comida, ele teve 6,65 mais chance de dançar. É uma métrica extremamente elegante e de fácil compreensão para expressar o efeito que você conseguiu!

16.4.6 Relatando os resultados do qui-quadrado ①

Quando relatamos o qui-quadrado de Pearson, simplesmente relatamos o valor da estatística teste com seus graus de liberdades associados e o valor da significância. A estatística teste, como vimos, é representada por χ^2 . A saída do SPSS nos informa que o valor de χ^2 era de 25,36 e que os graus de liberdade nos quais ele foi baseado foram 1 e que ele era significativo com ($p < 0,001$). Também é útil reproduzir a tabela de contingência e avaliar o risco relativo. Desse modo, podemos relatar:

- ✓ Houve uma associação significativa entre o tipo de treinamento e se os gatos dançaram ou não, $\chi^2 (1) = 25,36$, ($p < 0,001$). Isso parece representar o fato de que, baseado

Dica da Samanta Ferrinho



- Se você quiser testar o relacionamento entre duas variáveis categóricas, pode fazê-lo com o teste do qui-quadrado de Pearson ou a estatística da razão de verossimilhança.
- Olhe para a tabela *Testes qui-quadrado*. Se o Valor da Significância Exata é menor do que 0,05 para a linha do qui-quadrado de Pearson, existe um relacionamento significativo entre suas duas variáveis.
- Verifique as notas da tabela para ter certeza de que frequências esperadas não são menores do 5.
- Olhe para a tabela de tabulação cruzada para saber qual é o relacionamento entre as variáveis – melhor ainda, calcule o risco relativo.
- Relate a estatística χ^2 , os graus de liberdade e o valor da significância. Também apresente a tabela de contingência.

no risco relativo, os gatos tiveram 6,65 mais chances de dançar se treinados com comida do que se treinados com afeto.

16.5 DIVERSAS VARIÁVEIS CATEGÓRICAS: ANÁLISE LOGLINEAR ③

Até agora vimos situações nas quais há somente duas variáveis categóricas. Entretanto, geralmente queremos analisar tabelas de contingência mais complexas em que há três ou mais variáveis. Por exemplo, que tal se tomarmos o exemplo que acabamos de usar, mas também coletarmos dados de uma amostra de 70 cachorros? Teremos agora três variáveis: **animal** (animal) (cachorro ou gato), **training** (treinamento) (comida ou afeto como recompensa) e **dance** (dança) (eles dançaram ou não?). Isso não pode ser analisado com o qui-quadrado de Pearson, mas com uma técnica chamada de *análise loglinear*.

16.5.1 Qui-quadrado como regressão ④



Para começar, vamos ver como o nosso simples qui-quadrado pode ser expresso na forma de um modelo de regressão.

Embora já saibamos o suficiente sobre o teste qui-quadrado, para entender situações mais complexas, é mais simples considerar nosso modelo como um modelo linear geral (isto é, de regressão). Todos os modelos gerais lineares que consideramos neste livro tomam a forma geral de:

$$\text{Saída}_i = (\text{Modelo}_i) + \text{erro}_i$$

Por exemplo, quando apresentamos a regressão múltipla no Capítulo 5, vimos que esse modelo estava escrito como (veja a Equação (5.9)):

$$Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n) + \varepsilon_i$$

Também, quando apresentamos a ANOVA de um fator, adaptamos esse modelo de regressão para conceitualizar nosso exemplo do Viagra como (veja a Equação (8.2)):

$$\text{Libido}_i = b_0 + b_2\text{Alta}_i + b_1\text{Baixa}_i + \varepsilon_i$$

O teste *t* foi concebido de maneira similar. Em todos os casos, a mesma equação básica é usada, somente a complexidade do modelo muda. Com os dados categóricos, podemos usar o mesmo modelo da mesma maneira que usamos com a regressão para produzir um modelo linear. No nosso exemplo atual, temos duas variáveis categóricas: treinamento (comida ou afeto) e dança (sim, eles dançaram, ou não, eles não dançaram). As duas variáveis têm duas categorias cada e, assim, podemos representar cada uma com uma simples variável auxiliar (*dummy*) (veja a Seção 5.10.1), em que uma categoria é codificada como 0 e a outra, como 1. Assim, para treinamento podemos codificar “comida” como 0 e “afeto” como 1, e podemos codificar a variável dança como 0 para “sim” e 1 para “não” (veja a Tabela 16.2).

Essa situação pode parecer familiar se você lembrar da ANOVA fatorial (Seção 10.8), na qual também tínhamos duas variáveis previsoras. Naquela situação, vimos que quando existem duas variáveis, o modelo linear geral torna-se (lembre da Equação (10.1)):

$$\text{Saída}_i = (b_0 + b_1A_i + b_2B_i + b_3AB_i) + \varepsilon_i$$

em que A representa a primeira variável, B representa a segunda variável e AB representa a interação entre as duas variáveis. Portanto, podemos construir um modelo linear usando essas variáveis auxiliares que são exatamente

Tabela 16.2 Esquema de códigos para os gatos dançantes

Treinamento	Dança	Treino (Auxiliar)	Dança (Auxiliar)	Interação	Frequência
Comida	Sim	0	0	0	28
Comida	Não	0	1	0	10
Afeto	Sim	1	0	0	48
Afeto	Não	1	1	1	114

as mesmas que utilizamos na ANOVA fatorial (acima). O termo de interação será, simplesmente, a variável treinamento multiplicada pela variável dança (veja a Tabela 16.2 e, se não fizer sentido, retorne à Seção 10.8, pois o código é exatamente o mesmo desse exemplo):

$$\begin{aligned} \text{Saída}_i &= (\text{Modelo}_i) + \text{erro}_i \\ \text{Saída}_{ij} &= (b_0 + b_1 \text{Treinamento}_i \\ &\quad + b_2 \text{Dança}_j + b_3 \text{Interação}_{ij}) + \varepsilon_{ij} \quad (16.3) \end{aligned}$$

No entanto, em virtude de estarmos usando dados categóricos, para tornar o modelo linear temos que trabalhar com o logaritmo (veja o Capítulo 6) e, assim, o modelo real se torna:²

$$\begin{aligned} \ln(O_i) &= \ln(\text{Modelo}_i) + \ln(\varepsilon_i) \\ \ln(O_{ij}) &= (b_0 + b_1 \text{Treinamento}_i \\ &\quad + b_2 \text{Dança}_j + b_3 \text{Interação}_{ij}) \\ &\quad + \ln(\varepsilon_{ij}) \quad (16.4) \end{aligned}$$

As variáveis treinamento e dança e a interação podem ter os valores 0 e 1 dependendo de qual combinação de categorias estamos olhando (Tabela 16.2). Portanto, para saber o que os valores b representam nesse modelo, podemos fazer o mesmo que foi feito para o teste t e a ANOVA e ver o que acontece quando trocamos treinamento e dança pelos valores de 0 e 1. Para começar, vamos ver o que acontece quando treinamento e dança são zero. Isso representa a categoria de gatos que tiveram comida como recompensa e dançaram. Quando usamos esse tipo de modelo para o teste t e a ANOVA, as saídas que usamos foram tiradas dos dados observados: usamos as médias do grupo (por exemplo, veja as Seções 7.8 e 8.2.2). Entretanto, com variáveis categóricas as médias são menos significativas porque não medimos nada numa escala ordinal ou intervalar; em vez disso, temos apenas dados de frequência. Portanto, simplesmente usamos as frequências observadas (em vez das médias observadas) como nos

resultados. Na Tabela 16.1, vimos que havia 28 gatos que tiveram comida como recompensa e dançaram; se usarmos isso como resultado observado, o modelo pode ser escrito como (se ignorarmos o termo erro no momento):

$$\ln(O_{ij}) = b_0 + b_1 \text{Treinamento}_i + b_2 \text{Dança}_j + b_3 \text{Interação}_{ij}$$

Para gatos que tiveram como recompensa comida e dançaram, as variáveis treinamento e dança e a interação serão 0, e a equação se reduz a:

$$\begin{aligned} \ln(O_{\text{comida, sim}}) &= b_0 + (b_1 \times 0) + (b_2 \times 0) \\ &\quad + (b_3 \times 0) \\ \ln(O_{\text{comida, sim}}) &= b_0 \\ \ln(28) &= b_0 \\ b_0 &= 3,332 \end{aligned}$$

Portanto, b_0 no modelo representa o logaritmo natural (\ln) do valor observado quando todas as categorias são zero. Desse modo, ele é o logaritmo natural (\ln) do valor observado da categoria base (nesse caso, gatos que ganharam comida e dançaram). Agora, vejamos o que acontece quando olhamos os gatos que tiveram afeto como recompensa e dançaram. Nesse caso, a variável treinamento é 1 e a variável dança e a interação são ainda 0. Também, nossa saída agora muda para ser o valor observado para gatos que receberam afeto e dançaram (da Tabela 16.1, podemos ver que o valor é 48). Portanto, a equação se torna:

$$\begin{aligned} \ln(O_{\text{Afeto, Sim}}) &= b_0 + (b_1 \times 1) + (b_2 \times 0) \\ &\quad + (b_3 \times 0) \\ \ln(O_{\text{Afeto, Sim}}) &= b_0 + b_1 \\ b_1 &= \ln(O_{\text{Afeto, Sim}}) - b_0 \end{aligned}$$

Lembrando que b_0 é o valor esperado para gatos que ganharam comida e dançaram, temos:

$$\begin{aligned} b_1 &= \ln(O_{\text{Afeto, Sim}}) - \ln(O_{\text{comida, sim}}) \\ &= \ln(48) - \ln(28) \\ &= 3,871 - 3,332 \\ &= 0,539 \end{aligned}$$

² Na verdade, a convenção é usar o b_0 como θ e os valores b como λ , mas acredito que essas mudanças nos símbolos servem somente para confundir as pessoas, assim, permaneço com o b porque quero enfatizar as semelhanças entre a regressão e a ANOVA.

O importante é que b_1 é a diferença entre o logaritmo natural (\ln) da frequência observada para gatos que receberam afeto e dançaram e o logaritmo natural (\ln) dos valores observados para gatos que receberam comida e dançaram. Em outras palavras, dentro do grupo de gatos que dançaram ele representa a diferença entre aqueles treinados usando comida e aqueles treinados com afeto.

Agora, vejamos o que acontece quando olhamos os gatos que tiveram comida como recompensa e não dançaram. Nesse caso, a variável treinamento é 0, a variável dança é 1 e a interação é novamente 0. Nossa saída agora muda para ser a frequência observada para gatos que receberam comida, mas não dançaram (da Tabela 16.1, podemos ver que o valor é 10). Assim, a equação fica:

$$\ln(O_{\text{Comida, Não}}) = b_0 + (b_1 \times 0) + (b_2 \times 1) + (b_3 \times 0)$$

$$\ln(O_{\text{Comida, Não}}) = b_0 + b_2$$

$$b_2 = \ln(O_{\text{Comida, Não}}) - b_0$$

Lembrando que b_0 é o valor esperado para gatos que ganharam comida e dançaram, temos:

$$\begin{aligned} b_2 &= \ln(O_{\text{Comida, Não}}) - \ln(O_{\text{Comida, Sim}}) \\ &= \ln(10) - \ln(28) \\ &= 2,303 - 3,332 \\ &= -1,029 \end{aligned}$$

O importante é que b_2 é a diferença entre o logaritmo natural (\ln) da frequência observada para gatos que receberam comida e dançaram e o logaritmo natural da frequência observada para gatos que receberam comida e não dançaram. Em outras palavras, dentro do grupo dos gatos que receberam comida como recompensa, ele representa a diferença entre gatos que não dançaram e aqueles que dançaram. Finalmente, podemos olhar os gatos que ganharam afeto e dançaram. Nesse caso, ambas as variáveis treinamento e dança são 1 e a interação (o valor do treinamento multiplicado pelo valor da dança) também é 1. Podemos, também, trocar b_0 , b_1 e b_2 pelo que sabemos que eles

representam. A saída é o logaritmo natural (\ln) da frequência observada para gatos que receberam afeto, mas não dançaram (esse valor esperado é 114 – veja a Tabela 16.1); portanto, a equação fica (abreviei A para afeto, C para comida, S para sim e N para não):

$$\log(O_{A,N}) = b_0 + (b_1 \times 1) + (b_2 \times 1) + (b_3 \times 1)$$

$$\ln(O_{A,N}) = b_0 + b_1 + b_2 + b_3$$

$$\ln(O_{A,N}) = \ln(O_{C,S}) + [\ln(O_{A,S}) - \ln(O_{C,S})] + [\ln(O_{C,N}) - \ln(O_{C,S})] + b_3$$

$$\ln(O_{A,N}) = \ln(O_{A,S}) + \ln(O_{C,S}) - \ln(O_{C,S}) + b_3$$

$$b_3 = \ln(O_{A,N}) - \ln(O_{C,N}) + \ln(O_{C,S}) - \ln(O_{A,S})$$

$$= \ln(114) - \ln(10) + \ln(28)$$

$$= \ln(48)$$

$$= 1,895$$

Assim, o b_3 no modelo realmente compara a diferença entre afeto e comida quando os gatos não dançaram e a diferença entre comida e afeto quando os gatos dançaram. Em outras palavras, ele compara o efeito do treinamento quando os gatos não dançaram ao efeito do treinamento quando eles dançaram.

O modelo final é, portanto:

$$\begin{aligned} \ln(O_{ij}) &= 3,332 + 0,539\text{Treinamento} \\ &\quad - 1,029\text{Dança} + 1,895\text{Interação} + \ln(\varepsilon_{ij}) \end{aligned}$$

O importante a ser observado aqui é que tudo é exatamente igual à ANOVA fatorial exceto que agora estamos lidando com valores transformados logaritmicamente (compare essa seção com a Seção 10.8 para ver como isso tudo é parecido). No caso de você ainda não acreditar que isso funciona como um MLG (Modelo Linear Geral), preparei um arquivo chamado **CatRegression.sav** que contém as duas variáveis **dance** (dança) e **training** (treinamento) (uma variável auxiliar (*dummy*) codificada como zero e a outra como um, como está descrito acima) e a interação (**int**). Existe, também, uma variável chamada de **observed** (observado), que contém as frequências observadas na Tabela 16.1 para cada combinação de

dança e treinamento. Finalmente, existe uma variável chamada de **lno**, que é o logaritmo natural dessas frequências observadas (lembre que nessa seção discutimos os logaritmos dos valores observados). Execute uma análise de regressão múltipla usando esse arquivo com **lno** como a saída e **training**, **dance** e **interaction** como seus três previsores. A Saída 16.4 do SPSS mostra a tabela dos coeficientes resultantes dessa regressão. Note que a constante b_0 é 3,332, como foi calculado acima, o valor beta para o tipo de treinamento b_1 é 0,539, e para dança b_2 é -1,030, ambos estão dentro do erro de arredondamento o qual foi calculado acima. Ainda, o coeficiente para a interação, b_3 , é 1,895 como previsto. Existe um ponto importante: todos os erros padrões são zero ou, colocando de outra forma, *não* existe erro nesse modelo. Isso porque as diversas combinações de variáveis codificadas explicam completamente os valores observados. Isso é conhecido como *modelo saturado* e devo retornar a esse ponto mais tarde, portanto, lembre-se dele. Por ora, espero que isso o convença de que o qui-quadrado pode ser conceitualizado como um modelo linear.

Tudo isso está muito bem, mas o título dessa seção subentende que vou mostrar como o teste do qui-quadrado pode ser conceitualizado como um modelo linear. Basicamente, o teste do qui-quadrado verifica se duas variáveis são independentes; portanto, ele não tem interesse no efeito combinado das duas variáveis, somente no seu efeito único. Assim, po-

demos conceitualizar o qui-quadrado da mesma forma que o modelo saturado, apenas não incluímos o termo de interação. Se removermos o termo de interação, nosso modelo fica:

$$\ln(\text{Modelo}_{ij}) = b_0 + b_1 \text{Treinamento}_i + b_2 \text{Dança}_j$$

Com esse novo modelo, não podemos prever os valores observados como fizemos com o modelo saturado porque perdemos algumas informações (a saber, o termo de interação). Portanto, a saída do modelo muda e, assim, os betas mudam também. Vimos anteriormente que o teste qui-quadrado é baseado em “frequências esperadas”. Portanto, se estamos conceitualizando o teste qui-quadrado como um modelo linear, nossas saídas serão esses valores esperados. Se você voltar ao início deste capítulo, verá que já temos as frequências esperadas baseadas nesse modelo. Podemos recalcular os betas baseados no valor esperado:

$$\ln(E_{ij}) = b_0 + b_1 \text{Treinamento}_i + b_2 \text{Dança}_j$$

Para gatos que tiveram comida como recompensa e dançaram, as variáveis de treinamento e dança serão 0 e, assim, a equação se reduz a:

$$\begin{aligned} \ln(E_{\text{Comida, Sim}}) &= b_0 + (b_1 \times 0) + (b_2 \times 0) \\ \ln(E_{\text{Comida, Sim}}) &= b_0 \\ b_0 &= \ln(14,44) \\ &= 2,67 \end{aligned}$$

Saída 16.4 do SPSS

Coefficients ^a (Coeficientes)						
Model (Modelo)		Unstandardized Coefficients (Coeficientes não-padronizados)		Standardized Coefficients (Coeficientes padronizados)	t	Sig.
		B	Std. Erro (Erro Padrão)	Beta		
1	(Constant) (Constante)	3.332	0.000		6.3E+08	0.000
	Type of Training (Tipo de Treinamento)	0.539	0.000	0.307	8.1E+07	0.000
	Did they dance? (Eles dançaram?)	-1.030	0.000	-0.725	-1.0E+08	0.000
	Interaction (Interação)	1.895	0.000	1.361	1.7E+08	0.000

a Variável Dependente: LN (Valores Observados).

Portanto, b_0 no modelo representa o logaritmo do valor esperado quando todas as categorias são 0.

Quando olhamos os gatos que tiveram afeto como recompensa e dançaram, a variável treinamento é 1 e a variável dança e a interação são ainda 0. Também, nossa saída agora muda para ser o valor esperado para gatos que receberam afeto e dançaram:

$$\begin{aligned}\ln(E_{\text{Afeto, Sim}}) &= b_0 + (b_1 \times 1) + (b_2 \times 0) \\ \ln(E_{\text{Afeto, Sim}}) &= b_0 + b_1 \\ b_1 &= \ln(E_{\text{Afeto, Sim}}) - b_0 \\ &= \ln(E_{\text{Afeto, Sim}}) - \ln(E_{\text{Comida, Sim}}) \\ &= \ln(61,56) - \ln(14,44) \\ &= 1,45\end{aligned}$$

O importante é que b_1 é a diferença entre o logaritmo da frequência esperada para gatos que receberam afeto e dançaram e o logaritmo dos valores esperados para gatos que receberam comida e dançaram. Na verdade, os valores são os mesmos da coluna marginal; isto é, a diferença entre o número total de gatos que ganharam afeto e o número total de gatos que ganharam comida: $\ln(162) - \ln(38) = 1,45$.

Quando olhamos os gatos que ganharam comida como recompensa e não dançaram, a variável treinamento é 0 e a variável dança é 1. Nossa saída agora muda para ser a frequência esperada para gatos que receberam comida, mas não dançaram:

$$\begin{aligned}\ln(E_{\text{Comida, Não}}) &= b_0 + (b_1 \times 0) + (b_2 \times 1) \\ \ln(E_{\text{Comida, Não}}) &= b_0 + b_2 \\ b_2 &= \ln(O_{\text{Comida, Não}}) - b_0 \\ &= \ln(O_{\text{Comida, Não}}) - \ln(O_{\text{Comida, Não}}) \\ &= \ln(23,56) - \ln(14,44) \\ &= 0,49\end{aligned}$$

Portanto, b_2 é a diferença entre o logaritmo das frequências esperadas para gatos que receberam comida e dançaram ou não. Na verdade, o valor é o mesmo do que o da coluna marginal; isto é, a diferença entre o número total de gatos que dançaram ou não: $\ln(124) - \ln(76) = 0,49$.

Podemos verificar novamente tudo isso olhando a célula final:

$$\begin{aligned}\ln(E_{\text{Afeto, Não}}) &= b_0 + (b_1 \times 1) + (b_2 \times 1) \\ \ln(E_{\text{Afeto, Não}}) &= b_0 + b_1 + b_2 \\ \ln(100,44) &= 2,67 + 1,45 + 0,49 \\ 4,61 &= 4,61\end{aligned}$$

O modelo final qui-quadrado é, portanto:

$$\begin{aligned}\ln(O_{ij}) &= \text{Modelo}_{ij} + \ln(\varepsilon_{ij}) \\ \ln(O_{ij}) &= 2,67 + 1,45\text{Treinamento} \\ &\quad + 0,49\text{Dança} + \ln(\varepsilon_{ij})\end{aligned}$$

Podemos reorganizar isso para conseguir os resíduos (o termo erro):

$$\ln(\varepsilon_{ij}) = \ln(O_{ij}) - \text{Modelo}_{ij}$$

Nesse caso, o modelo é apenas as frequências esperadas calculadas para o teste do qui-quadrado, assim, os resíduos são as diferenças entre as frequências observadas e esperadas.

Isso demonstra como o qui-quadrado pode funcionar como um modelo linear, como a regressão e a ANOVA, nos quais os betas nos informam algo sobre as diferenças relativas em frequências através de categorias das nossas duas variáveis. Se você não estiver entendendo nada, quero que pelo menos você aprenda com essa seção que o qui-quadrado (e a análise de dados categóricos, geralmente) pode ser expresso como um modelo linear (embora tenhamos que usar valores de logaritmos). Podemos expressar categorias de uma variável usando variáveis auxiliares (*dummy*) assim como fizemos com a regressão e a ANOVA e os valores betas resultantes podem ser calculados exatamente da mesma que foram calculados para a regressão e a ANOVA. Na ANOVA, os valores betas representaram as diferenças entre as médias de uma categoria particular comparada com uma categoria base. Com dados categóricos, os valores beta representam o mesmo, a única diferença sendo que em vez de tratar de médias, estamos tratando de valores esperados. Entender essa ideia (que regressão, testes t , ANOVAs e



análise de dados categóricos são basicamente os mesmos) irá lhe ajudar consideravelmente na próxima seção.

16.5.2 Análise Loglinear ③

Na seção anterior, tentando explicar que a análise de dados categóricos é apenas outra forma de regressão, processei os dados através de uma regressão ordinária no SPSS para provar que eu não estava falando besteira. Naquele momento, avoadamente disse que “oh, a propósito, não existe erro no modelo, estranho, não?” e dei importância ao fato afirmando que era um modelo “saturado” e que não era preciso se preocupar com isso porque eu iria explicar mais tarde – assim que eu descobrisse o que estava acontecendo. Parecia uma boa tática para evitar o assunto naquele momento, mas agora devo explicar isso.

Para começar, espero que você esteja feliz com a ideia de que dados categóricos podem ser expressos na forma de um modelo linear desde que usemos valores de logaritmos (por isso, a propósito, a técnica que estamos usando é chamada de análise loglinear). Do que você já sabe sobre a ANOVA e modelos lineares em geral, deveria estar deitado na cama bem aconchegado com a ideia de que podemos estender qualquer modelo linear para incluir qualquer quantidade de previsores e qualquer termo de interação resultante entre previsores. Portanto, se podemos representar uma simples análise categórica de duas variáveis em termos de um modelo linear, não deveria impressioná-lo o fato de que se temos mais do que duas variáveis, isso não é um problema: podemos estender o modelo simplesmente adicionando qualquer variável e os termos de interação resultantes. Isso é tudo que você realmente precisa saber. Assim, exatamente como na regressão múltipla e na ANOVA, se pensarmos nas coisas em termos de um modelo linear, conceitualmente, é muito fácil entender como o modelo se expande para incorporar novas variáveis. Assim, por exemplo, se tivermos três previsores (A, B e C) na ANOVA (lembre-se da Seção 12.3), acabamos com três interações de dois fatores (AB, AC, BC) e uma interação

de três fatores (ABC). Portanto, o modelo linear resultante disso é apenas:

$$\text{Saída}_j = (b_0 + b_1A + b_2B + b_3C + b_4AB + b_5AC + b_6BC + b_7ABC) + \varepsilon_j$$

Exatamente da mesma maneira, se temos três variáveis numa análise de dados categóricos, conseguimos um modelo idêntico, mas com uma saída em termos de logaritmos:

$$\ln(O_{ijk}) = (b_0 + b_1A_i + b_2B_j + b_3C_k + b_4AB_{ij} + b_5AC_{ik} + b_6BC_{jk} + b_7ABC_{ijk}) + \ln(\varepsilon_{ijk})$$

Obviamente, o cálculo dos valores de beta e dos valores esperados do modelo se torna consideravelmente mais confuso, mas é por isso que inventamos computadores: para que não precisemos nos preocupar com isso! A análise loglinear funciona com base nesses princípios. Entretanto, como vimos no caso de duas variáveis, quando nossos dados são categóricos e incluímos todos os termos disponíveis (efeitos principais e interações), não erramos: nossos previsores podem perfeitamente prever nossa saída (os valores esperados). Assim, se começarmos com o modelo mais complexo possível, não erramos. A tarefa da análise loglinear é tentar ajustar um modelo mais simples aos dados sem uma perda substancial do poder preditivo. Portanto, a análise loglinear tipicamente trabalha num princípio de eliminação de trás pra frente (sim, o mesmo tipo de eliminação de trás para frente que podemos usar na regressão múltipla – veja a Seção 5.5.3.3). Assim, começamos com o modelo saturado e, depois, removemos um preditor do modelo e usando esse novo modelo prevemos nossos dados (calculando as frequências esperadas como no teste do qui-quadrado). Então, verificamos quão bem o modelo se ajusta aos dados (isto é, se as frequências esperadas estão próximas das frequências observadas). Se a aderência do novo modelo não for muito diferente do modelo complexo, abandonamos o modelo complexo em favor do novo. Colocando de outra maneira, presumimos que o termo que removemos não estava tendo um impacto significativo na habilidade do nosso modelo para prever os dados observados.

Entretanto, a análise não apenas remove os termos ao acaso, ela faz isso de forma hierárquica. Assim, começamos com o modelo saturado e, depois, removemos a interação da ordem mais alta e avaliamos o efeito que ela tinha. Se a remoção desse termo de interação não afetou o modelo, ele obviamente não estava sendo muito útil; portanto, nos livramos dele e continuamos removendo outras interações de ordem mais baixa agora. Se a remoção dessas interações não tem impacto, continuamos com os efeitos principais até encontrarmos um efeito que realmente afete a adequação do modelo se for removido.

Colocando isso de forma mais objetiva, no início desta seção de análise loglinear pedi para você imaginar que estendemos nossos exemplos de treinamento e dança para incorporar uma amostra de cachorros. Assim, sabemos que temos três variáveis: animal (cachorro ou gato), treinamento (comida ou afeto) e dança (eles dançaram ou não?). Da mesma forma que na ANOVA, isso resulta em três efeitos principais, três interações envolvendo duas variáveis e uma interação envolvendo todas as três variáveis. Assim, temos os seguintes efeitos: animal, treinamento, dança, e animal \times treinamento, animal \times dança, treinamento \times dança, e animal \times treinamento \times dança. Eliminação de trás para frente significa que a análise loglinear começa incluindo todos esses efeitos e, depois, pega a ordem de interação mais alta (nesse caso, a interação de três fatores animal \times treinamento \times dança) e a remove. Ela constrói um novo modelo sem essa interação e a partir do modelo calcula as frequências esperadas. Depois, compara essas frequências esperadas (ou modelo de frequências) com as frequências observadas usando a equação padrão para a estatística da razão da verossimilhança (veja a Seção 16.2.2). Se o novo modelo muda de forma significativa a estatística da razão da verossimilhança, remover esse termo de interação tem um efeito significativo na aderência do modelo e sabemos que esse efeito é estatisticamente importante. Se esse for o caso, o SPSS irá parar e dirá que você tem uma interação dos três fatores signi-

ficativa! Ele não testará outros efeitos porque com dados categóricos, todos os efeitos de baixa ordem são consumidos dentro os efeitos de alta ordem. Se, entretanto, remover a interação de três fatores não afetar de modo significativo a aderência do modelo, o SPSS continua com as interações de ordem mais baixa. Portanto, ele olha para as interações animal \times treinamento, animal \times dança e treinamento \times dança em turnos e constrói modelos em que esses termos não estão presentes. Para cada modelo, ele calcula novamente os valores esperados e os compara aos dados observados usando a razão de verossimilhança.³ Novamente, se qualquer um desses modelos resulta numa mudança significativa na razão de verossimilhança, o termo é retido e o SPSS para de olhar para qualquer efeito principal envolvido naquela interação (assim, se a interação animal \times treinamento é significativa, ele não olhará os efeitos principais de animal ou treinamento). Entretanto, se a razão de verossimilhança não mudou, a análise remove o termo de interação prejudicial e continua olhando os efeitos principais.

Mencionei que as estatísticas de verossimilhança (veja a Seção 16.2.2) são usadas para avaliar cada modelo. A partir da equação deveria ficar claro como essa equação pode ser adaptada para se ajustar à qualquer modelo: os valores observados são os mesmos todo o tempo e as frequências do modelo são simplesmente as esperadas do modelo que está sendo testado. Para o modelo saturado, essa estatística será também 0 (porque as frequências observadas e as do modelo são as mesmas, assim, a razão das frequências observadas para as do modelo será 1 e $\ln(1) = 0$), mas, como vimos, em outros casos ela dará uma medida de quão bem o modelo adere às frequências. Para testar se o modelo mudou a razão de verossimilhança, basta tomar a razão

³ Vale a pena mencionar que para cada modelo, o cálculo dos valores esperados difere, e quando o projeto fica mais complexo, o cálculo fica incrivelmente tedioso e incompreensível (pelo menos para mim); entretanto, você não precisa fazer os cálculos para saber o que está acontecendo.

de verossimilhança de um modelo e subtraí-la da estatística de verossimilhança do modelo anterior (desde que os modelos estejam hierarquicamente estruturados).

$$L\chi^2_{\text{Mudança}} = L\chi^2_{\text{Modelo Atual}} - L\chi^2_{\text{Modelo Anterior}}$$

Nesta seção, tentei lhe dar uma ideia de como a análise loglinear funciona sem entrar muito nos detalhes dos cálculos. Mostrei como podemos conceitualizar a análise do qui-quadrado como um modelo linear e, com base no que falei previamente sobre a ANOVA, espero que você possa extrapolar essas ideias conceituais e entender o que está acontecendo. Os curiosos entre vocês podem querer saber exatamente como tudo é calculado e, para essas pessoas, tenho duas coisas a dizer: “Eu não sei” e “Eu conheço um bom lugar onde você pode comprar uma camisa de força”. Se você estiver ainda muito interessado, Tabachnick e Field (2001) escreveram, como sempre, um capítulo maravilhosamente detalhado sobre o assunto, o qual humilha esta mísera seção. Contudo, assumindo que você está feliz em viver em relativa ignorância, agora veremos como fazer uma análise loglinear.

16.6 HIPÓTESES NA ANÁLISE LOGLINEAR ②

A análise loglinear é uma extensão do teste do qui-quadrado e, assim, tem hipóteses similares: isto é, um valor deve pertencer a apenas uma célula da tabela de contingência (isto é, as células da tabela devem ser independentes) e as frequências esperadas devem ser grandes o suficiente para uma análise confiável. Na análise loglinear com mais do que duas variáveis, é possível ter até 20% das células com a frequência esperada menor do que 5, mas todas as células devem ter frequências esperadas maiores do que 1. Se essa hipótese não é satisfeita, o resultado é uma redução tão radical no poder do teste que talvez seja melhor nem fazer a análise. Soluções para problemas com frequências esperadas são (1) agregue os dados de uma das variáveis (de preferência a que você menos espera ter um efeito!); (2) agregue

as categorias de uma das variáveis; (3) colete mais dados ou (4) aceite a perda de poder.

Se você quer agregar dados em uma das variáveis, alguns fatores devem ser considerados:

1. A ordem de interação mais alta não deve ser significativa.
2. Pelo menos um dos termos de interação de baixa ordem envolvendo a variável sendo agregada deve ser não-significativo.

Voltemos ao exemplo que estamos usando. Digamos que queremos eliminar a variável animal; para isso ser válido, a variável animal \times treinamento \times dança deve ser não-significativa e tampouco a interação animal \times treinamento ou animal \times dança deve ser significativa. Você pode, também, agregar categorias de uma das variáveis. Assim, se você tiver a variável “estações” com as categorias primavera, verão, outono e inverno e tem algumas observações em inverno, poderia reduzir a variável para três categorias apenas, primavera, verão, outono/inverno. Entretanto, você deve tentar agregar somente categorias que faça sentido teórico combinar.

Finalmente, algumas pessoas superam o problema simplesmente adicionando uma constante a todas as células da tabela, mas não há razão para tanto porque isso não resolve a questão do poder.

16.7 ANÁLISE LOGLINEAR UTILIZANDO O SPSS ②

16.7.1 Considerações iniciais ②

Os dados para a análise loglinear são inseridos da mesma maneira que para o teste qui-quadrado (veja as Seções 16.4.1 e 16.4.2). Os dados para o exemplo do gato e cachorro estão no arquivo **CatsandDogs.sav**. Abra esse arquivo. Note que ele tem três variáveis (**animal**, **training** e **dance**) (animal, treinamento e dança) e cada uma contém códigos representando as diferentes categorias dessas variáveis. Para começar, devemos usar o comando *crosstabs* (tabulação cruzada) a fim de obter uma tabela de contingência dos dados. Use a Seção 16.4.3

para ajudá-lo a fazer isso. A caixa de diálogo final será semelhante à da Figura 16.4.

A tabela de tabulação cruzada produzida pelo SPSS (Saída 16.5 do SPSS) contém o número de casos que se encaixam em cada combinação de categorias. A metade superior dessa tabela é a mesma da Saída 16.1 do SPSS porque os dados são os mesmos (apenas acrescentamos alguns cachorros), e se você reler o capítulo terá um resumo sobre o que isso nos informa. Para os cachorros, podemos resumir os dados de uma maneira semelhante. No total, 49 cachorros dançaram (70% do total) e desses, 20 foram treinados usando comida (40,8% do total que dançou) e 29 foram treinados com afeto (59,2% do total que dançou). Vinte e um cachorros não dançaram (30% do total) e, desses que não dançaram, 14 foram treinados usando comida como recompensa (66,7% do total que não dançou) e 7 foram treinados pelo afeto (33,3% do total que não dançou). Os números dos cachorros podem ser lidos nas linhas rotuladas de **Count** (Contagem) e as porcentagens são lidas nas linhas chamadas de **% within Did they dance?** (% Eles dançaram?). Resumindo, mais cachorros dançaram (70%) do que não dançaram (30%). Aproximadamente metade daqueles que dançaram foram

treinados com afeto e aproximadamente metade com comida como recompensa. Ou seja, os cachorros parecem estar mais propensos a dançar do que os gatos (70% comparado com 38%) e eles estão pouco preocupados com o tipo de treinamento usado.

Antes de continuarmos a olhar os testes estatísticos, é importante conferir se a hipótese da análise loglinear foi satisfeita: mais especificamente, não deverão existir frequências esperadas menores do que 1 e não mais do que 20% das células podem ter frequências esperadas menores do que 5. Se você olhar para a tabela de tabulação cruzada, deverá notar que o menor valor esperado é 10,2 (para cachorros que foram treinados com comida e não dançaram). Esse valor é maior do que 5 e, assim, a hipótese é verificada.

16.7.2 A Análise Loglinear ②

Tendo estabelecido que a hipótese foi verificada, podemos seguir com a análise principal. A melhor maneira de processar uma análise loglinear é usar o menu **Analyze**⇒**Loglinear**⇒**Statistics**⇒**Model Selection...** (Analisar⇒Loglinear⇒Estatísticas⇒Seleção do Modelo...) para acessar a cai-

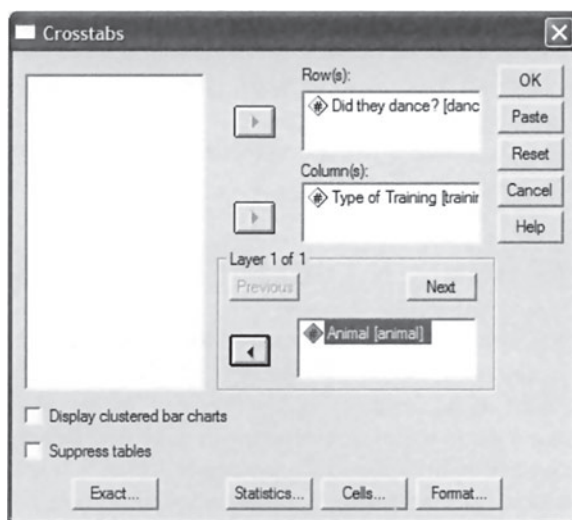




Figura 16.4 Caixa de diálogo *Crosstabs*.

Saída 16.5 do SPSS

Did They dance? * Type of Training * Animal Crosstabulation (Eles dançaram?
***Tipo de treinamento * Tabulação cruzada do Animal)**

Animal (Animal)				Type of Training (Tipo de treinamento)		Total	
				Food as Reward (Comida como recompensa)	Affection as Reward (Afeto como recompensa)		
Gato	Did they dance? (Eles dançaram?)	Yes (Sim)	Count (Contagem)	28	48	76	
			Expected Count (Contagem Esperada)	14.4	61.6	76.0	
			% within Did they dance? (% dentro de Eles dançaram?)	36.8%	63.2%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	73.7%	29.6%	38.0%	
			% of total (% do total)	14.0%	24.0%	38.0%	
	No (Não)		Count (Contagem)	10	114	124	
			Expected Count (Contagem Esperada)	23.6	100.4	124.0	
			% within Did they dance? (% dentro de Eles dançaram?)	8.1%	91.9%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	26.3%	70.4%	62.0%	
			% of total (% do total)	5.0%	57.0%	62.0%	
	Total		Count (Contagem)	38	162	200	
			Expected Count (Contagem Esperada)	38.0	162.0	200.0	
			% within Did they dance? (% dentro de Eles dançaram?)	19.0%	81.0%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	100.0%	100.0%	100.0%	
			% of total (% do total)	19.0%	81.0%	100.0%	
Cão	Did they dance? (Eles dançaram?)	Yes (Sim)	Count (Contagem)	20	29	49	
			Expected Count (Contagem Esperada)	23.8	25.2	49.0	
			% within Did they dance? (% dentro de Eles dançaram?)	40.8%	59.2%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	58.8%	80.6%	70.0%	
			% of total (% do total)	28.6%	41.4%	70.0%	
	No (Não)		Count (Contagem)	14	7	21	
			Expected Count (Contagem Esperada)	10.2	10.8	21.0	
			% within Did they dance? (% dentro de Eles dançaram?)	66.7%	33.3%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	41.2%	19.4%	30.0%	
			% of total (% do total)	20.0%	10.0%	30.0%	
	Total		Count (Contagem)	34	36	70	
			Expected Count (Contagem Esperada)	34.0	36.0	70.0	
			% within Did they dance? (% dentro de Eles dançaram?)	48.6%	51.4%	100.0%	
			% within Type of Training (% dentro do Tipo de Treinamento)	100.0%	100.0%	100.0%	
			% of total (% do total)	48.6%	51.4%	100.0%	

xa de diálogo principal como a da Figura 16.5. Marque qualquer variável que você queira incluir na análise selecionando-a com o mouse e transfira-a para a caixa **Factor(s)** (Fatores)

clicando em . Quando existe uma variável nessa caixa a tecla  (Definir Intervalo) é ativada. Da mesma forma como no teste *t* e vários dos testes não-paramétricos que discuti-

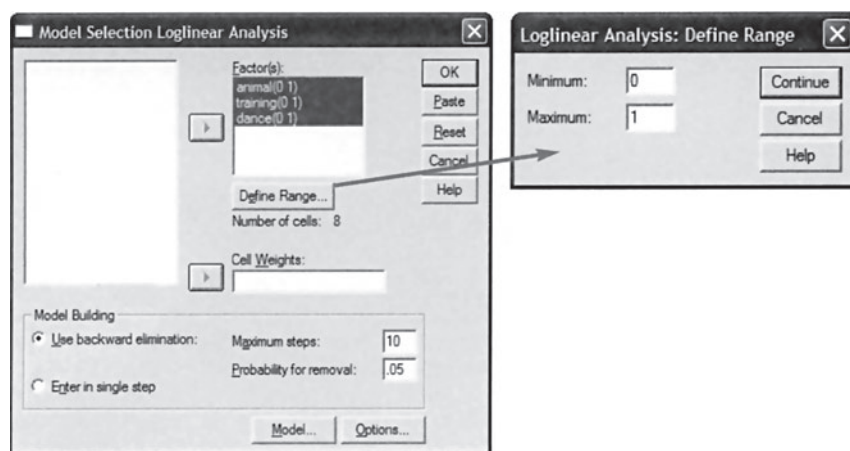


Figura 16.5 Caixa de diálogo principal da análise loglinear.

mos no Capítulo 13, precisamos dizer ao SPSS os códigos que usamos para definir nossas variáveis categóricas. Selecione uma variável na caixa **Factor(s)** (Fatores) e clique em **Define Range** (Definir Intervalo) para ativar a caixa de diálogo que permite especificar o valor dos códigos mínimo e máximo que você usou para aquela variável. Todas as três variáveis nesse exemplo têm os mesmos códigos (todas elas apresentam duas categorias e foram codificadas como 0 e 1), assim, podemos selecionar as três; clique em **Define Range** (Definir Intervalo) e digite 0 na caixa *Minimum* (Mínimo) e 1 na caixa *Maximum* (Máximo). Depois de fazer isso, clique em **Continue** (Continuar) para retornar à caixa de diálogo principal.

As opções por padrão na caixa principal são adequadas; o mais importante a ser notado é que, por omissão, o SPSS usa a eliminação para trás (*Backwards*). Você pode selecionar **Enter in a single step** (Entrar de uma única vez), que é um método não hierárquico (no qual todos os efeitos entram e são avaliados, como a entrada forçada na regressão múltipla). Na análise loglinear, os efeitos combinados tem prioridade sobre os efeitos de baixa ordem e, assim, não há razão para recomendar métodos não hierárquicos.

Se você clicar em **Model...** (Modelo), abrirá uma caixa de diálogo semelhante a uma vista

na ANOVA (veja a Figura 10.2). Por padrão, o SPSS ajusta o modelo saturado e é esse que devemos tentar. Entretanto, você pode definir seu próprio modelo, se quiser, especificando efeitos especiais principais e termos de interação. A não ser que você tenha um bom motivo para não ajustar o modelo saturado, deixe as coisas como estão!

Clicar em **Options...** (Opções) abre uma caixa de diálogo semelhante à da Figura 16.6. Existem poucas opções (as opções por omissão são boas). As únicas duas coisas que você pode selecionar são **Parameter estimates** (Estimativas dos parâmetros), que produzirá uma tabela de estimativas paramétricas para

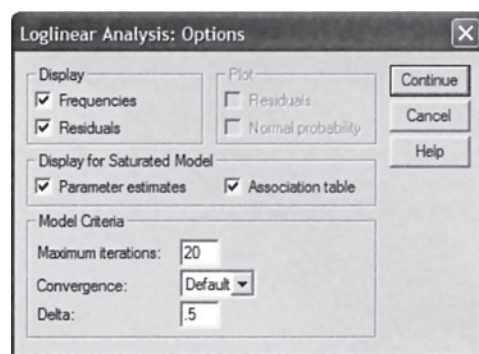


Figura 16.6 Opções para a análise Loglinear.

cada efeito (por estimativas paramétricas, quero dizer apenas o escore- z e o intervalo de confiança associado) e uma **Association Table** (Tabela de Associação), que produzirá as estatísticas qui-quadrado para todos os efeitos do modelo. Isso pode ser útil em algumas situações, mas, como escrevi anteriormente, se as interações de alta ordem são significativas, não deveríamos estar interessados nos efeitos de baixa ordem porque eles são confundidos com os efeitos de alta ordem. Quando você terminar com as opções, clique em (Continuar) para retornar à caixa de diálogo principal e clique em para executar a análise.

16.8 SAÍDAS DA ANÁLISE LOGLINEAR ③

A Saída 16.6 do SPSS mostra a saída inicial da análise loglinear. A saída nos informa que temos 270 casos (lembre que tínhamos 200 gatos e 70 cães; esse é um bom momento para verificar que nenhum caso se perdeu!). O SPSS lista todos os fatores do modelo com o número de níveis de cada um (nesse caso, todos têm dois níveis). Para começar, o SPSS ajusta o modelo saturado (todos os termos estão no modelo, incluindo a interação de alta ordem, nesse caso, a interação animal \times treinamento \times dança). O SPSS fornece a contagem observada e esperada para cada combinação de categorias do modelo. Esses valores devem ser os mesmos da tabela original de contingência, exceto que cada célula tem o valor 0,5 adicionado (esse é o valor padrão e ele é bom, mas se você quiser trocá-lo, pode fazê-lo mudando o delta como na Figura 16.6). A parte final dessa saída inicial fornece duas estatísticas de aderência (qui-quadrado de Pearson e a estatística da razão de verossimilhança, ambas vistas no começo deste capítulo). Nesse contexto, esses testes estão testando a hipótese de que as frequências previstas pelo modelo (as frequências esperadas) são significativamente diferentes das frequências reais dos nossos dados (as frequências observadas). Se o seu modelo está bem ajustado aos dados, as frequências observadas e esperadas devem ser muito parecidas

(isto é, não-significativamente diferentes). Portanto, queremos que essas estatísticas sejam não-significativas. Um resultado significativo significaria que nosso modelo foi significativamente diferente dos nossos dados (isto é, o modelo é um péssimo ajuste dos dados). Em amostras grandes, essas estatísticas deveriam dar os mesmos resultados, mas a estatística da razão de verossimilhança deve ser preferida em amostras pequenas. Nesse exemplo, ambas as estatísticas são 0 e fornecem um valor de probabilidade p , *of-INF*, que é uma maneira confusa de dizer que a probabilidade é muito alta. Em outras palavras, nesse estágio o modelo prevê perfeitamente os dados. Se você leu a seção teórica, isso não deveria surpreendê-lo, porque lá mostrei que o modelo saturado é um ajuste perfeito dos dados e mencionei que a razão de verossimilhança resultante seria zero. O interessante na análise loglinear é que podemos remover partes do modelo sem afetar de modo significativo o ajuste do modelo.

A próxima parte da saída (Saída 16.7 do SPSS) informa quais componentes do modelo podem ser removidos. A primeira parte da saída é rotulada de **Tests that K-way and higher order effects are zero** (Testa se o fator de ordem k ou maior é zero) e abaixo há uma tabela mostrando a razão de verossimilhança e a estatística qui-quadrado quando $K = 3, 2$ e 1 (descendo pelas linhas da tabela). A primeira linha ($K = 3$) testa se remover o efeito de três fatores e os efeitos de alta ordem irá afetar de modo significativo o ajuste do modelo. Agora, é claro, a interação de três fatores é o efeito de alta ordem que temos, assim, simplesmente testamos se a remoção da interação de três fatores (isto é, a interação animal \times treinamento \times dança) irá afetar significativamente o ajuste do nosso modelo. Se você olhar as duas colunas rotuladas de **Prob**, poderá ver que ambos os testes, qui-quadrado e a razão de verossimilhança, concordam que remover essa interação afetará de forma significativa o ajuste do modelo (porque a probabilidade é menor do que 0,05). A próxima linha da tabela ($K = 2$) nos diz se remover as interações de dois fatores (isto é, as interações animal \times

Saída 16.6 do SPSS

***** HIERARCHICAL LOG LINEAR *****					
(LOG LINEAR HIERÁRQUICO)					
DATA information (Informação dos dados)					
270 unweighted cases accepted (270 casos não ponderados aceitos).					
0 cases rejected because of out-of-range factor values					
(0 casos rejeitados porque estão fora do intervalo de valores)					
0 cases rejected because of missing data (0 casos rejeitados por dados omitidos).					
270 weighted cases will be used in the analysis (270 casos ponderados serão usados na análise).					
FACTOR information (Informação dos Fatores)					
Factor (Fator)	Level (Nível)	Label (Rótulo)			
ANIMAL (ANIMAL)	2	Animal (Animal)			
TRAINING (TREINAMENTO)	2	Type of Training (Tipo de Treinamento)			
DANCE (DANÇA)	2	Did they dance? (Eles dançaram?)			

DESIGN 1 has generating class (O DELINEAMENTO 1 tem classe geradora)					
ANIMAL*TRAINING*DANCE (ANIMAL*TREINAMENTO*DANÇA)					
Note: For saturated models .500 has been added to all observed cells.					
(Nota: Para modelos saturados, 0,500 foi acrescentado a todas as células observadas).					
The value may be changed by using the CRITERIA - DELTA subcommand.					
(Esse valor pode ser trocado usando o subcomando CRITÉRIO = DELTA)					
The Iterative Proportional Fit algorithm converged at iteration 1.					
(O algoritmo iterativo de aderência proporcional convergiu na interação 1).					
The maximum difference between observed and fitted marginal totals is .000 and					
the convergence criterion is .250 (A diferença máxima entre os totais marginais					
observados e ajustados é 0,000 e o critério de convergência é 0,250).					

Observed, Expected Frequencies and Residuals (Frequências observadas, esperadas e					
resíduos).					
Factor	Code	OBS Count	EXP Count	Residual	Std Resid
(Fator)	(Código)	(Contagem OBS)	(Contagem ESP)	(Resíduos)	(Resíduos PAD)
ANIMAL cat (gato)					
TRAINING Food As					
(TREINAMENTO) (Com Comida)					
DANCE (Dança)	Yes (Sim)	28.5	28.5	0.00	0.00
DANCE (Dança)	No (Não)	10.5	10.5	0.00	0.00
TRAINING Affectio					
(TREINAMENTO) (Com Afeto)					
DANCE	Yes (Sim)	48.5	48.5	0.00	0.00
DANCE	No (Não)	114.5	114.5	0.00	0.00
ANIMAL Dog (Cão)					
TRAINING Food As					
(TREINAMENTO) (Com Comida)					
DANCE (Dança)	Yes (Sim)	20.5	20.5	0.00	0.00
DANCE (Dança)	No (Não)	14.5	14.5	0.00	0.00
TRAINING Affectio					
(TREINAMENTO) (Com Afeto)					
DANCE	Yes (Sim)	29.5	29.5	0.00	0.00
DANCE	No (Não)	7.5	7.5	0.00	0.00

Goodness-of-fit test Statistics (Estatísticas de Aderência do Teste)					
Likelihood ratio chi square =		0.00000	DF = 0	P = -INF	
(Qui-quadrado Razão de verossimilhança =		0.00000	GL = 0	P = - INF)	
Pearson chi square =		0.00000	DF = 0	P = -INF	
(Qui-quadrado de Pearson =		0.00000	GL = 0	P = -INF)	

Saída 16.7 do SPSS

Tests that K-way and higher order effects are zero.						
(Testa se os efeitos do fator k ou de ordem mais alta são zero).						
K	DF	L. R. Chisq	Prob	Pearson Chisq	Prob	Iteration
(K)	(GL)	(χ^2 Razão de Ver.)	(Probab.)	(χ^2 de Pearson)	(Probab.)	(Iteração)
3	1	20.305	0.0000	20.778	0.0000	4
2	4	72.267	0.0000	67.174	0.0000	2
1	7	200.163	0.0000	253.556	0.0000	0

Tests that K-way effects are zero (Testa se os efeitos de ordem k são zero).						
K	DF	L. R. Chisq	Prob	Pearson Chisq	Prob	Iteration
(K)	(GL)	(χ^2 Razão de Ver.)	(Probab.)	(χ^2 de Pearson)	(Probab.)	(Iteração)
3	3	127.896	0.0000	186.382	0.0000	0
2	3	51.962	0.0000	46.396	0.0000	0
1	1	20.305	0.0000	20.778	0.0000	0

treinamento, animal \times dança e treinamento \times dança) e qualquer efeito de alta ordem irá afetar o modelo. Nesse caso, existe um efeito de alta ordem (a interação de três fatores), assim, isso testa se a remoção das interações de dois fatores e da interação de três fatores irá afetar o ajuste do modelo. Isso é altamente significativo (o que não deveria surpreendê-lo porque o teste contém a interação de três fatores que, como já vimos, é altamente significativa) indicando que se removermos as interações de dois fatores e a interação de três fatores, isso terá um efeito detrimental significativo no modelo. A linha final ($K = 1$) nos diz que remover os efeitos de um fator (isto é, os efeitos principais de animal, treinamento e dança) e efeitos de alta ordem irá significativamente afetar o ajuste do modelo. Existem muitos efeitos de alta ordem aqui – existem interações de dois fatores e interação de três fatores – e, assim, isso é basicamente testar se remover tudo do modelo terá um efeito significativo no ajuste do modelo. Isso, é altamente significativo (o que deveríamos esperar porque, como já vimos, a interação de três efeitos é altamente significativa e esse teste inclui aquela interação). Se esse teste não fosse significativo (se os valores de **Prob** estivessem acima de 0,05), isso lhe informaria que remover tudo do seu modelo não afetaria o ajuste do modelo (em

outras palavras, o efeito combinado de suas variáveis e as interações não seriam significativas).

A próxima parte da tabela expressa a mesma coisa, mas não inclui os efeitos de alta ordem. Ela é rotulada de **Tests that K-way effects are zero** (Testa se os efeitos de ordem K são zero), e lista testes onde $K = 1, 2$ e 3 . A primeira linha ($K = 1$), portanto, testa se ao remover os efeitos principais (os efeitos de um fator) ocorre um efeito detrimental significativo no modelo. A probabilidade é menor do que 0,05, indicando que se removermos o efeito principal de animal, treinamento e dança do nosso modelo, isso irá afetar significativamente o ajuste do modelo (em outras palavras, esses efeitos são precursores significativos dos dados). A segunda linha ($K = 2$) testa se ao remover as interações de dois fatores há um efeito detrimental significativo no modelo. A probabilidade é menor do que 0,05, indicando que se removermos as interações animal \times treinamento, animal \times dança e treinamento \times dança, isso irá reduzir significativamente o ajuste do modelo aos dados. A linha final ($K = 3$) testa se ao remover a interação de três fatores há um efeito detrimental significativo no modelo. A probabilidade é menor do que 0,05, indicando que se removermos a interação animal \times treinamento \times dança, o ajuste

dos dados ao modelo irá reduzir significativamente. Em outras palavras, essa interação de três fatores é um preditor significativo dos dados. Essa linha deve ser idêntica à primeira linha da tabela anterior (Testes onde efeitos de ordem K são zero) porque ela é um efeito de alta ordem e, assim, na tabela anterior não havia efeitos de alta ordem para incluir no teste (veja a saída e você perceberá que os resultados são idênticos).

Isso nos mostra que a interação de três fatores é significativa: removê-la do modelo tem um efeito significativo no ajuste do modelo aos dados. Também sabemos que remover todas as interações de dois fatores tem um efeito significativo no modelo, mas você deve lembrar que a análise loglinear deve ser feita hierarquicamente. Assim, essas interações de dois fatores não são do nosso interesse porque a interação de três fatores é significativa (somente olharíamos para esses efeitos se a interação de três fatores fosse não-significativa).

Se você selecionou *Association table* (Tabela de associação) como na Figura 16.6, terá a tabela de Saída 16.8 do SPSS. Isso simplesmente desmembra a tabela que acabamos de olhar em suas partes componentes. Assim, por exemplo, embora saibamos da saída anterior que remover todas as interações de dois fatores afeta significativamente o modelo, não sabemos qual das interações de dois fatores está ocasionando o efeito. Essa tabela nos diz. Temos o teste qui-quadrado de Pearson para

cada interação de dois fatores e para os efeitos principais, e a coluna rotulada de **Prob** informa qual desses efeitos é significativo (valores menores do que 0,05 são significativos). Podemos afirmar disso que as interações animal \times dança, treinamento \times dança e animal \times treinamento são todas significativas. Da mesma forma, vimos na saída anterior que a remoção dos efeitos de um fator (os efeitos principais de animal, treinamento e dança) afetou significativamente o ajuste do modelo e essas descobertas estão confirmadas aqui porque os efeitos principais de animal e treinamento são significativos. Entretanto, o efeito principal de dança não é (a probabilidade é maior do que 0,05). Mesmo que essas descobertas sejam interessantes, devemos ignorá-las devido à natureza hierárquica da análise loglinear: esses efeitos são confundidos com a interação de alta ordem de animal \times treinamento \times dança.

Se você selecionou *Parameter estimates* (Estimativas dos parâmetros) conforme a Figura 16.6, terá a tabela da Saída 16.9 do SPSS. Ela informa o mesmo da tabela anterior (isto é, ela fornece estimativas individuais para cada efeito), mas faz isso usando o escore-*z* em vez do teste qui-quadrado. Isso pode ser útil porque temos intervalos de confiança e, também, porque o valor de *z* fornece uma comparação útil entre os efeitos (se você ignorar o sinal de mais ou menos, quanto maior o *z*, mais significativo o efeito é). Assim, se você olhar os valores de *z* poderá ver que o efeito principal

Saída 16.8 do SPSS

* * * * * HIERARCHICAL LOG LINEAR (LOG LINEAR HIERÁRQUICO) * * * * *				
Tests of PARTIAL association (Testes de associação PARCIAL)				
Effect Name (Nome do Efeito)	DF (GL)	Partial Chisq (χ^2 Parcial)	Prob	Iter (Iteração)
ANIMAL*TRAINING (ANIMAL*TREINAMENTO)	1	13.760	0.0002	2
ANIMAL*DANCE (ANIMAL*DANÇA)	1	13.748	0.0002	2
TRAINING*DANCE (TREINAMENTO*DANÇA)	1	8.611	0.0033	2
ANIMAL (ANIMAL)	1	65.268	0.0000	2
TRAINING (TREINAMENTO)	1	61.145	0.0000	2
DANCE (DANÇA)	1	1.483	0.2233	2

Saída do SPSS 16.9

Note: For saturated models .500 has been added to all observed cells.
(Nota: Para modelos saturados, 0,500 foi somado a todas as células observadas).

This value may be changed by using the CRITERIA - DELTA subcommand.
(Esse valor pode ser trocado usando o sub-comando CRITÉRIO = DELTA)

Estimates for Parameters (Estimativas para os Parâmetros)

ANIMAL*TRAINING*DANCE (ANIMAL*TREINAMENTO*DANÇA)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	0.3600938577	0.08335	4.32010	0.19672	0.52347

ANIMAL*TRAINING (ANIMAL*TREINAMENTO)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	0.4020174410	0.08335	-4.82306	-0.56539	-0.23865

ANIMAL*DANCE (ANIMAL*DANÇA)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	-0.1970307093	0.08335	-2.36381	-0.36040	-0.03366

TRAINING*DANCE (TREINAMENTO*DANÇA)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	0.1042911061	0.08335	1.25120	-0.05908	0.26766

ANIMAL (ANIMAL)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	0.4036938932	0.08335	4.84318	0.20032	0.56707

TRAINING (TREINAMENTO)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	-0.3281973781	0.08335	-3.93743	-0.49157	-0.16483

DANCE (DANÇA)

Parameter (Parâmetro)	Coeff. (Coeficiente)	Std. Err. (Erro Padrão)	Z-Value (Valor de Z)	95 CI Lower (Lim. Inf. do IC de 95%)	95 CI Upper (Lim. Sup. Do IC de 95%)
1	0.2319101606	0.08335	2.78226	0.06854	0.39528

de animal é o efeito mais importante do modelo ($z = 4,84$), seguido pela interação animal \times treinamento ($z = -4,82$) e, depois, pela interação animal \times treinamento \times dança ($z = 4,32$), e assim por diante. Entretanto, vale a pena reiterar que, nesse caso, não precisamos

nos preocupar com nada além da interação de três fatores.

A parte final da saída (Saída 16.10 do SPSS) trata da eliminação para trás. O SPSS começará com o efeito de mais alta ordem (nesse caso, a interação animal \times treinamento \times dança) tiran-

Saída 16.10 do SPSS

***** HIERARCHICAL LOG LINEAR (LOG LINEAR HIERÁRQUICO) *****					
Backward Elimination (p = .050) for DESIGN 1 with generating class					
(Eliminação para trás (p = 0.050) para o DELINEAMENTO 1 com classe gerada)					
ANIMAL*TRAINING*DANCE (ANIMAL*TREINAMENTO*DANÇA)					
Likelihood Ratio Chi Square = 0.00000		DF = 0	P = -INF		
(χ² da razão da verossimilhança = 0.00000		GL = 0	P = -INF)		

If Deleted Simple Effect is	DF	L.R. Chisq Change	Prob.	Iter.	
(Se o efeito simples eliminado é)	(GL)	(χ² da razão da	(Probab.)	(Iteração)	
		verossimilhança)			
ANIMAL*TRAINING*DANCE	1	20.305	0.0000	4	
(ANIMAL*TREINAMENTO*DANÇA)					
Step 1 (Passo 1)					
The Best model has generating class (O melhor modelo tem classe geradora)					
ANIMAL*TRAINING*DANCE (ANIMAL*TREINAMENTO*DANÇA)					
Likelihood Ratio Chi Square = 0.00000		DF = 0	P = -INF		
(χ² da razão da verossimilhança = 0.00000		GL = 0	P = -INF)		

The Final model has generating class (O modelo final tem classe geradora)					
ANIMAL*TRAINING*DANCE (ANIMAL*TREINAMENTO*DANÇA)					
The Iterative Proportional Fit algorithm converged at iteration 0.					
(O algoritmo iterativo de aderência proporcional convergiu na iteração 1).					
The maximum difference between observed and fitted marginal totals is .000 and					
the convergence criterion is .250 (A diferença máxima entre os totais marginais					
observados e ajustados é 0,000 e o critério de convergência é 0,250).					

Observed, Expected Frequencies and Residuals (Frequências observadas, esperadas e					
resíduos).					
Factor	Code	OBS Count	EXP Count	Residual	Std Resid
(Fator)	(Código)	(Contagem OBS)	(Contagem ESP)	(Resíduos)	(Resíduos PAD)
ANIMAL cat (gato)					
TRAINING Food As					
(TREINAMENTO) (Com Comida)					
DANCE (Dança)	Yes (Sim)	28.0	28.0	0.00	0.00
DANCE (Dança)	No (Não)	10.0	10.0	0.00	0.00
TRAINING Affection					
(TREINAMENTO) (Com Afeto)					
DANCE	Yes (Sim)	48.0	48.0	0.00	0.00
DANCE	No (Não)	114.0	114.0	0.00	0.00
ANIMAL Dog (Cão)					
TRAINING Food As					
(TREINAMENTO) (Com Comida)					
DANCE (Dança)	Yes (Sim)	20.0	20.0	0.00	0.00
DANCE (Dança)	No (Não)	14.0	14.0	0.00	0.00
TRAINING Affection					
(TREINAMENTO) (Com Afeto)					
DANCE	Yes (Sim)	29.0	29.0	0.00	0.00
DANCE	No (Não)	7.0	7.0	0.00	0.00

Goodness-of-fit test Statistics (Estatísticas de Aderência do Teste)					
Likelihood ratio chi square =		0.00000	DF = 0	P = -INF	
(Qui-quadrado Razão de verossimilhança =		0.00000	GL = 0	P = - INF)	
Pearson chi square =		0.00000	DF = 0	P = -INF	
(Qui-quadrado de Pearson =		0.00000	GL = 0	P = -INF)	

do-o do modelo e verificando que efeito isso acarreta. Se isso não tiver um efeito significativo, ele vai para os efeitos mais altos seguintes (nesse caso, a interação de dois fatores). Entretanto, já vimos que a remoção da interação de três fatores terá um efeito significativo e isso é confirmado, nesse estágio, pela tabela **If Deleted Simple Effect is** (Se o efeito simples eliminado é), a qual confirma que a remoção da interação de três fatores tem um efeito significativo no modelo. Portanto, a análise acaba aqui: a interação de três fatores não é removida e o SPSS avalia esse modelo final. A parte da saída rotulada de **The Best model has generating class** (O melhor modelo tem uma classe geradora) informa que o modelo que melhor se ajusta aos dados inclui a interação de três fatores e, por causa da natureza hierárquica da análise, esse é o único efeito que precisamos interpretar. Finalmente, o SPSS avalia esse modelo final com a estatística da razão de verossimilhança, e estamos buscando um teste estatístico não-significativo que indique que os valores esperados gerados pelo modelo não são significativamente diferentes dos dados observados (colocando de outra maneira, o modelo se ajusta bem aos dados). Nesse caso, o resultado é não-significativo, indicando que o modelo se ajusta bem aos dados.⁴

Não preciso da análise loglinear para me dizer que os gatos são muito superiores aos cães!



O próximo passo é tentar interpretar essa interação. Primeiro devemos representar as frequências de todas as categorias. Você deve apresentar as frequências em termos de percentuais do total (esses valores podem ser en-

contrados na tabulação cruzada apresentada na Saída 16.5 do SPSS nas linhas rotuladas de %

of Total (% do total)). O gráfico resultante é mostrado na Figura 16.7, e revela o que já sabemos sobre gatos: eles irão dançar (ou fazer qualquer coisa) quando há comida envolvida, mas se você treiná-los com afeto eles não estão interessados. Cães, por outro lado, dançarão quando ganharem afeto (na verdade, mais cães dançaram do que não dançaram independentemente do tipo de recompensa, mas o efeito é mais pronunciado quando o afeto foi utilizado como método de treinamento). Os dois tipos de animais mostram respostas similares ao treinamento com comida, mas os gatos nada fazem por afeto. Assim, gatos são criaturas sensíveis que fazem coisas estúpidas somente quando há uma boa recompensa para eles (isto é, comida), já os cães são simplesmente estúpidos!

16.9 CONTINUAÇÃO DA ANÁLISE LOGLINEAR ②

Uma maneira alternativa de interpretar uma interação de três fatores é conduzir uma análise qui-quadrado em diferentes níveis de uma das variáveis. Por exemplo, para interpretar nossa interação animal \times treinamento \times dança, podemos executar um teste qui-quadrado em treinamento e dança, mas fazê-lo separadamente para cães e gatos (a análise para gatos será a mesma que do exemplo usado para o qui-quadrado). Você pode, então, comparar os resultados nos diferentes animais. Tente fazer isso usando o comando de divisão de arquivo (veja a Seção 3.4). Os resultados e a interpretação para gatos estão na Saída 16.2 do SPSS e para cães os resultados estão apresentados na Saída 16.11 do SPSS. Para cães existe ainda um relacionamento significativo entre tipos de treinamento e se dançaram ou não, mas não é tão forte (o qui-quadrado é 3,93 comparado com 25,2 dos gatos)⁵. Isso se reflete no fato de que cães tem uma probabilidade maior de dançar se ganharem afeto do que comida: o oposto dos gatos!

⁴ O fato de que a análise parou aqui não é útil porque não posso mostrar como ela procederia no caso de uma interação não-significativa de três fatores. Entretanto, ela realmente simplifica as coisas e se você estiver interessado em explorar mais a análise loglinear, as tarefas no final do capítulo e as respostas disponíveis no site www.artmed.com.br mostram o que acontece quando a interação de mais alta ordem é não-significativa.

⁵ A estatística qui-quadrado depende do tamanho da amostra, desse modo, você realmente precisa calcular os tamanhos dos efeitos e compará-los para fazer esse tipo de afirmação (a não ser que você tenha um número igual de cães e gatos!).

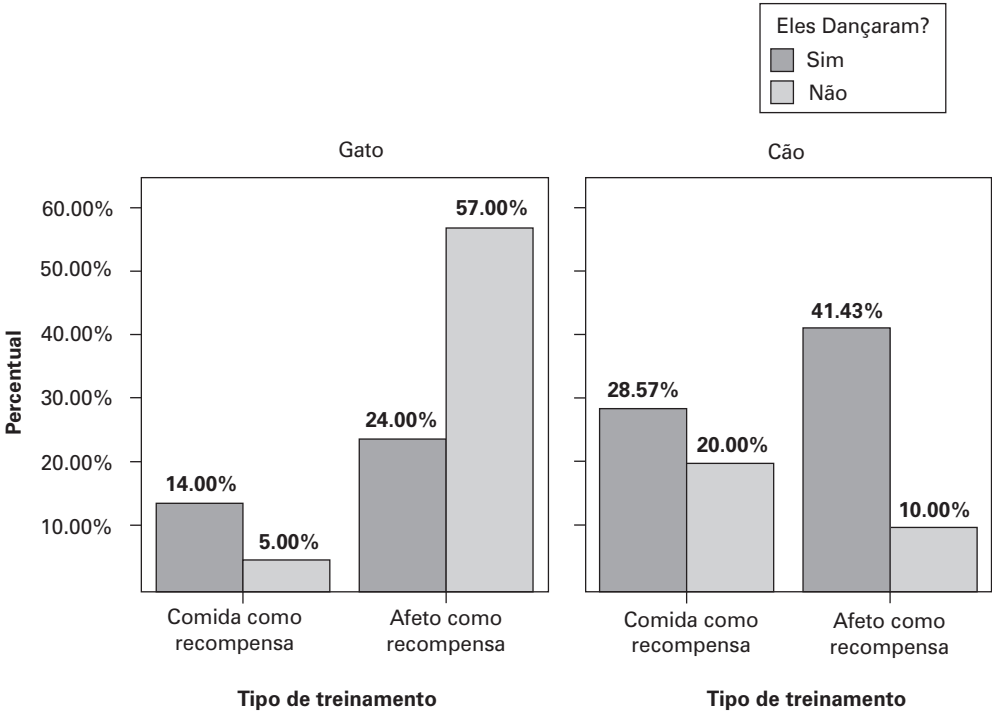


Figura 16.7 Categorias expressas em percentual.

Saída 16.11 do SPSS

Chi-Square Tests^c (Testes Qui-quadrado)

	Value (Valor)	df (gl)	Asymp. Sig. (Sig. Assintótica)	Exact Sig. (2-sided) (Sig. Exata – Bilateral)	Exact Sig. (1-sided) (Sig. Exata Unilateral)
Pearson Chi-Square (Qui-quadrado de Pearson)	3.932 ^b	1	0.047		
Continuity Correction ^a (Correção de continuidade)	2.966	1	0.085		
Likelihood Ratio (Razão de verossimilhança)	3.984	1	0.046		
Fisher's Exact Test (Teste Exato de Fisher)				0.068	0.042
Linear-by-Linear Association (Associação Linear-por-Linear)	3.876	1	0.49		
N of Valid Cases (Casos Válidos)	70				

a Calculado somente para tabelas 2 × 2.
b. 0 células (0%) tem valor esperado menor do que 5. A contagem mínima esperada é 10,20.
c. Animal = Cão.

16.10 TAMANHOS DE EFEITO NA ANÁLISE LOGLINEAR ③

Como com o qui-quadrado de Pearson, uma das maneiras mais elegantes de relatar seus efeitos é em termos de razão de chan-

ces. É mais fácil entender a razão de chances para tabelas de contingência 2 × 2 e, assim, se você tiver interações significativas de alta ordem ou suas variáveis tem mais do que duas categorias, tente desmembrar esses efeitos em tabelas lógicas 2 × 2 e calcular a razão

de chances que reflete a natureza da interação. Assim, para esse exemplo podemos calcular a razão de chances para cães e gatos separadamente. Já temos a razão de chances para gatos (Seção 16.4.5) e para cães, temos:

$$\begin{aligned}\text{Chances}_{\text{Dançar após comer}} &= \frac{\text{número dos que comeram e dançaram}}{\text{número dos que comeram mas não dançaram}} \\ &= \frac{20}{14} \\ &= 1,43 \\ \text{Chances}_{\text{Dançar após afeto}} &= \frac{\text{número dos que ganharam afeto e dançaram}}{\text{número dos que ganharam afeto mas não dançaram}} \\ &= \frac{29}{7} \\ &= 4,14 \\ \text{Razão de chances} &= \frac{\text{Chances}_{\text{Dançar após comer}}}{\text{Chances}_{\text{Dançar após afeto}}} \\ &= \frac{1,43}{4,14} \\ &= 0,35\end{aligned}$$

Isso nos diz que se um cão foi treinado com comida ele teve 0,35 mais probabilidade de dançar (isto é, ele teve menos probabilidade de dançar). Outra maneira de dizer isso é que foi $1/0,35 = 2,90$ vezes menos provável de dançar. Compare isso com gatos que tiveram 6,65 mais probabilidades de dançar. Como você pode ver, comparar a razão das chances entre cães e gatos é uma maneira elegante de apresentar o termo de interação de três fatores no modelo.

16.11 RELATANDO OS RESULTADOS DA ANÁLISE LOGLINEAR ②

Quando relatar a análise linear, você precisa informar a estatística da razão de verossi-

milhança para o modelo final, geralmente indicado por χ^2 . Para qualquer termo significativo você deve relatar a mudança do qui-quadrado ou você deve relatar o escore-z para o efeito e seu intervalo de confiança associado. Se você precisar desmembrar uma interação de alta ordem em análises subsequentes, obviamente você precisa relatar a estatística qui-quadrado relevante (e a razão de chances). Para esse exemplo podemos relatar:

- ✓ A análise loglinear de três fatores produziu um modelo final que retém todos os efeitos. A razão de verossimilhança desse modelo foi $\chi^2(0) = 0$, $p = 1$. Isso indicou que a interação de mais alta ordem (a interação animal \times treinamento \times dança) foi significativa $\chi^2(1) = 20,31$ $p < 0,001$. Para desmembrar esse efeito, separe os testes do qui-quadrado das variáveis treinamento e dança que foram realizadas separadamente para cães e gatos. Para gatos, houve uma associação significativa entre o tipo de treinamento e se os gatos dançaram ou não, $\chi^2(1) = 25,36$ $p < 0,001$; isso foi verdadeiro para cães também, $\chi^2(1) = 3,93$ $p < 0,05$. A razão das chances indicou que gatos tinham 6,65 mais chances de dançar se treinados com comida do que treinados com afeto, mas cães tinham somente 0,35 mais chances de dançar (ou 2,90 vezes menos probabilidade de dançar). Portanto, a análise revela uma diferença fundamental entre cães e gatos: gatos têm uma probabilidade maior de dançar por comida do que por afeto e cães têm uma probabilidade maior de dançar por afeto do que por comida.

16.12 O QUE DESCOBRIMOS SOBRE ESTATÍSTICA? ①

Na primeira edição deste livro, queria escrever um capítulo sobre análise loglinear, mas já havia escrito 300 páginas a mais do que o combinado e tinha despendido tanto esforço no restante do livro que pensar em mais um capítulo era impossível. Desta vez, decidi que escreveria um capítulo sobre a análise loglinear. Entretanto, quando chegou o momento,

Dica da Samanta Ferrinho



- Se você quer testar o relacionamento entre mais do que duas variáveis categóricas, pode fazer isso com a *análise loglinear*.
- A análise loglinear é hierárquica: o modelo inicial contém todos os efeitos principais e interações. Iniciando com a interação de alta ordem, os termos são removidos para ver se sua remoção afeta de modo significativo o ajuste do modelo. Se ela afeta, esse termo não é removido e todos os efeitos de baixa ordem são ignorados.
- Olhe para a tabela **The Best model has generating class** (O melhor modelo tem uma classe geradora) para ver quais efeitos ficaram retidos no modelo final. Consulte a tabela **If Deleted Simple Effect is** (Se o efeito simples eliminado é) para ver a significância individual dos efeitos retidos (olhe para a coluna **Prob**: valores menores do que 0,05 indicam significância).
- Olhe para as estatísticas teste de aderência para o modelo final: se o modelo se ajusta bem aos dados, essa estatística deve ser não-significativa (Prob. deve ser maior do que 0,05).
- Olhe para a tabulação cruzada a fim de interpretar qualquer efeito significativo (% do total das células é o melhor fator para consultar).

havia escrito 200 páginas a mais do que deveria para essa nova edição e, com o prazo encurtando, a história estava se repetindo. Portanto, você não sabe como estou muito feliz de ter finalmente escrito esta droga de capítulo! Este capítulo forneceu uma breve visão sobre a análise de dados categóricos. Mostrei que abordamos dados categóricos da mesma maneira que qualquer outro tipo de dados: ajustamos o modelo, calculamos o desvio entre nosso modelo e os dados observados e usamos isso para avaliar o modelo que ajustamos. Também tentei ensinar que o modelo que ajustamos é o mesmo que vimos ao longo de todo o livro: o modelo linear (regressão). Quando temos somente duas variáveis, podemos usar o teste qui-quadrado de Pearson ou o teste da razão de verossimilhança para ver se as duas variáveis estão associadas. Em situações mais complexas, simplesmente estendemos esses modelos para algo conhecido como modelo loglinear. Isso é parecido com a ANOVA para dados categóricos: para cada variável que temos, obtemos um efeito principal, mas também obtemos interações entre variáveis. A análise loglinear avalia todos esses efeitos hierarquicamente para nos dizer quais deles preveem melhor nossa saída.

16.13 TERMOS-CHAVE QUE DESCOBRIMOS

- Variável Categórica
- Teste qui-quadrado
- Tabela de contingência
- O V de Cramer
- O λ de Goodman e Kruskal
- Análise Loglinear
- Razão de chances
- Fi
- Modelo saturado
- Correção de continuidade de Yates

16.14 TAREFAS DO ALEX ESPERTO



- **Tarefa 1:** Alguns editores das Sage Publications acham que sabem tudo sobre futebol. Para ver se eles são melhores do que os professores de Sussex e alunos da pós-graduação, convidamos vários funcionários da Sage para participarem dos nossos jogos de futebol (quero dizer, nós os convidamos para importantes reuniões sobre livros). Cada jogador poderia jogar somente uma partida. Ao longo das partidas, contamos o número de jogadores que mar-

caram gols. Os dados estão no arquivo **SageEditorsCan'tPlayFootball.sav**; faça o teste qui-quadrado para ver quem marcou mais gols: editores ou acadêmicos. Nossa hipótese é que o time de Sussex irá marcar mais do que o da Sage. ③

- **Tarefa 2:** Queria verificar se o horóscopo é somente uma invenção da cabeça das pessoas. Portanto, peguei 2201 pessoas, anotei seus signos (essa variável, obviamente, tem 12 categorias: Capricórnio, Aquário, Peixes, Áries, Touro, Gêmeos, Câncer, Leão, Virgem, Libra, Escorpião e Sagitário) e se elas acreditavam em horóscopos (essa variável tem duas categorias: crente ou não-crente). Depois, enviei um horóscopo para elas com a informação do que iria acontecer durante o próximo mês: todas, independentemente do seu signo, receberam o mesmo horóscopo, que dizia: “Agosto é um ótimo mês para você. Você irá fazer amizade com um vagabundo na primeira semana do mês e fará uma omelete de queijo para ele. A curiosidade é sua grande virtude e, na segunda semana, você irá se interessar por um assunto que antes julgava entediante, estatística, talvez. Talvez você compre um livro nessa mesma semana que o guiará em direção a essa sabedoria. Seu novo conhecimento o fará mudar de carreira por volta

da terceira semana, quando você largará o seu emprego e se tornará contador. Na última semana, você não terá mais amigos, sua namorada(o) te abandonará por um(a) dançarino(a) russo(a) de balé com um olho de vidro e você passará seus finais de semana fazendo análise loglinear à mão com a companhia de um pombo chamado Hephzibah”. No final de agosto, entrevistei todas as pessoas em relação a como suas vidas combinaram com o horóscopo fictício e classifiquei o horóscopo como verdadeiro ou falso. Os dados estão no arquivo **Horoscope.sav**. Realize uma análise loglinear para ver se existe um relacionamento entre o signo da pessoa, se eles acreditam em horóscopo e se o horóscopo foi verdadeiro.

As respostas podem ser encontradas no site www.artmed.com.br no arquivo **Answers(Chapter16).pdf**.

16.15 LEITURAS COMPLEMENTARES

- HUTCHESON, G., SOFRONIOU, N. *The multivariate social scientist*. London: Sage, 1999.
- TABACHNICK, B. G., FIDELL, L. S. *Using multivariate statistics*. Boston: Allyn e Bacon. 2001. 4ª ed. O Capítulo 7 sobre Análise Loglinear é muito bom.

17

EPÍLOGO

*Eis algumas perguntas que o autor enviou
Um observador pode ser um participante?
Eu tenho visto demais?
É válido se não posso tocar?
Se a visão é tudo o que eu posso assegurar
O entendimento está fora do alcance.*

(Fugazi, 2001)

Então, você chegou ao final do livro, hein? Se você leu o livro de cabo a rabo, investiu grande parte da sua vida (possivelmente toda a sua vida, a não ser que você seja um leitor rápido) neste livro e eu estou lisonjeado. Você deve ter aprendido algo sobre estatística (se não, eu certamente aprendi muito escrevendo o livro), mas o mais importante é que tivemos nossa parte justa de sexo, drogas e *rock'n'roll* (em exemplos, obviamente).

Um dos grandes problemas em ser apaixonado por lecionar é que você constantemente se dá conta de como é um péssimo professor. O dia em que eu obtive a cópia da primeira edição deste livro eu estava muito entusiasmado: o abri para admirar meu trabalho e a primeira coisa que encontrei foi um erro! Era um erro pequeno, mas desde então tenho encontrado imperfeições irritantes. Assim, quando os editores decidiram que queriam uma nova versão eu pensei “ótimo”, agora posso escrevê-lo corretamente e acrescentar todas as coisas que as pessoas me perguntam. Obviamente, fiquei obsessivo com isso e, quase 700 páginas depois, espero nunca mais ter que escrever outra palavra sobre estatística novamente! Assim, se você enviasse o se-

guinte modelo de carta a Michael Carmichael, a/c Sage Publications, eu ficaria muito grato.

Prezado Sr. Carmichael

Estou escrevendo para dizer que comprei o livro de Andy Field, *Descobrimo Estatística Usando o SPSS*, e achei-o perfeito. Se você for publicar uma terceira edição, peço que não mude uma única palavra, apenas crie uma nova capa e pague alguém, que não o Andy, para atualizar o SPSS. Mudar uma palavra, mesmo que pequena, seria um crime hediondo e iria colocar em risco todas as futuras vendas deste livro. Não é justo, também, que um homem tão legal como o Andy tenha que trabalhar para que os amigos dos editores possam passear por Londres em limusines, bebendo champanhe e cheirando coca; você deveria dar mais dinheiro para ele e pagar por umas férias bem longas, as quais ele, com certeza, merece após todo esse trabalho árduo.

Atenciosamente,

(coloque seu nome aqui)
Cliente valorizado

GLOSSÁRIO

Este glossário contém definições breves de todos os termos-chave do livro. Palavras em *itálico* referem-se a itens que estão definidos em outros locais do glossário

-2LL: a *verossimilhança log* multiplicada por menos 2. Essa versão da verossimilhança é utilizada na regressão logística.

Aderência: um índice de quão bem um modelo se ajusta aos dados a partir do qual ele foi gerado. É normalmente baseado em quão bem os dados previstos pelo modelo correspondem aos dados que foram de fato coletados.

AFD: um acrônimo para Análise da Função Discriminante (ver *Análise Discriminante*).

Aleatorização: o processo de fazer coisas de uma maneira não-sistemática ou aleatória. No contexto da pesquisa experimental, a palavra, em geral, se aplica à atribuição ao acaso dos participantes às diferentes condições do tratamento.

Alfa (α) de Cronbach: uma medida de confiabilidade de uma escala definida por:

$$\alpha = \frac{N^2 \overline{\text{Cov}}}{\sum s_{\text{item}}^2 \sum \text{Cov}_{\text{item}}}$$

em que o numerador da equação é simplesmente o número de itens (N) ao quadrado multiplicado

pela covariância média entre os itens (a média dos elementos fora da diagonal na *matriz das variâncias-covariâncias*). O denominador é a soma de todos os elementos da *matriz das variâncias-covariâncias*.

Amostra: uma pequena (mas com sorte representativa) coleção de unidades de uma *população*, usada para determinar verdades sobre essa população (por exemplo, como uma dada população se comporta em determinadas condições).

Análise de covariância: um procedimento estatístico que utiliza a razão F para testar a aderência de um modelo linear controlando o efeito que uma ou mais *covariáveis* apresentam sobre os resultados da *variável saída*. Na pesquisa experimental, esse modelo linear tende a ser definido em termos das médias dos grupos, e a ANOVA resultante é, dessa forma, um teste global se a média dos grupos difere após a variância na *variável saída* explicada por qualquer *covariável* ter sido removida.

Análise de fatores: uma técnica *multivariada* para identificar se a correlação entre um conjunto de variáveis observadas é devido ao relacionamento com uma ou mais *variáveis latentes* existentes nos dados, cada uma das quais assumindo a forma de um *modelo linear*.

Análise de Fatores Confirmatória (AFC): uma versão da *análise de fatores* em que hipóteses

específicas sobre a estrutura e relações entre as *variáveis latentes* subjacentes aos dados são testadas.

Análise de variância multivariada: família de testes que estende a *análise de variância* básica a situações nas quais mais do que uma *variável de saída* foi mensurada.

Análise de variância: um procedimento estatístico que utiliza a razão F para testar a aderência de um modelo linear. Na pesquisa experimental, esse modelo linear tende a ser definido em termos da média do grupo e a ANOVA resultante é, dessa forma, um teste global para verificar se as médias dos grupos diferem.

Análise discriminante: também conhecida como análise da função discriminante. Essa análise identifica e descreve a função discriminante de um conjunto de variáveis (combinação linear de variáveis) e é utilizada para dar seguimento a um teste MANOVA como uma forma de ver como o conjunto de variáveis permite discriminar o conjunto de casos.

Análise dos componentes principais (ACP): uma *técnica multivariada* para identificar os componentes lineares de um conjunto de variáveis.

Análise dos efeitos simples: Essa análise observa os efeitos de uma *variável independente* (*variável previsora* categórica) em níveis individuais de outra *variável independente*.

Análise Loglinear: um procedimento usado como uma extensão do teste *qui quadrado* para analisar situações nas quais você tem mais do que duas *variáveis categóricas* e você quer testar os relacionamentos entre essas variáveis. Essencialmente, um *modelo linear* é ajustado aos dados que preveem frequências esperadas (isto é, o número de casos esperados numa categoria dada). A esse respeito, é quase o mesmo que a *análise de variância*, mas para dados totalmente categóricos.

ANCOVA: acrônimo de *análise de covariância*.

ANOVA de Friedman: um teste não-paramétrico para verificar se dois ou mais grupos diferem. É a versão não-paramétrica da *ANOVA de medidas repetidas* de um fator.

ANOVA de medidas repetidas: uma *análise da variância* conduzida em qualquer delineamento no qual a *variável independente* (previsora) ou as *variáveis* (previsoras) foram mensuradas usando os mesmos participantes em todas as condições.

ANOVA Fatorial: uma análise de variância envolvendo duas ou mais *variáveis independentes* ou *previsores*.

ANOVA independente: *análise da variância* conduzida em qualquer delineamento no qual *variáveis independentes* ou *previsores* foram manipuladas usando participantes diferentes (isto é todos os dados vêm de entidades diferentes).

ANOVA Mista: *análise de variância* usada quando você tem um *delineamento misto*.

ANOVA: acrônimo para *análise de variância*.

Assimetria negativa: veja *assimetria*.

Assimetria positiva: veja *assimetria*.

Assimetria: uma medida da simetria de uma *distribuição de frequências*. Distribuições simétricas têm um coeficiente de assimetria igual a 0. Quando a maioria dos escores está concentrada no final da distribuição e a cauda aponta na direção do escore mais alto ou positivo, a assimetria é positiva. Contrariamente, quando a maioria dos escores está concentrada na parte final da distribuição e a cauda aponta na direção dos escores mais baixos ou negativos, a assimetria é negativa.

Autocorrelação: quando os *resíduos* de duas observações num modelo de regressão são correlacionados.

b_i : coeficiente de regressão não-padronizado. Indica a força do relacionamento entre um dado preditor, i , e uma saída em unidades de medidas do preditor. Ele é a mudança na saída associada à mudança de uma unidade no preditor.

β_i : coeficiente de regressão padronizado. Indica a força do relacionamento entre um dado preditor, i , e uma saída de forma *padronizada*. Ele é a mudança na saída (em desvios padrão) associada a uma alteração de um desvio padrão no preditor.

Bimodal: uma descrição de uma distribuição de observações que apresenta duas *modas*.

Carga de um fator: o *coeficiente de regressão* de uma variável para o *modelo linear* que descreve a *variável latente* ou *fator* na *análise de fatores*.

Chance: a probabilidade de um evento ocorrer dividido pela probabilidade daquele evento não ocorrer.

Chorar: é o que você faz depois de escrever um livro de estatística.

Coeficiente de correlação de Pearson: também chamado de coeficiente de correlação momento-produto de Pearson, é uma medida *padronizada* da força do relacionamento entre duas variáveis. Ele pode ter qualquer valor de -1 (à medida que uma variável muda, a outra muda na direção oposta pela mesma quantia) passando por 0 (à medida que uma variável muda, a outra não muda) até $+1$ (à medida que uma variável muda, a outra muda na mesma direção pela mesma quantia).

Coeficiente de correlação de Spearman: uma medida padronizada da força do relacionamento entre duas variáveis que não depende das hipóteses de um *teste paramétrico*. Ele é o *coeficiente de correlação de Pearson* calculado para dados que foram convertidos em postos.

Coeficiente de determinação: a proporção da variância em uma variável explicada pela segunda variável. Ele é o *coeficiente de correlação de Pearson* elevado ao quadrado.

Coeficiente de regressão: veja b_i e β_i .

Colinearidade perfeita: existe quando pelo menos um preditor num *modelo de regressão* é uma combinação linear perfeita de outros (o exemplo mais simples sendo dois preditores que são perfeitamente correlacionados – eles tem um coeficiente de correlação de 1).

Comparações pareadas: comparações de pares de médias.

Comparações Planejadas: outro nome para *contrastes planejados*.

Comunalidade: a proporção da variância de uma variável que é comum. Esse termo é utilizado principalmente na *análise de fatores*. Uma variável que não tem uma *variância exclusiva* (ou *variância aleatória*) terá uma comunalidade de 1 ,

enquanto uma variável que nada compartilha de sua variância com qualquer outra variável terá uma comunalidade de 0 .

Confiabilidade meio a meio: uma medida de *confiabilidade* obtida dividindo os itens de uma medida em duas metades (de uma maneira aleatória) e obtendo um escore de cada metade da escala. A correlação entre os dois escores, corrigidos levando em consideração o fato de que as correlações são baseadas em somente a metade dos itens, é usada como uma medida da confiabilidade. Existem duas maneiras de fazer isso. Spearman (1910) e Brown (1910) desenvolveram uma fórmula que não precisa do desvio padrão entre os itens:

$$r_{sh} = \frac{2r_{12}}{1 + r_{12}}$$

onde r_{12} é a correlação entre as duas metades da escala. Flanagan (1937) e Rulon (1939), entretanto, propuseram uma medida que leva em consideração a variância dos itens:

$$r_{sh} = \frac{4r_{12} \times s_1 \times s_2}{s_T^2}$$

no qual s_1 e s_2 são desvios padrão de cada metade da escala e s_T^2 é a variância de todo o conjunto. Veja Cortina (1993) para mais detalhes.

Confiabilidade Teste-Reteste: a habilidade de uma mensuração produzir resultados consistentes quando as mesmas entidades são testadas em dois pontos diferentes no tempo.

Confiabilidade: a habilidade de uma medida produzir resultados consistentes quando as mesmas entidades estão sendo mensuradas sob as mesmas condições.

Contrabalanceamento: o processo de sistematicamente variar a ordem em que condições experimentais são conduzidas. No caso mais simples, de existirem duas condições (A e B), o contrabalanceamento implica que metade dos participantes submete-se à condição A e em seguida à condição B, enquanto os restantes submetem-se à condição B seguida da A. O objetivo é remover tendenciocidades causadas por efeitos práticos ou de tédio.

Contraste diferença: um *contraste planejado não-ortogonal* que compara a média de cada condição (exceto a primeira) com a média global de todas as condições anteriores combinadas.

Contraste Helmert: um *contraste planejado não-ortogonal* que compara a média de cada condição (exceto a última) à média geral de todas as condições subsequentes combinadas.

Contraste polinomial: um contraste que testa por tendências nos dados. Na sua forma mais básica, ele procura por uma tendência linear (isto é, que as médias dos grupos aumentam proporcionalmente).

Contraste repetido: um *contraste planejado não-ortogonal* que compara as médias em cada condição (exceto a primeira) à média da condição precedente.

Contraste reverso de Helmert: outro nome para um *contraste diferença*.

Contraste simples: um *contraste planejado não-ortogonal* que compara a média em cada condição à média da primeira ou última condição dependendo em como o contraste for especificado.

Contrastes planejados: um conjunto de comparações entre grupos de médias projetado antes que qualquer dado seja coletado. São comparações teóricas baseadas na ideia de particionar a variância criada pelo efeito geral das diferenças dos grupos gradualmente em pequenas porções da variância. Esses testes têm mais poder do que os *testes post hoc*.

Correção contínua de Yates: um ajuste feito para o teste *qui-quadrado* quando a *tabela de contingência* é duas linhas por duas colunas (isto é, existem duas variáveis categóricas, ambas consistindo em apenas duas categorias). Em amostras grandes, o ajuste faz pouca diferença e é levemente duvidoso (veja Howell, 2002).

Correção de Bonferroni: uma correção aplicada ao nível α para controlar a *taxa de erro do Tipo I* global quando vários testes de significância são executados. Cada teste conduzido deve utilizar um critério de significância de nível α (normalmente 0,05) dividido pelo número de testes realizados. Essa é uma correção simples, porém efetiva, mas

tende a ser muito rígida quando muitos testes são executados.

Correção de Greenhouse e Geisser: uma estimativa do afastamento da *esfericidade*. O valor máximo é 1 (os dados atendem completamente a hipótese de *esfericidade*) e o mínimo é o *limite inferior*. Valores abaixo de 1 indicam distanciamento da *esfericidade* e são utilizados para corrigir os *graus de liberdade* associados à correspondente *razão F* multiplicando o seu valor pelo da estimativa. Alguns dizem que a correção de Greenhouse e Geisser é muito conservadora (estrita) e recomendam a *correção de Huynh-Feldt*.

Correção de Huynh-Feldt: uma estimativa do desvio da *esfericidade*. O valor máximo é 1 (os dados atendem completamente à hipótese da *esfericidade*). Valores abaixo desse indicam desvios da *esfericidade* e são utilizados para corrigir os *graus de liberdade* associados às *razões F* correspondentes multiplicando-as pelo valor da estimativa. É menos conservador do que a estimativa de *Greenhouse e Geisser*, mas alguns dizem que ela é muito liberal.

Correção Sidak: uma variante levemente menos conservadora da *correção de Bonferroni*.

Correlação bisserial: uma medida padronizada da força do relacionamento entre duas variáveis quando uma das duas é *dicotômica*. O coeficiente de correlação bisserial é utilizado quando uma variável é dicotômica contínua (isto é, tem um contínuo subjacente entre as categorias).

Correlação bivariada: uma correlação entre duas variáveis.

Correlação parcial: uma medida do relacionamento entre variáveis “controlando” o efeito que uma ou mais variáveis adicionais podem ter em ambas.

Correlação ponto bisserial: uma medida padrão da força do relacionamento entre duas variáveis quando uma das duas variáveis é *dicotômica*. O coeficiente de correlação ponto bisserial é usado quando a dicotomia é discreta ou verdadeira (isto é, quando não há um *continuum* subjacente entre as categorias). Um exemplo disso é a gravidez: a mulher está ou não grávida, não há meio termo.

Correlação por partes: outro nome para uma *correlação semiparcial*.

Correlação semiparcial: uma medida do relacionamento entre duas variáveis “controlando” o efeito que uma ou mais variáveis tem sobre uma dessas variáveis. Se nomearmos nossas variáveis x e y , ele fornece uma medida da variância em y compartilhada apenas com x .

Corridas de Wald-Wolfowitz: outra variante do teste de Mann-Whitney. Os escores são ordenados como no teste de Mann-Whitney, mas em vez de analisar as classificações esse teste procura por “corridas” de escores do mesmo grupo dentro de uma determinada ordem. Agora, se não houver diferença entre os grupos, obviamente classificações dos dois grupos devem ser entremeadas aleatoriamente. Entretanto, se os grupos são diferentes, você deve ver mais postos de um grupo na parte inferior e mais postos de outro grupo na parte superior. Procurando aglomerados de escores dessa maneira, o teste pode determinar se os grupos diferem.

Covariância: uma medida do relacionamento médio entre duas variáveis. É o produto cruzado médio (isto é, o produto cruzado dividido pelo número de observações menos um).

Covariável: a variável que tem um relacionamento com (em termos de *covariância*) ou tem o potencial de estar relacionada à *variável de saída* que está sendo mensurada.

Crítério de Kaiser: um método de *extração* na análise de fatores baseado na ideia de reter fatores com autovalores associados maiores do que 1. Esse método parece ser preciso quando o número das variáveis na análise é menor do que 30 e as *comunalidades* resultantes (após a *extração*) são maiores do que 0,7; ou quando o tamanho da amostra excede a 250 e a média da comunalidade é maior ou igual a 0,6.

Curtose: mensura o grau em que os escores se amontoam nas caudas de uma *distribuição de frequências*. Uma distribuição *platicúrtica* é uma que têm muitos escores nas caudas e, assim, é bem plana. Em contraste, as distribuições *leptocúrticas* são relativamente finas nas caudas, portanto, parecem bem pontudas.

Dados de razão: *dados intervalares*, mas com a propriedade adicional de que as razões apresentam significados. Por exemplo, as avaliações das pessoas para este livro na Amazon.com podem variar de 1 a 5; para esses dados serem de razão, eles não somente devem ter as propriedades dos *dados intervalares*, mas também uma avaliação de 4 deve representar genuinamente alguém que apreciou este livro duas vezes mais do que alguém cuja avaliação foi 2. Da mesma forma, alguém que avaliou o livro com 1 deve estar apenas meio impressionado em relação a alguém cuja avaliação foi 2.

Dados intervalares: dados mensurados numa escala ao longo da qual os intervalos são iguais. Por exemplo, as avaliações das pessoas sobre este livro na Amazon.com podem variar de 1 a 5; para esses dados serem intervalares, deve ser verdade que o aumento da avaliação deste livro de 1 para 2 deva ser o mesmo que um aumento de 4 para 5.

Dados nominais: onde os números apenas representam nomes, por exemplo, os nomes nas camisas dos jogadores: um jogador com o número 1 nas costas não é necessariamente pior do que um jogador com um 2 nas suas costas. Os números não têm outro significado do que representar o tipo de jogador (isto é, zagueiro, centro-avante, etc.).

Dados ordinais: dados que nos informam o que aconteceu e também a ordem em que ocorreu. Esses dados não dizem nada sobre as diferenças entre os valores. Por exemplo, as medalhas de ouro, prata e bronze são ordinais: elas nos dizem que o medalhista de ouro foi melhor do que o medalhista de prata; entretanto, eles não nos dizem quanto melhor (o competidor de ouro foi muito melhor do que o de prata ou os resultados foram muito próximos?).

Delineamento de medidas repetidas: um delineamento experimental em que diferentes condições de tratamento utilizam os mesmos organismos (isto é, em psicologia isso significaria que todas as pessoas tomariam parte em todas as condições experimentais), assim, os dados resultantes são relacionados (também conhecido como relacionamentos ou delineamentos entre sujeitos).

Delineamento entre sujeitos: outro nome para *delineamento de medidas repetidas*.

Delineamento entre grupos: outro nome para o *delineamento independente*.

Delineamento entre participantes: outro nome para o *delineamento independente*.

Delineamento fatorial independente: um delineamento experimental com dois ou mais *previsores* (ou *variáveis independentes*) em que todos foram manipulados usando diferentes participantes (ou qualquer entidade sendo testada).

Delineamento fatorial relacionado: um delineamento experimental incorporando dois ou mais *previsores* (ou *variáveis independentes*), onde todos foram manipulados usando os mesmos participantes (ou qualquer entidade que esteja sendo testada).

Delineamento independente: um delineamento experimental no qual condições diferentes de tratamento utilizam diferentes organismos (por exemplo, em psicologia isso significaria pessoas diferentes em condições de tratamento distintas), assim, os dados resultantes são independentes (também conhecido como delineamento entre grupo ou delineamento entre participantes ou sujeitos).

Delineamento misto: um delineamento experimental incorporando dois ou mais *previsores* (ou *variáveis independentes*), onde pelo menos um foi manipulado usando participantes diferentes (ou qualquer entidade que esteja sendo usada) e pelo menos um foi manipulado usando os mesmos participantes (ou entidades). Também conhecido como um delineamento por divisão em lotes (*split-plot*) porque Fisher desenvolveu a ANOVA para analisar dados agrícolas envolvendo lotes de terra contendo cultivos.

Delineamento relacionado: outro nome para um *delineamento de medidas repetidas*.

Desvio padrão: uma estimativa da variabilidade média (espalhamento) de um conjunto de dados mensurado na mesma unidade de mensuração dos dados originais. Ele é a raiz quadrada da variância.

Desvios dos contrastes: um *contraste não-ortogonal planejado* que compare as médias de cada grupo (exceto a primeira e a última de-

pendendo de como o contraste é planejado) com a média global (médias de todas as médias).

DFBeta padronizado: uma versão *padronizada* do DFBeta. Esses valores padronizados são mais fáceis de usar do que o próprio DFBeta porque pontos de corte universais podem ser aplicados. Stevens (1992) sugere observar casos com valores absolutos maiores do que 2.

DFBeta: uma medida da influência de um caso nos valores de b_i em um *modelo de regressão*. Se estimarmos um parâmetro de regressão b_i e eliminarmos um caso (valor) específico e estimarmos novamente o mesmo parâmetro de regressão b_i , a diferença entre essas duas estimativas será o DFBeta para o caso que foi eliminado. Pelo exame dos valores do DFBeta, é possível identificar casos que tem uma grande influência nos parâmetros do modelo de regressão; entretanto, o tamanho do DFBeta dependerá da unidade de medida dos parâmetros da regressão.

DFFit padronizado: uma versão *padronizada* do DFFit.

DFFit: uma medida da influência de um caso. É a diferença entre o *valor previsto ajustado* e o *valor original previsto* para um caso específico. Se o caso não exerce influência, o DFFit deve ser zero, entretanto, espera-se que casos que não exerçam influência tenham pequenos valores DFFit. Existe um problema com essa estatística, pois ela depende das unidades de medida da variável de saída e, assim, um DFFit de 0,5 será pequeno se a saída variar de 0 a 100, mas grande se ela variar apenas de 0 a 1.

Diagrama de barras de erros: uma representação gráfica da média de um conjunto de observações que inclui o intervalo de 95% para a média. A média é normalmente representada como um ponto, quadrado ou retângulo estendendo o valor. O intervalo de confiança é representado por uma linha projetada da média (para cima, para baixo ou ambos) até uma pequena linha horizontal representando os limites do intervalo de confiança. Barras de erro podem ser desenhadas utilizando o erro padrão ou desvio padrão em vez do intervalo de 95% de confiança.

Diagrama de Caixa e Bigodes (Boxplot): uma representação gráfica de algumas características im-

portantes de um conjunto de observações. No centro do diagrama está a *mediana*, que é rodeada por um retângulo (caixa), onde as arestas superior e inferior são os limites dentro dos quais encontram-se 50% das observações (o *intervalo interquartilico*). A partir dos extremos dos retângulos existem duas linhas (os bigodes) que se estendem até os valores extremos do conjunto respectivamente.

Diagrama de declividade: um gráfico traçando cada fator numa *análise de fatores* (eixo x) contra seu autovalor associado (eixo y). Ele mostra a importância relativa de cada fator. Esse gráfico tem uma forma muito característica (existe uma descida acentuada na curva seguida por uma cauda suave) e o ponto de inflexão dessa curva é geralmente usado como ponto de corte do número de fatores *extraídos*. Com uma amostra de mais de 200 participantes, ele fornece um critério confiável para a *extração dos fatores* (Stevens, 1992).

Diagrama de dispersão: um gráfico que traça os valores de uma variável contra o valor correspondente da outra variável (o valor correspondente de uma terceira variável pode também ser incluído num diagrama de dispersão tridimensional).

Dicotômica: descrição de uma variável que consiste em apenas duas categorias (por exemplo, a variável gênero é dicotômica porque ela é formada apenas pelas categorias masculino e feminino).

Distância de Cook: uma medida da influência global de um caso (valor) em um modelo. Cook e Winberg (1982) sugeriram que valores superiores a 1 podem ser motivo de preocupação.

Distância de Mahalanobis: mede a influência de um caso examinando a distância dos casos da(s) média(s) da(s) variável(is) previsora(s). Você deve buscar casos com valores altos. Não é fácil estabelecer um ponto de corte a partir do qual se preocupar, embora Barnett e Lewis (1978) produziram uma tabela de valores críticos dependendo do número de previsores e do tamanho da amostra. Do seu trabalho está claro que mesmo com grandes amostras ($N = 500$) e cinco previsores, valores acima de 25 são preocupantes. Em amostras menores ($N = 100$) e com poucos previsores (a saber, três), valores maiores do que 15 são problemáticos e em amostras muito pequenas ($N = 30$) com so-

mente dois previsores, valores maiores do que 11 devem ser examinados. Entretanto, para um conselho mais específico, consulte a tabela de Barnett e Lewis (1978).

Distribuição amostral: a *distribuição de probabilidade* de uma estatística. Você pode pensar nisso da seguinte forma: se tirarmos uma amostra de uma população e calcularmos uma estatística (por exemplo, a *média*), o valor dessa estatística irá depender de algum modo da amostra que tiramos. Assim, a estatística irá variar levemente de amostra para amostra. Se, hipoteticamente, tirarmos muitas amostras da população e calcularmos a estatística de interesse, será possível criar uma distribuição de frequência dos valores que conseguimos. A distribuição resultante é o que a distribuição amostral representa: a distribuição de possíveis valores de uma estatística que podemos esperar de uma determinada população.

Distribuição de frequências: o conjunto de valores de uma variável com as respectivas frequências (absolutas ou relativas). Se a variável é discreta, a distribuição é apresentada em um diagrama de barras e se contínua, em um histograma. Em qualquer caso, a variável é representada no eixo x e as frequências são representadas no eixo y.

Distribuição de probabilidade: o conjunto de todos os valores de uma variável junto com as suas probabilidades, se a variável for discreta, ou densidades, se a variável for contínua.

Distribuição normal: uma *distribuição de probabilidade* de uma variável aleatória que se sabe ter certas propriedades. Ela é perfeitamente simétrica (tem uma assimetria de 0) e tem um curtose de 0.

Distribuição qui-quadrado: a *distribuição de probabilidade* da soma dos quadrados de diversas variáveis normalmente distribuídas. Ela é geralmente utilizada para testar hipóteses sobre dados categóricos.

Editor de dados (Data editor): a janela principal no SPSS na qual se entram com os dados e se executam as funções estatísticas.

Editor de sintaxe: uma janela no SPSS para escrever e editar a sintaxe.

Efeito “álcool” (*beer-goggles*): fenômeno em que pessoas do sexo oposto (ou do mesmo, dependendo da orientação sexual) tendem a parecer mais atrativas após alguns goles de bebida alcoólica.

Efeito de interação: o efeito combinado de duas ou mais *variáveis previsoras* numa *variável de saída*.

Efeito principal: o efeito único de uma *variável previsora* (ou *variável independente*) numa *variável de saída*. O termo é geralmente usado no contexto da ANOVA.

Efeito supressor: quando um previsor tem um efeito significativo, mas somente quando outra variável é mantida constante.

Efeito tédio: refere-se à possibilidade de que o desempenho em tarefas possa ser influenciado (a hipótese é que ele seja uma influência negativa) por tédio ou falta de concentração se muitas tarefas existirem, ou se elas durarem muito tempo.

Efeitos práticos: relacionado à possibilidade de que o desempenho de um participante em uma tarefa seja influenciado (positiva ou negativamente) se a tarefa for repetida em virtude da familiaridade com a situação experimental e/ou com as medidas sendo utilizadas.

Encolhimento: a perda do poder preditivo de um modelo de regressão se o modelo foi derivado de uma população da qual a amostra foi retirada em vez da própria amostra.

Equamax: um método de *rotação ortogonal* híbrido do *quantimax* e *varimax*. Ele pode se comportar erraticamente (ver Tabachnick e Fidell, 2001), portanto, talvez seja melhor evitá-lo.

Erro padrão das diferenças: se você fosse tirar muitos pares de amostras de uma população e calcular suas médias, poderia também calcular a diferença entre suas médias. Se você traçar essas diferenças entre as médias das amostras como uma *distribuição de frequências*, terá a *distribuição amostral* das diferenças. O desvio padrão dessa distribuição amostral é o *erro padrão das diferenças*. Desse modo, ele é a medida da variabilidade das diferenças entre as médias das amostras.

Erro padrão: o desvio padrão da *distribuição amostral* de uma estatística. Para uma determinada estatística (por exemplo, a *média*), ele informa quanta variabilidade existe nessa estatística através das *amostras* da mesma *população*.

Erro SSCP (E): a soma dos quadrados dos erros e a matriz produto cruzado. Ele é a *soma dos quadrados e matriz produto cruzado* para o erro em um *modelo linear* preditivo ajustado a dados *multivariados*. Ele representa a *variância não-sistemática* e é o equivalente *multivariado* da *soma dos quadrados dos resíduos*.

Erro tipo I: ocorre quando acreditamos que exista um efeito genuíno na nossa população, mas de fato, ele não existe.

Erro tipo II: ocorre quando acreditamos que não exista um efeito na população, mas, na verdade, ele existe.

Erros independentes: para quaisquer duas observações em regressão, os *resíduos* não devem estar correlacionados (devem ser independentes).

Escore discriminante: um escore para um caso individual de uma CL (*variate*) função discriminante obtido pela substituição do escore do caso na variável medida na equação que define a CL em questão.

Escore do fator: um único escore de uma entidade representando seu desempenho em alguma *variável latente*. O escore pode ser grosseiramente conceitualizado da seguinte forma: pegue o escore de uma entidade em cada uma das variáveis que compõe um fator e multiplique-o pela *carga* correspondente para a variável e então some esses valores (ou faça a média deles).

Escore-z: o valor de uma observação expresso em unidades de desvios padrão. Ele é calculado pegando a observação, subtraindo dela a média de todas as observações e dividindo o resultado pelo desvio padrão de todas as observações. Convertendo a distribuição das observações em escores-z, você cria uma nova distribuição que tem uma média de 0 e um desvio padrão de 1.

Esfericidade: uma forma menos restritiva da *simetria composta* que assume que as variâncias das diferenças entre os dados obtidos de um mesmo

participante (ou qualquer outra entidade sendo testada) sejam iguais. Essa hipótese é mais comumente encontrada na *ANOVA de medidas repetidas* e se aplica somente quando há mais do que dois pontos dos dados do mesmo participante (veja também a *correção de Greenhouse e Geisser* e a *correção de Huynh-Feldt*).

Estatística de Wald: um teste estatístico com uma distribuição de probabilidade conhecida (uma distribuição qui-quadrado) usado para testar se o coeficiente b para um predictor num modelo de regressão logística é significativamente diferente de zero. Ele é análogo à estatística t num modelo de regressão no qual é simplesmente o coeficiente b dividido pelo seu erro padrão. A estatística de Wald não é precisa quando o coeficiente de regressão (b) for grande porque o erro padrão tende a inflacionar, assim subestimando a estatística de Wald.

Estatística do escore eficiente de Roa: uma estatística mensurando o mesmo que a estatística de Wald, mas computacionalmente mais fácil de ser calculada.

Estatística t : o t do estudante é um teste estatístico com uma distribuição de probabilidade (a distribuição t) conhecida. No contexto de regressão, ele é usado para testar se um coeficiente de regressão b é significativamente diferente de zero; no contexto de um trabalho experimental, ele é usado para testar se as diferenças entre duas médias são significativamente diferentes de zero (veja também o teste t independente e teste t dependente).

Estatística teste: uma estatística para a qual sabemos quão frequentemente os valores ocorrem. O valor observado de tal estatística é tipicamente usado para testar hipóteses.

Estimativa de máxima verossimilhança: uma maneira de estimar parâmetros escolhendo os parâmetros que tornam os dados mais prováveis de terem acontecido. Imagine que calculamos a probabilidade de conseguir os dados observados para um conjunto de parâmetros: se essa probabilidade é alta, esses parâmetros geram uma boa aderência dos dados; inversamente, se a probabilidade for baixa, esses parâmetros não produzem uma boa aderência para os nossos dados. A estimativa de

máxima verossimilhança escolhe os parâmetros que maximizam a probabilidade.

Eta ao quadrado: uma medida do tamanho de efeito que é a razão entre a soma dos quadrados do modelo para a soma dos quadrados totais – assim, é o coeficiente de determinação com outro nome. Francamente, é uma perda de tempo, não somente porque ele é viciado, mas ele também mede o efeito global da ANOVA e assim não pode ser interpretado de forma significativa.

Exp(B): um indicador de mudanças nas chances resultante da mudança de uma unidade no predictor na regressão logística. Se o valor é maior que 1, ele indica que à medida que o predictor aumenta, as chances do resultado ocorrer também aumentam. Contrariamente, um valor menor que 1 indica que à medida que o predictor aumenta, a chance do resultado acontecer diminui.

Extração: um termo utilizado no processo de decidir se um fator na análise de fatores é estatisticamente importante para ser “extraído” dos dados e interpretado. A decisão é baseada na magnitude do autovalor associado ao fator. Veja diagrama de declividade, critério de Kaiser.

F de Brown-Forsythe: uma versão da razão F projetada para ser utilizada quando a hipótese da homogeneidade da variância for violada.

F de Welch: uma versão da razão F projetada para ser precisa quando a hipótese de homogeneidade das variâncias for violada.

Fator de inflação da variância (FIV): uma medida da multicolinearidade. O FIV indica se um predictor tem um forte relacionamento linear com o(s) outro(s) predictor(es). Myers (1990) sugere que o valor de 10 é preocupante. Bowerman e O’Connell (1990) sugerem que se a média do FIV é maior do que 1, a multicolinearidade pode tornar o modelo de regressão tendencioso.

Fator: outro nome para uma variável independente ou predictor tipicamente utilizado na descrição de delineamentos experimentais. No entanto, ele também é utilizado como sinônimo de variável latente na análise de fatores.

Fatoração alfa: um método de análise de fatores.

Fi (Phi): uma medida da força da associação entre duas *variáveis categóricas*. O *fi* é usado com *tabelas de contingência 2x2* (tabelas nas quais há duas variáveis categóricas e cada variável tem somente duas categorias). Ele é uma variante do *teste do qui quadrado* χ^2 : $\phi = \sqrt{\chi^2/n}$, onde *n* é o número total de observações.

FIV: veja *fator de inflação da variância*.

Generalização: a habilidade de um modelo estatístico informar algo além do conjunto de dados em que ele foi baseado. Se um modelo generaliza, assume-se que previsões desse modelo podem ser aplicadas não apenas à amostra na qual ele foi baseado, mas para a população de onde a amostra foi extraída.

Glossário: uma coleção de definições imprecisas (escritas tarde da noite, quando você já deveria estar dormindo) sobre coisas que você pensa que já entendeu até que um editor de livros o obrigue a defini-las.

Gráfico de interação: um gráfico mostrando as médias de duas ou mais *variáveis independentes* no qual as médias de uma variável são mostradas em diferentes níveis da outra variável. Excepcionalmente, as médias estão conectadas por linhas ou estão expostas como barras. Esses gráficos são usados para ajudar a entender os *efeitos de interação*.

Graus de liberdade (Degrees of freedom): algo impossível de explicar em algumas páginas que dirá em poucas linhas. Essencialmente, é o número de “entidades” que estão livres para variar quando se estima algum tipo de parâmetro estatístico. Ele é usado em vários testes de significância para muitas das estatísticas mais utilizadas como o *teste F*, o *teste t* e o *teste qui-quadrado*, e serve para determinar a *distribuição de probabilidade* da estatística teste utilizada. A explicação envolvendo jogadores de rúbi, no Capítulo 8, é mais interessante...

HE⁻¹: uma matriz funcionalmente equivalente à *hipótese SSCP* dividida pelo *erro SSCP* na MANOVA. Conceitualmente, ela representa a razão da *variância sistemática* para a *não-sistemática*, e é a análoga *multivariada* da razão *F*.

Heterocedasticidade: o contrário de *homocedasticidade*. Isso ocorre quando os resíduos em cada

nível da(s) variável(s) previsora(s) têm variâncias diferentes. Colocando de outra maneira, em cada ponto ao longo de qualquer variável previsora, a distribuição dos resíduos é diferente.

Hipótese experimental: a previsão de que a manipulação experimental terá algum efeito ou que certas variáveis se relacionarão entre si.

Hipótese nula: reverso da *hipótese experimental* de que sua previsão está errada e que o efeito previsto não existe.

Hipótese SSCP (H): a hipótese soma dos quadrados e matriz do produto cruzado. Essa é a *matriz da soma dos quadrados e do produto cruzado* para o *modelo linear* preditivo ajustado aos dados *multivariados*. Ela representa a *variância sistemática* e a equivalente *multivariada* ao *da soma dos quadrados do modelo*.

Hipótese: uma previsão sobre o estado do mundo (veja *hipótese experimental* e *hipótese nula*).

Histograma: veja *distribuição de frequência*.

Homocedasticidade: uma hipótese da análise de regressão em que os resíduos em cada nível da(s) variável(s) previsora(s) têm variâncias similares. Ou seja, em cada ponto ao longo de qualquer variável previsora, a distribuição dos resíduos deve ser constante.

Homogeneidade da variância: uma hipótese de que a variância de uma variável é estável (isto é, relativamente similar) em todos os níveis de outra variável.

Homogeneidade das inclinações da regressão: uma hipótese da *análise da covariância*. Essa é uma hipótese em que o relacionamento entre a *covariância* e a *variável da saída* é constante ao longo dos diferentes níveis de tratamento. Assim, para três condições de tratamento, se existe um relacionamento positivo entre a covariável e a saída em um grupo, presumimos que exista um relacionamento positivo de tamanho semelhante entre a covariável e a saída nos outros dois grupos também.

Homogeneidade das matrizes de covariâncias: uma hipótese de alguns testes *multivariados* como a MANOVA. É uma extensão da *hipótese da homogeneidade da variâncias* na análise univa-

riada. Entretanto, assim como se assume que as *variâncias* são as mesmas ao longo dos grupos, ela também assume que os relacionamentos (covariâncias) entre essas *variáveis dependentes* são aproximadamente iguais. Ela é testada comparando as *matrizes de variâncias-covariâncias* dos diferentes grupos na análise.

Independência: a hipótese de que um ponto de dados não influencia outro. Quando os dados são provenientes de pessoas, isso basicamente significa que o comportamento de um participante não influencia o comportamento de outro.

Influência (Leverage): estatísticas de influências (ou valores chapéu) calculam a influência do valor observado da saída da variável sobre os valores previstos. O valor médio da influência é $(k + 1)/n$, no qual k é o número de previsores no modelo e n é o número de participantes. Os valores da influência podem estar entre 0 (o caso não tem influência alguma) e 1 (o caso tem completa influência sobre a previsão). Se nenhum dos casos exerce desmedida influência sobre o modelo, devemos esperar que todos os valores da estatística influência estejam próximos ao valor médio. Hoaglin e Welsch (1978) recomendam investigar casos com valores maiores do que duas vezes a média $2(k + 1)/n$ e Stevens (1992) recomenda usar três vezes a média $3(k + 1)/n$ como um ponto de corte para identificar casos com influência desmedida.

Intervalo de confiança: para uma dada estatística obtida de uma amostra (por exemplo, a média), o intervalo de confiança é um conjunto de valores em torno da estatística que se acredita que contenha, com uma determinada probabilidade (por exemplo, 95%), o verdadeiro valor da estatística (isto é, o valor populacional).

Intervalo interquartilico: os limites que contém 50% de um conjunto de observações.

Lambda (λ) de Goodmann and Kruskal: mede a redução proporcional no erro obtida quando um membro de uma categoria de uma variável é utilizado para prever a inclusão em uma categoria de outra variável. Um valor de 1 significa que uma variável prevê perfeitamente a outra enquanto um valor de 0 indica que uma variável não prevê a outra de forma alguma.

Lambda de Wilks (Λ): um teste estatístico na MANOVA. Ele é o produto da variância não explicada em cada uma das *combinações lineares da função discriminante*, assim, ele representa a razão entre a variância do erro e a variância total (SS_R/SS_T) para cada combinação linear.

Leptocúrtica: veja *curtose*.

Limite inferior: o nome dado para o valor mais baixo possível para a estimativa de *esfericidade* de *Greenhouse e Geisser*. Seu valor é $1/k - 1$, onde k é o número das condições de tratamento.

Maior raiz de Roy: um teste estatístico na MANOVA. Ele é o autovalor para a primeira *combinação linear da função discriminante* de um conjunto de observações. Assim, ela é o mesmo que o *traço de Hotelling-Lawley*, mas apenas para a primeira variável. Ele representa a proporção da variância explicada em relação à variância não-explicada (SS_M/SS_R) para a primeira função discriminante.

MANOVA: acrônimo para *análise de variância multivariada* (*multivariate analysis of variance*).

Matriz componente: termo genérico para uma *matriz estrutura* na *análise de componentes principais* no SPSS.

Matriz de fatores: um termo genérico para a estrutura matricial na *análise de fatores* do SPSS.

Matriz de transformação dos fatores, Λ : uma matriz utilizada na *análise de fatores*. Ela pode ser pensada como contendo os ângulos pelos quais os fatores são girados na *rotação* de fatores.

Matriz de variâncias-covariâncias: uma matriz quadrada (isto é, mesmo número de linhas e colunas) representando as variáveis mensuradas. As diagonais representam as *variâncias* dentro de cada variável enquanto os elementos fora da diagonal representam as *covariâncias* entre pares de variáveis.

Matriz estrutural: uma matriz na *análise dos fatores* contendo os *coeficientes de correlação* para cada variável em cada *fator* dos dados. Quando a *rotação ortogonal* é usada, ela é a mesma que a *matriz padrão*, mas quando a rotação oblíqua é usada, essas matrizes são diferentes.

Matriz identidade: uma matriz quadrada (isto é, com o mesmo número de linhas e colunas) em que os elementos da diagonal são iguais a 1 e todos os demais são iguais a zero. As seguintes são exemplos desse tipo de matriz:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matriz padrão: uma matriz na *análise de fatores* contendo os *coeficientes de regressão* para cada variável em cada *fator* dos dados (veja também *matriz estrutural*).

Matriz quadrada: uma *matriz* que tem um número igual de colunas e linhas.

Matriz SSCP total (T): a soma total dos quadrados e matriz do produto cruzado. Ela é a *soma dos quadrados* e *matriz dos produtos cruzados* para um conjunto completo de observações. Ela é a versão *multivariada* equivalente à *soma total dos quadrados*.

Matriz: uma coleção de números organizados em linhas e colunas. Os valores dentro da matriz são tipicamente referidos como *componentes* ou *elementos*.

Média: um modelo estatístico simples do centro da distribuição dos escores. Uma estimativa hipotética do escore “típico”.

Média Ajustada: no contexto da *análise de covariância*, esse é o valor da média do grupo ajustada para o efeito da *covariável*.

Média dos Quadrados: uma medida de variabilidade média. Para cada *soma dos quadrados* (que mensuram a variabilidade total), é possível criar a média dos quadrados dividindo pelo número de coisas usadas para calcular a soma dos quadrados (ou alguma função dela).

Média geral: a *média* de conjunto completo de observações.

Média Harmônica: é o inverso da média aritmética dos inversos.

$$H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}}$$

Mediana: o escore do meio de um conjunto ordenado de observações. Quando existe um número par de observações a mediana é a média dos dois escores que estão nos lados do valor que supostamente estaria no meio.

Medida da adequação da amostra de Kaiser-Meyer-Olkin (KMO): pode ser calculada para variáveis individuais ou múltiplas e representa a razão das correlações ao quadrado entre variáveis para a *correlação parcial* ao quadrado entre as variáveis. Ela varia entre 0 e 1: um valor de 0 indica que a soma das correlações parciais é grande em relação à soma das correlações, indicando difusão no padrão das correlações (entretanto, a *análise dos fatores* é provavelmente inapropriada); um valor próximo a 1 indica que padrões de correlações são relativamente compactos, assim, a análise dos fatores deve produzir fatores distintos e confiáveis. Valores entre 0,5 e 0,7 são medíocres, valores entre 0,7 e 0,8 são bons, valores entre 0,8 e 0,9 são muito bons e valores acima de 0,9 são soberbos (veja Hutcheson e Sofroniou, 1999).

Método de Anderson-Rubin: uma forma de calcular *escores de fatores* que produz escores que são não-correlacionados e *padronizados* com média 0 e desvio padrão 1.

Método Monte Carlo: um termo aplicado ao processo de usar simulações de dados para resolver problemas estatísticos.

Moda: o escore mais frequente num conjunto de dados.

Modelo de regressão: veja *regressão simples* e *regressão múltipla*.

Modelo linear: um modelo baseado numa linha reta.

Modelo saturado: um modelo que se ajusta perfeitamente aos dados e, portanto, não tem erro. Ele contém todos os *efeitos principais* e *interações* possíveis entre as variáveis.

Multicolinearidade: uma situação em que duas ou mais variáveis apresentam um relacionamento linear próximo.

Multimodal: uma distribuição de dados que apresenta mais do que *duas modas*.

Multivariada: significa “muitas variáveis” e é geralmente usada quando se refere a análises nas quais existe mais do que uma *variável de saída* (por exemplo, MANOVA, *análise de componentes principais*, etc.).

Nível α : a probabilidade de cometer *erro do tipo I* (normalmente esse valor é 5%).

Nível β : A probabilidade de cometer *erro do Tipo II* (Cohen, 1992, sugere um valor máximo de 0,20).

Normalidade multivariada: uma extensão de uma distribuição normal a múltiplas variáveis. Ela é uma *distribuição de probabilidade* de um conjunto de variáveis $v' = [v_1, v_2, \dots, v_n]$ dado por:

$$f(v_1, v_2, \dots, v_n) = \frac{1}{2\pi^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (v - \mu)' \Sigma^{-1} (v - \mu) \right]$$

onde qual μ é o vetor das médias das variáveis e Σ é a *matriz variâncias-covariâncias*. Se isso fizer sentido para você, então você é mais inteligente do que eu.

O Traço (T^2) de Hotelling-Lawley: um teste estatístico na MANOVA. É a soma dos autovalores para cada CL (*variate*) da função discriminante dos dados, assim, conceitualmente é o mesmo que a razão F na ANOVA: ele é a soma da razão da *variância sistemática* pela *não-sistemática* (SS_M/SS_R) para cada uma das variáveis.

Oblimin Direto: um método de rotação oblíqua.

Ômega ao quadrado: uma medida do *tamanho de efeito* associado à ANOVA menos tendenciosa do que o *eta ao quadrado*. Ela é uma função (às vezes horrenda) da *soma dos quadrados do modelo* e da *soma dos quadrados dos resíduos* e não é muito útil, pois mede o efeito geral da ANOVA e, assim, não pode ser interpretada de uma maneira significativa. Em todos os outros sentidos ela é ótima.

Ortogonal: isso significa perpendicular (ângulo reto) a algo. Ele tende a ser equiparado com *independência* em estatística por causa da conotação de que os *modelos lineares perpendiculares* no espaço geométrico são completamente independentes (um não é influenciado pelo outro).

Padronização: o processo de converter uma variável em uma unidade padrão de medida. A unidade de medida tipicamente usada é o *desvio padrão* (veja também escores- z). A padronização nos permite comparar dados quando diferentes unidades de medidas foram usadas (podemos comparar medidas de peso em quilogramas com altura medida em polegadas).

Padronizado: veja *padronização*.

Parcializar: para parcializar o efeito de uma variável, devemos controlar o efeito que a variável tem no relacionamento entre duas outras variáveis (veja *correlação parcial*).

Pesos: um número pelo qual algo (geralmente uma variável em estatística) é multiplicado. O peso atribuído a uma variável determina a influência que aquela variável tem dentro de uma equação matemática: muito peso dá à variável muita influência.

Platicúrtica: veja *curtose*.

Poder: a habilidade de um teste detectar um efeito de um tamanho específico (um valor de 0,8 é um bom nível para se atingir).

População: em termos estatísticos, isso geralmente se refere à coleta de unidades (sejam pessoas, plâncton, plantas, cidades, autores suicidas, etc.) para a qual queremos generalizar o conjunto de constatações ou o modelo estatístico.

Produto-cruzado dos desvios: uma medida do relacionamento total entre duas variáveis. Ele é o produto dos desvios de uma variável em relação a sua média multiplicada pelos desvios da outra variável em relação à média.

Promax: um método de *rotação oblíqua* computacionalmente mais rápido do que *oblimin* e útil para grandes conjuntos de dados.

Q de Cochran: esse teste é uma extensão do *teste de McNemar* e ele é basicamente uma ANOVA de *Friedman* para dados *dicotômicos*. Por exemplo, se perguntássemos a 10 pessoas se elas gostariam de atirar em Justin Timberlake, David Beckham e Tony Blair e elas só pudessem responder sim ou não, e as respostas fossem codificados como 0 (não) e 1 (sim), poderíamos utilizar o teste de Cochran nesses dados.

Quartimax: um método de *rotação ortogonal*. Ele tenta maximizar a dispersão das cargas de um fator para uma variável ao longo de todos os fatores. Geralmente, isso resulta em diversas variáveis com altas cargas em um único fator.

R múltiplo: o coeficiente de correlação múltipla. Ele é a correlação entre os valores observados de uma saída e os valores de uma saída prevista por um modelo de regressão múltipla.

R² ajustado: uma medida da perda do poder preditivo ou *encolhimento* na regressão. O R² ajustado nos informa quanta variância na saída será de responsabilidade do modelo derivado da população de onde a amostra foi retirada.

R²_{CS} de Cox e Snell: uma versão do *coeficiente de determinação* para a regressão logística. Ele é baseado na verossimilhança log de um modelo (LL(Novo)) e a verossimilhança log do modelo original (LL(base)), e o tamanho da amostra, n . Entretanto, ele se destaca por não alcançar seu valor máximo de 1 (ver R^2_N de Nagelkerke).

R²_L de Hosmer e Lemeshow: Uma versão do *coeficiente de determinação* para a regressão logística. É uma tradução bem literal no sentido de que ele é o $-2LL$ do modelo dividido pelo $-2LL$ original; em outras palavras, ele é a razão do que o modelo pode explicar comparado ao que havia para ser explicado em primeiro lugar!

R²_N de Nagelkerke: Uma versão do *coeficiente de determinação* para a regressão logística. Ele é uma variação do R^2_{CS} de Cox e Snell que supera o problema que essa estatística tem de não ser capaz de alcançar seu valor máximo.

Razão de Chances: a razão entre a *probabilidade* de um evento ocorrer em um grupo comparado com outro. Assim, por exemplo, se a chance de morrer após escrever o glossário é 4 e a chance de morrer se não escrever o glossário é de 0,25, a razão de chances é de $4/0,25 = 16$. Isto significa que se você escrever um glossário, tem 16 vezes mais chances de morrer do que se não escrevê-lo. Uma razão de chances de 1 indicaria que a *chance* de uma saída particular é igual em ambos os grupos.

Razão de covariância (RCV): uma medida do quanto um caso (valor individual) influencia a variância das estimativas dos parâmetros em um

modelo de regressão. Quando essa razão estiver próxima de 1, o caso tem pouca influência sobre as variâncias das estimativas do modelo. Balsey e colaboradores. (1980) recomendam o seguinte: se o RCV de um caso for maior que $1 + [3(k + 1)/n]$, eliminar esse caso afetará a precisão da estimativas de alguns dos parâmetros dos modelos, mas se ele for menor do que $1 - [3(k + 1)/n]$, retirar o caso da análise melhorará a precisão das estimativas (k é o número de previsores e n é o tamanho da amostra).

Razão de variâncias: a variância do grupo com a maior variação dividida pela variância do grupo com a menor variação. Se essa razão for menor do que 2, é seguro supor a *homogeneidade das variâncias*.

Razão F: uma estatística teste com uma *distribuição de probabilidade* conhecida (a distribuição F). É a razão entre variabilidade média nos dados que um dado modelo pode explicar para a variabilidade média que não pode ser explicada pelo mesmo modelo. Ela é utilizada para testar o ajuste global dos modelos de *regressão simples* e *múltipla* e para testar a diferença global entre grupos de médias em experimentos.

Reações extremas de Moses: um teste não-paramétrico que compara a variabilidade dos escores em dois grupos, assim, é parecido com o *teste não-paramétrico de Levene*.

Regra da soma das variâncias: isso afirma que a variância da diferença entre duas variáveis independentes é igual à soma das suas variâncias.

Regressão Hierárquica: um método de *regressão múltipla* no qual a ordem em que os previsores entram no modelo de regressão é determinada pelo pesquisador baseado em pesquisas prévias: variáveis que já sabemos serem previsoras entram primeiro, novas variáveis entram depois.

Regressão logística: uma versão da *regressão múltipla* em que a saída é *dicotômica*.

Regressão múltipla: uma extensão da *regressão simples* na qual uma saída é prevista por uma combinação linear de duas ou mais variáveis previsoras. A forma do modelo é $Y_i = (b_0 + b_1X_1 + b_2X_2 + \dots + b_nX_n) + \Delta_1$, no qual a saída é representada por Y e cada preditor é representado por X . Cada preditor tem um coeficiente de regressão b_i asso-

ciado a ele e b_0 é o valor da saída quando todos os previsores são zero.

Regressão passo a passo: um método de regressão múltipla no qual as variáveis entram no modelo baseadas num critério estatístico (a correlação semi-parcial com a variável de saída). Uma vez que uma nova variável é colocada no modelo, todas as variáveis do modelo são avaliadas para ver se elas devem ser removidas.

Regressão por blocos: outro nome para a *regressão hierárquica*.

Regressão simples: um *modelo linear* no qual uma variável ou saída é prevista a partir de uma única variável previsora. O modelo toma a forma $Y_i = (b_0 + b_1X_i) + \Delta_i$, na qual Y é a variável de saída, X é a previsora, b_1 é o coeficiente de regressão associado com a previsora, e b_0 é o valor da saída quando o preditor é zero.

Resíduo eliminado: uma medida da influência de um caso (observação) específico dos dados. Ele é a diferença entre o *valor previsto ajustado* para o caso e o valor original observado para esse mesmo caso.

Resíduo estudentizado eliminado: uma medida da influência de um caso específico dos dados. Ele é a versão padronizada do *resíduo eliminado*.

Resíduo: a diferença entre o valor que um modelo prevê e o valor observado nos dados em que o modelo é baseado. Quando o resíduo é calculado para cada observação no conjunto de dados, a coleção resultante é referida como *resíduos*.

Resíduos estudentizados: uma variação dos resíduos padronizados. Resíduos estudentizados são os *resíduos não-padronizados* divididos por uma estimativa de seu desvio padrão que varia de ponto a ponto. Esses resíduos têm as mesmas propriedades que os resíduos padronizados, mas geralmente fornecem estimativas mais precisas da variância do erro de um caso específico.

Resíduos não-padronizados: os *resíduos* de um modelo expressos nas unidades em que a variável de saída original foi mensurada.

Resíduos ou Desvios: a diferença entre valores observados de uma variável e os valores dessa mesma variável previstos pelo modelo estatístico.

Resíduos padronizados: os *resíduos* de um modelo expresso em unidades de desvios padrão.

Resíduos padronizados com valor absoluto maior do que 3,29 (na verdade, geralmente usamos 3) são causa de preocupação porque numa amostra padrão um valor tão alto é improvável de acontecer por acaso; se mais de 1% de suas observações tem resíduos padronizados com valor absoluto maior do que 2,58 (geralmente dizemos 2,5), há evidências de que o nível de erro dentro do nosso modelo não é aceitável (o modelo é um ajuste muito pobre dos dados da amostra); se mais do que 5% das observações tem resíduos padronizados em valores absolutos maiores do que 1,96 (ou 2, por conveniência), há também evidências de que o modelo é uma representação pobre dos dados reais.

Resíduos: veja *resíduo*.

Rotação oblíqua: um método de *rotação* na *análise de fatores* que permite que os fatores subjacentes sejam correlacionados.

Rotação ortogonal: um método de *rotação* na *análise de fatores* que mantém os fatores subjacentes independentes (isto é, não-correlacionados).

Rotação: um processo na *análise de fatores* para melhorar a interpretação dos *fatores*. Em essência, é uma tentativa de transformar os fatores que emergem da análise de modo a maximizar as maiores *cargas* e minimizar as menores. Existem duas abordagens gerais: *rotação ortogonal* e *rotação oblíqua*.

Separação completa: uma situação na *regressão logística* quando uma variável de saída pode ser perfeitamente prevista por um preditor ou uma combinação de preditores! Essa situação faz o seu computador ter o equivalente a um ataque de nervos: ele começará a gaguejar, chorar e dizer que ele não sabe o que fazer.

Simetria composta: uma condição que se torna verdadeira quando ambas as variâncias das condições são iguais (é o mesmo que a hipótese da *homogeneidade da variância*) e as *covariâncias* entre pares de condições são, também, iguais.

Sintaxe: comandos escritos predefinidos que instruem o SPSS a fazer o que você gostaria que ele fizesse (escrever “Me deixa em paz!” não funciona).

Soma dos erros ao quadrado: outro nome para a *soma dos quadrados*.

Soma dos quadrados (SS): uma estimativa do total da variabilidade (espalhamento) de um conjunto de dados. Em primeiro lugar, um *desvio* para cada escore é calculado e, depois, esse valor é elevado ao quadrado. A SS é a soma desses desvios elevados ao quadrado.

Soma dos quadrados do modelo: uma medida da variabilidade total que um modelo pode explicar. É a diferença entre a *soma dos quadrados total* e a *soma dos quadrados dos resíduos*.

Soma dos quadrados dos resíduos: uma medida da variabilidade que não pode ser explicada pelo modelo ajustado aos dados. Ela é o total dos *desvios* ao quadrado entre as observações e o valor dessas observações previsto por qualquer modelo que foi ajustado aos dados.

Soma dos quadrados e matriz dos produtos cruzados (SQPC): uma *matriz quadrada* em que os elementos diagonais representam a *soma dos quadrados* para uma variável específica e os elementos fora da diagonal representam os *produtos cruzados* entre os pares de variáveis. A matriz SQPC é basicamente a mesma que a *matriz das variâncias-covariâncias* exceto que ela expressa variabilidade e relacionamentos entre variáveis como valores totais enquanto a matriz de variâncias-covariâncias os expressa como valores médios.

Soma total dos quadrados: uma medida da variabilidade dentro de um conjunto de observações. Ela é soma total dos quadrados dos *desvios* entre cada observação e a média geral das observações.

Tabela de contingência: uma tabela representando uma classificação cruzada de duas ou mais *variáveis categóricas*. Os níveis de cada variável são alocados em uma grade e o número de observações em cada categoria é apresentado nas

células da tabela. Por exemplo, se tomarmos a variável categórica **glossário** (com duas categorias: se um autor escreveu um glossário ou não) e **estado mental** (com três categoriais: normal, soluçando incontrolavelmente e totalmente psicótico), podemos construir uma tabela como a mostrada a seguir. Isso nos informa que 127 autores que escreveram um glossário acabaram totalmente psicóticos, comparados com somente 2 que não escreveram um glossário.

Tamanho de efeito: uma medida objetiva e padronizada da magnitude de um efeito observado. As medidas incluem o *d* de Cohen, *g* de Glass e o coeficiente de correlação *r* de Pearson.

Tau de Kendall: um coeficiente de correlação não-paramétrico similar ao *coeficiente de correlação de Spearman*, mas deve ser preferido quando você tem um conjunto pequeno de dados com um número grande de postos empatados.

Taxa de erro com experimentos passo a passo: a probabilidade de cometer *erro do tipo I* em um experimento envolvendo uma ou mais comparações estatísticas quando a hipótese nula é verdadeira em cada caso.

Taxa de erro de conjunto: A probabilidade de cometer um *Erro do Tipo I* em uma família ou conjunto de testes quando a hipótese nula é verdadeira em cada caso. Uma “família de testes” pode ser definida como um conjunto de testes conduzidos sobre o mesmo conjunto de dados e tratando da mesma questão empírica.

Tendência cúbica: se for conectado às médias de condições ordenadas com uma linha, uma tendência cúbica é mostrada por duas mudanças na direção dessa linha. Deve-se ter pelo menos quatro condições ordenadas.

Tendência quadrática: se você conectar as médias em condições ordenadas com uma linha, a

		Glossário		
		Escreveu um glossário	Não escreveu um glossário	Total
Estado mental	Normal	5	423	428
	Soluçando incontrolavelmente	23	46	69
	Totalmente psicótico	127	2	129
	Total	155	471	626

tendência quadrática é caracterizada por uma mudança na direção dessa linha (por exemplo, a linha é curva em um lugar): a linha apresenta uma forma de U. Para que isso aconteça, é necessário, pelo menos, três condições ordenadas.

Tendência quártica: se você conectar as médias ordenadas com uma linha, uma tendência quártica é caracterizada por três mudanças na direção dessa linha. Você deve ter, pelo menos, cinco condições ordenadas.

Teste bilateral: um teste de uma hipótese não-direcional. Por exemplo, a hipótese “Escrever esse glossário tem algum efeito no que eu quero fazer com as genitais do meu editor” requer um teste bilateral porque ele não sugere a direção do relacionamento (veja também o *teste unilateral*).

Teste da mediana: um teste não-paramétrico para verificar se as amostras foram retiradas de uma população com a mesma mediana. Assim, ele faz o mesmo que o *teste de Kruskal-Wallis*. Ele produz uma tabela de contingência dividida em dois grupos de valores que estão acima ou abaixo da mediana. Se os grupos são da mesma população você espera que essas frequências sejam as mesmas em todas as condições (em torno de 50% acima e 50% abaixo).

Teste da soma dos postos de Wilcoxon: um *teste não-paramétrico* que procura por diferenças entre duas amostras independentes. Isto é, ele testa se as populações das quais duas amostras foram retiradas tem a mesma localização. Ele é funcionalmente o mesmo que o *teste de Mann-Whitney* e ambos os testes são os equivalentes não-paramétricos do *teste t independente*.

Teste de Box: um teste sobre a hipótese da *homogeneidade das matrizes de covariâncias*. Ele não deve ser significativo se as matrizes forem as mesmas. O teste de Box é bastante suscetível a desvios da *normalidade multivariada* e pode, então, ser não-significativo não porque as *matrizes de variâncias-covariâncias* sejam parecidas entre os grupos, mas porque a hipótese de normalidade multivariada não é satisfeita. Assim, é vital ter alguma ideia de se os dados satisfazem a hipótese de normalidade multivariada (que é bastante difícil) antes de interpretar os resultados do teste de Box.

Teste de Durbin-Watson: testa a correlação serial entre erros nos *modelos de regressão*. Especificamente, ele testa se resíduos adjacentes são correlacionados, o que é útil para testar a hipótese de *independência dos erros*. A estatística teste pode variar entre 0 e 4 com um valor de 2 indicando que os resíduos não são correlacionados. Um valor maior do que 2 indica uma correlação negativa e um valor abaixo de 2 indica uma correlação positiva. O tamanho da estatística de Durbin-Watson depende do número de previsores no modelo e do número de observações. Para ser preciso, você deve verificar os valores de aceitação no artigo original de Durbin e Watson de 1951. Como uma regra prática, bastante conservadora, valores menores do que 1 e maiores do que 3 são definitivamente motivo de atenção, no entanto, valores próximos de 2 podem ser problemáticos dependendo da amostra e do modelo.

Teste de esfericidade de Bartlett: um teste sobre a suposição de *esfericidade*. Ele examina se a *matriz de variâncias-covariâncias* é proporcional a uma *matriz identidade*. Dessa forma, ele testa se os elementos da diagonal da *matriz de variâncias-covariâncias* são iguais (isto é, se as variâncias dos grupos são iguais) e se os elementos fora da diagonal são aproximadamente zero (isto é, as *variáveis dependentes* não são *correlacionadas*). Jeremy Miles, que faz muitas análises multivariadas, declara que nunca viu uma matriz que não tenha tido significância quando submetida a esse teste, assim, questiona a sua utilidade prática.

Teste de Jonckheere-Terpstra: essa estatística testa um padrão ordenado de medianas entre grupos independentes. Essencialmente, ele faz o mesmo que o *teste de Kruskal-Wallis* (isto é, testa a diferença entre as medianas dos grupos), mas incorpora informação se a ordem do grupo é significativa. Assim, você deve usar esse teste quando espera que os grupos que está comparando produzam uma ordenação significativa das medianas.

Teste de Kolmogorov-Smirnov: um teste para ver se a distribuição dos escores é significativamente diferente de uma *distribuição normal*. Um valor significativo indica um desvio da normalidade, mas esse teste é notoriamente afetado por grandes amostras nas quais pequenos desvios da normalidade geram resultados significativos.

Teste de Kruskal-Wallis: um teste não-paramétrico para verificar se mais do que dois grupos independentes diferem. É a versão não-paramétrica da ANOVA independente de um fator.

Teste de Levene: testa a hipótese de que as variâncias em diferentes grupos são iguais (isto é, a diferença entre as variâncias é zero). Um resultado significativo indica que as variâncias são significativamente diferentes – portanto, a hipótese de *homogeneidade das variâncias* foi violada. Quando os tamanhos das amostras são grandes, pequenas diferenças em variâncias de grupo podem produzir um teste de Levene significativo, assim, a razão das variâncias é uma checagem dupla útil.

Teste de Mann-Whitney: um teste não-paramétrico que procura por diferenças entre duas amostras independentes. Isto é, ele testa se a população de onde as duas amostras foram retiradas tem a mesma localização. Ele é funcionalmente o mesmo que o teste da soma dos postos de Wilcoxon e ambos os testes são equivalentes não-paramétricos do teste *t* independente.

Teste de Mauchly: um teste da hipótese da *esfericidade*. Se esse teste for significativo, a hipótese da esfericidade não foi satisfeita e uma correção apropriada deve ser aplicada aos graus de liberdade da razão *F* na ANOVA de medidas repetidas. O teste compara a matriz de variâncias-covariâncias dos dados com uma matriz de identidade, se a matriz de variâncias-covariâncias é um múltiplo escalar de uma matriz de identidade, a esfericidade foi satisfeita.

Teste de McNemar: testa as diferenças entre dois grupos relacionados (veja o teste de postos com sinais de Wilcoxon e teste dos sinais), quando você tem dados nominais. Ele é tipicamente usado quando você está procurando por mudanças nos escores das pessoas e ele compara a proporção das pessoas que mudaram sua resposta em uma direção (isto é, aumento dos escores) àquelas que mudaram na direção oposta (diminuição dos escores). Assim, esse teste deve ser utilizado quando existem duas variáveis dicotômicas relacionadas.

Teste de Shapiro-Wilk: um teste para ver se a distribuição dos escores é significativamente diferente de uma distribuição normal. Um valor significativo indica um desvio da normalidade, mas esse teste é afetado por amostras grandes nas quais

pequenos desvios da normalidade geram resultados significativos.

Teste dos postos com sinais de Wilcoxon: um teste não-paramétrico que procura por diferenças entre duas amostras relacionadas. Ele é o não-paramétrico equivalente ao teste *t* relacionado.

Teste dos sinais: testa se duas populações são diferentes utilizando amostras relacionadas. Ele faz o mesmo que o teste dos postos com sinais de Wilcoxon. Diferenças entre as condições são calculadas e o sinal da diferença (positivo ou negativo) é analisado como indicador da direção das diferenças. A magnitude da mudança é completamente ignorada (diferente do teste de Wilcoxon, onde o posto revela algo sobre a relativa magnitude da mudança) e, por essa razão, ele apresenta perda de poder. Entretanto, sua simplicidade computacional faz dele um truque simples para ser usado em festas na eventualidade de algum bêbado abordar você querendo uma análise rápida de alguns dados sem a ajuda de um computador – realizar o teste dos sinais, de cabeça, realmente impressiona as pessoas. Na verdade, ele não impressiona, as pessoas irão achá-lo um bobo triste.

Teste paramétrico: um teste que requer dados de um grande catálogo de distribuições que os estatísticos descreveram. Normalmente, esse termo é usado para testes paramétricos baseados na distribuição normal que requer quatro hipóteses básicas que precisam ser encontradas para o teste ser preciso: dados normalmente distribuídos (veja distribuição normal), homogeneidade da variância, dados por intervalo ou de razão e independência.

Teste qui-quadrado: embora esse termo possa ser aplicado a qualquer estatística teste tendo uma distribuição qui-quadrado, ele geralmente se refere ao teste qui-quadrado de independência de Pearson entre duas variáveis categóricas. Essencialmente, ele testa se duas variáveis categóricas dispostas em uma tabela de contingência estão associadas.

Teste *t* dependente: um teste utilizando a estatística *t* que determina se duas médias coletadas de amostras relacionadas (emparelhadas) diferem significativamente.

Teste *t* independente: um teste usando a estatística *t* que estabelece se duas médias coletadas

de amostras independentes diferem de forma significativa.

Teste unilateral: um teste de hipóteses direcional. Por exemplo, a hipótese “quanto mais eu escrevo este glossário mais eu quero colocar os genitais do meu editor na boca de um crocodilo faminto” requer um teste unilateral porque estabeleci a direção do relacionamento (veja também *teste bilateral*).

Testes não-paramétricos: uma família de procedimentos estatísticos que não depende das hipóteses restritivas dos testes paramétricos. Especificamente, eles não presumem que os dados sejam provenientes de uma distribuição normal.

Testes *post hoc*: um conjunto de comparações entre as médias dos grupos que não foram planejadas antes dos dados serem coletados. Tipicamente, esses testes envolvem a comparação das médias de todas as combinações de pares das condições experimentais. Para compensar pelo número de testes conduzidos, cada teste usa um critério rígido para a significância. Assim, eles tendem a ter menos poder do que os *contrastes planejados*. Eles geralmente são usados em trabalhos exploratórios para os quais nenhuma hipótese firme estava disponível para ser base dos contrastes planejados.

Tolerância: a estatística de tolerância mensura a *multicolinearidade* e é simplesmente o recíproco do *fator da inflação da variância* (1/FIV). Valores abaixo de 0,1 indicam problemas sérios, embora Menard (1995) sugira que valores abaixo de 0,2 são dignos de preocupação.

Traço de Pillai-Bartlett (V): um teste estatístico na MANOVA. Ele é a soma da proporção da variância explicada na *função discriminante*. Assim, ele é semelhante à razão SS_M/SS_T .

Transformação: o processo de aplicar uma função matemática a todas as observações num conjunto de dados geralmente para corrigir alguma anormalidade da distribuição tal como a *assimetria* ou a *curtose*.

Uma vida: o que você não tem quando escreve um livro de Estatística.

Univariada: isso significa “uma variável” e é geralmente usado para referir a situações nas quais

somente uma variável de saída foi medida. (isto é, ANOVA, testes *t*, testes de Mann-Whitney, etc.).

V de Cramer: uma medida da força do relacionamento de uma associação entre duas *variáveis categóricas*. Utilizado quando uma das variáveis tem mais do que duas categorias. É uma variação do *fi*, usado quando uma ou ambas as variáveis contém mais de duas categorias, mas o *fi* falha ao não alcançar seu valor mínimo de 0 (indicando ausência de relação).

Validação cruzada: Determinação da precisão de um modelo utilizando amostras diferentes. Ele é uma etapa importante na *generalização*. Em um *modelo de regressão*, existem dois métodos principais de validação cruzada: R^2 ajustado ou divisão dos dados, em que os dados são aleatoriamente separados em metades e o modelo de regressão é estimado então com cada metade e os resultados comparados.

Valor atípico (*outlier*): uma observação muito diferente da maioria. Valores atípicos podem introduzir tendenciosidades como a média, por exemplo.

Valor Chapéu: outro nome para a Influência (*Leverage*).

Valor previsto ajustado: uma medida da influência de um caso específico de dados. É o valor previsto de um caso de um modelo estimado sem tal caso incluído nos dados. O valor é calculado estimando novamente o modelo sem o caso em questão e depois o modelo recalculado é utilizado para fazer uma previsão do caso excluído. Se o caso não exerce uma grande influência sobre o modelo, o valor previsto deve ser semelhante não importa se o modelo foi estimado incluindo ou excluindo tal caso. A diferença entre o valor previsto de um caso de um modelo quando o caso está incluído e o previsto quando ele está excluído é o DFFit.

Variância: uma estimativa da variabilidade média (espalhamento) de um conjunto de dados. Ela é a soma dos quadrados dividida pelo número dos valores menos 1.

Variação não-sistemática: variação que não é devida ao efeito no qual estamos interessados (assim, ela pode ser devido às diferenças naturais entre pessoas em amostras diferentes, como as diferenças em inteligência e motivação). Você pode

pensar nessa variação como sendo a que não pode ser explicada por qualquer modelo que ajustamos aos dados.

Varição sistemática: variação devido a algum efeito genuíno (seja esse efeito de um pesquisador fazendo algo a todos os participantes em uma amostra, mas não em outras amostras ou a variação natural entre conjuntos de variáveis). Pense nisso como a variação que pode ser explicada pelo modelo que ajustamos aos dados.

Variância aleatória: uma variância única a uma determinada variável, mas que não é muito confiável.

Variância comum: variância compartilhada por duas ou mais variáveis.

Variância total: a *variância* dentro de um conjunto completo de observações.

Variância única: variância específica a uma determinada variável (isto é, não é dividida com outras variáveis). Temos a tendência de usar o termo “variância única” para nos referirmos à variância que pode ser atribuída com confiança a somente uma medida, de outro modo ela é chamada de *variância aleatória*.

Variate função discriminante: uma combinação linear de variáveis criadas de modo que a diferença entre as médias dos grupos da variável transformada é maximizada. Ela apresenta a seguinte forma geral: $Variate_i = b_1X_1 + b_2X_2 + b_nX_n$. É uma combinação linear (CL) de variáveis.

Variáveis de texto: variáveis que envolvem palavras (isto é, texto). Tais variáveis podem incluir respostas para questões abertas como: “O quanto você gosta de escrever um glossário?” a resposta poderia ser “Tanto quanto colocar meus testículos no fogo”.

Variáveis numéricas: variáveis envolvendo números.

Variável auxiliar (*dummy*): uma maneira de recodificar uma variável categórica com mais de duas categorias em uma série de variáveis *dicotômicas* e que podem assumir apenas os valores 0 ou 1. Existem sete passos básicos para criar tais variáveis: (1) conte o número de grupos que você quer recodificar e subtraia; (2) crie tantas variá-

veis quanto o valor que foi calculado no passo 1 (essas são as variáveis *dummy*); (3) escolha um dos grupos como base (isto é, um grupo contra o qual todos os demais serão comparados, como um grupo-controle); (4) atribua ao grupo-base o valor 0 para todas as variáveis auxiliares; (5) para a primeira variável auxiliar, atribua 1 para o primeiro grupo que você quer comparar contra o grupo-base (atribua em todos os outros grupos 0 para essa variável); (7) repita esse processo até não existirem mais variáveis auxiliares.

Variável categórica: qualquer variável apresentada em categorias de objetos/entidades. O conceito é um exemplo de tal tipo de variável. Uma apresentação típica é: A, B, C, D e E, onde D e E significam reprovação e os demais, aprovação.

Variável de confusão: uma variável (que pode ou não ter sido mensurada) além das *variáveis previsoras* em que estamos interessados que pode potencialmente afetar a *variável de saída*.

Variável de saída: uma variável cujos valores estamos tentando prever a partir de uma ou mais *variáveis previsoras*.

Variável dependente: outro nome para a *variável de saída*. Esse nome está normalmente associado à metodologia experimental (que é a única vez em que ela realmente faz sentido) e é assim denominada em virtude de ser a variável que não é manipulada pelo pesquisador e assim seus valores dependem das variáveis que foram manipuladas. Para ser honesto, eu sempre utilizo a variável de saída – faz mais sentido (para mim) e é menos confuso.

Variável independente: outro nome para uma *variável previsora*. Esse nome é geralmente associado à metodologia experimental (que é a única ocasião em que ele faz sentido) e é assim chamado porque ela é a variável manipulada pelo pesquisador, portanto, seu valor não depende de qualquer outra variável (apenas do pesquisador). Uso a *variável previsora* sempre porque o significado do termo não é restrito a uma metodologia específica.

Variável latente: uma variável que não pode ser mensurada diretamente, mas que supostamente está relacionada a muitas variáveis que podem ser mensuradas.

Variável previsor: uma variável usada para tentar prever valores de outra variável conhecida como *variável da saída*.

Varimax: um método de *rotação ortogonal*. Ele tenta maximizar a dispersão das *cargas dos fatores* dentro dos *fatores*. Portanto, ele tenta carregar um número pequeno de variáveis em cada fator resultando em uma melhor interpretação de aglomerados de fatores.

Verossimilhança log: uma medida do erro ou da variação não-explicada em modelos categóricos. Ela é baseada na soma das probabilidades associada às saídas previstas e reais e é análoga à *soma dos quadrados residuais* na regressão múltipla na qual ela é um indicador de quanta informação não-explicada há após o modelo ter sido ajustado. Valores altos da verossimilhança log indicam modelos estatísticos pouco ajustados porque quanto maior o valor da probabilidade log, mais observações não-explicadas existem. A verossimilhança log é o logaritmo da *verossimilhança*.

Verossimilhança: a probabilidade de obter um conjunto de observações dado os parâmetros de um modelo ajustado a essas observações.

Visualizador da variável: existem duas maneiras de ver os conteúdos da janela do *editor de dados*. O visualizador de variáveis permite definir propriedades das variáveis para as quais você quer entrar com dados (veja também visualizador de dados).

Visualizador de dados (Data view): existem duas maneiras de visualizar os dados na janela de edição de dados (*data editor window*). A janela de visualização de dados mostra uma planilha que pode ser utilizada para entrar com os dados (veja também *visualizador de variáveis*).

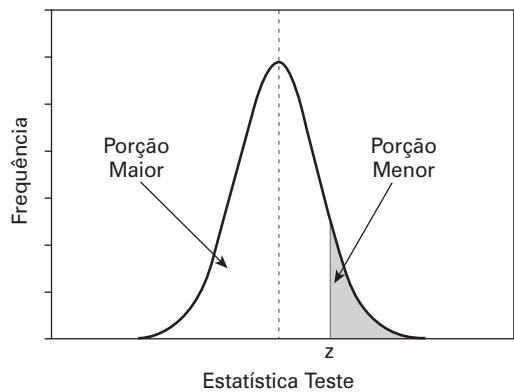
Visualizador: janela do SPSS na qual a saída de qualquer análise é apresentada.

W de Kendall: É muito parecido com a *ANOVA de Friedman*, mas é usado especialmente para observar a concordância entre avaliadores. Assim, se, por exemplo, pedimos a 10 mulheres diferentes avaliarem a atratividade de Justin Timberlake, David Beckham e Tony Blair, podemos usar esse teste para analisar a extensão da concordância. O *W* de Kendall varia de 0 (nenhuma concordância entre os avaliadores) a 1 (concordância completa entre os avaliadores).

Z de Kolmogorov-Smirnov: não deve ser confundido com o *teste de Kolmogorov-Smirnov* que testa se uma amostra vem de uma população normalmente distribuída. Esse testa se dois grupos foram retirados da mesma população (independentemente de qual seja a população). Ele faz praticamente o mesmo que o *teste de Mann-Whitney* e o *teste da soma dos postos de Wilcoxon*! Este teste tende a ter um poder maior do que o teste de Mann-Whitney quando o tamanho das amostras é menor do que 25 por grupo.

APÊNDICE

A.1 TABELA DA DISTRIBUIÇÃO NORMAL PADRÃO



z	Larger Portion	Smaller Portion	y
0.00	0.50000	0.50000	0.3989
0.01	0.50399	0.49601	0.3989
0.02	0.50798	0.49202	0.3989
0.03	0.51197	0.48803	0.3988
0.04	0.51595	0.48405	0.3986
0.05	0.51994	0.48006	0.3984
0.06	0.52392	0.47608	0.3982
0.07	0.52790	0.47210	0.3980
0.08	0.53188	0.46812	0.3977
0.09	0.53586	0.46414	0.3973
0.10	0.53983	0.46017	0.3970
0.11	0.54380	0.45620	0.3965
0.12	0.54776	0.45224	0.3961
0.13	0.55172	0.44828	0.3956
0.14	0.55567	0.44433	0.3951
0.15	0.55962	0.44038	0.3945
0.16	0.56356	0.43644	0.3939
0.17	0.56749	0.43251	0.3932
0.18	0.57142	0.42858	0.3925

z	Larger Portion	Smaller Portion	y
0.19	0.57535	0.42465	0.3918
0.20	0.57926	0.42074	0.3910
0.21	0.58317	0.41683	0.3902
0.22	0.58706	0.41294	0.3894
0.23	0.59095	0.40905	0.3885
0.24	0.59483	0.40517	0.3876
0.25	0.59871	0.40129	0.3867
0.26	0.60257	0.39743	0.3857
0.27	0.60642	0.39358	0.3847
0.28	0.61026	0.38974	0.3836
0.29	0.61409	0.38591	0.3825
0.30	0.61791	0.38209	0.3814
0.31	0.62172	0.37828	0.3802
0.32	0.62552	0.37448	0.3790
0.33	0.62930	0.37070	0.3778
0.34	0.63307	0.36693	0.3765
0.35	0.63683	0.36317	0.3752
0.36	0.64058	0.35942	0.3739
0.37	0.64431	0.35569	0.3725

(Continua)

z	Larger Portion	Smaller Portion	y
0.38	0.64803	0.35197	0.3712
0.39	0.65173	0.34827	0.3697
0.40	0.65542	0.34458	0.3683
0.41	0.65910	0.34090	0.3668
0.42	0.66276	0.33724	0.3653
0.43	0.66640	0.33360	0.3637
0.44	0.67003	0.32997	0.3621
0.45	0.67364	0.32636	0.3605
0.46	0.67724	0.32276	0.3589
0.47	0.68082	0.31918	0.3572
0.48	0.68439	0.31561	0.3555
0.49	0.68793	0.31207	0.3538
0.50	0.69146	0.30854	0.3521
0.51	0.69497	0.30503	0.3503
0.52	0.69847	0.30153	0.3485
0.53	0.70194	0.29806	0.3467
0.54	0.70540	0.29460	0.3448
0.55	0.70884	0.29116	0.3429
0.56	0.71226	0.28774	0.3410
0.57	0.71566	0.28434	0.3391
0.58	0.71904	0.28096	0.3372
0.59	0.72240	0.27760	0.3352
0.60	0.72575	0.27425	0.3332
0.61	0.72907	0.27093	0.3312
0.62	0.73237	0.26763	0.3292
0.63	0.73565	0.26435	0.3271
0.64	0.73891	0.26109	0.3251
0.65	0.74215	0.25785	0.3230
0.66	0.74537	0.25463	0.3209
0.67	0.74857	0.25143	0.3187
0.68	0.75175	0.24825	0.3166
0.69	0.75490	0.24510	0.3144
0.70	0.75804	0.24196	0.3123
0.71	0.76115	0.23885	0.3101
0.72	0.76424	0.23576	0.3079
0.73	0.76730	0.23270	0.3056
0.74	0.77035	0.22965	0.3034
0.75	0.77337	0.22663	0.3011
0.76	0.77637	0.22363	0.2989
0.77	0.77935	0.22065	0.2966
0.78	0.78230	0.21770	0.2943
0.79	0.78524	0.21476	0.2920
0.80	0.78814	0.21186	0.2897
0.81	0.79103	0.20897	0.2874
0.82	0.79389	0.20611	0.2850
0.83	0.79673	0.20327	0.2827
0.84	0.79955	0.20045	0.2803
0.85	0.80234	0.19766	0.2780
0.86	0.80511	0.19489	0.2756
0.87	0.80785	0.19215	0.2732
0.88	0.81057	0.18943	0.2709
0.89	0.81327	0.18673	0.2685
0.90	0.81594	0.18406	0.2661
0.91	0.81859	0.18141	0.2637
0.92	0.82121	0.17879	0.2613
0.93	0.82381	0.17619	0.2589
0.94	0.82639	0.17361	0.2565
0.95	0.82894	0.17106	0.2541
0.96	0.83147	0.16853	0.2516
0.97	0.83398	0.16602	0.2492

z	Larger Portion	Smaller Portion	y
0.98	0.83646	0.16354	0.2468
0.99	0.83891	0.16109	0.2444
1.00	0.84134	0.15866	0.2420
1.01	0.84375	0.15625	0.2396
1.02	0.84614	0.15386	0.2371
1.03	0.84849	0.15151	0.2347
1.04	0.85083	0.14917	0.2323
1.05	0.85314	0.14686	0.2299
1.06	0.85543	0.14457	0.2275
1.07	0.85769	0.14231	0.2251
1.08	0.85993	0.14007	0.2227
1.09	0.86214	0.13786	0.2203
1.10	0.86433	0.13567	0.2179
1.11	0.86650	0.13350	0.2155
1.12	0.86864	0.13136	0.2131
1.13	0.87076	0.12924	0.2107
1.14	0.87286	0.12714	0.2083
1.15	0.87493	0.12507	0.2059
1.16	0.87698	0.12302	0.2036
1.17	0.87900	0.12100	0.2012
1.18	0.88100	0.11900	0.1989
1.19	0.88298	0.11702	0.1965
1.20	0.88493	0.11507	0.1942
1.21	0.88686	0.11314	0.1919
1.22	0.88877	0.11123	0.1895
1.23	0.89065	0.10935	0.1872
1.24	0.89251	0.10749	0.1849
1.25	0.89435	0.10565	0.1826
1.26	0.89617	0.10383	0.1804
1.27	0.89796	0.10204	0.1781
1.28	0.89973	0.10027	0.1758
1.29	0.90147	0.09853	0.1736
1.30	0.90320	0.09680	0.1714
1.31	0.90490	0.09510	0.1691
1.32	0.90658	0.09342	0.1669
1.33	0.90824	0.09176	0.1647
1.34	0.90988	0.09012	0.1626
1.35	0.91149	0.08851	0.1604
1.36	0.91309	0.08691	0.1582
1.37	0.91466	0.08534	0.1561
1.38	0.91621	0.08379	0.1539
1.39	0.91774	0.08226	0.1518
1.40	0.91924	0.08076	0.1497
1.41	0.92073	0.07927	0.1476
1.42	0.92220	0.07780	0.1456
1.43	0.92364	0.07636	0.1435
1.44	0.92507	0.07493	0.1415
1.45	0.92647	0.07353	0.1394
1.46	0.92785	0.07215	0.1374
1.47	0.92922	0.07078	0.1354
1.48	0.93056	0.06944	0.1334
1.49	0.93189	0.06811	0.1315
1.50	0.93319	0.06681	0.1295
1.51	0.93448	0.06552	0.1276
1.52	0.93574	0.06426	0.1257
1.53	0.93699	0.06301	0.1238
1.54	0.93822	0.06178	0.1219
1.55	0.93943	0.06057	0.1200
1.56	0.94062	0.05938	0.1182
1.57	0.94179	0.05821	0.1163

(Continua)

z	Larger Portion	Smaller Portion	y	z	Larger Portion	Smaller Portion	y
1.58	0.94295	0.05705	0.1145	2.18	0.98537	0.01463	0.0371
1.59	0.94408	0.05592	0.1127	2.19	0.98574	0.01426	0.0363
1.60	0.94520	0.05480	0.1109	2.20	0.98610	0.01390	0.0355
1.61	0.94630	0.05370	0.1092	2.21	0.98645	0.01355	0.0347
1.62	0.94738	0.05262	0.1074	2.22	0.98679	0.01321	0.0339
1.63	0.94845	0.05155	0.1057	2.23	0.98713	0.01287	0.0332
1.64	0.94950	0.05050	0.1040	2.24	0.98745	0.01255	0.0325
1.65	0.95053	0.04947	0.1023	2.25	0.98778	0.01222	0.0317
1.66	0.95154	0.04846	0.1006	2.26	0.98809	0.01191	0.0310
1.67	0.95254	0.04746	0.0989	2.27	0.98840	0.01160	0.0303
1.68	0.95352	0.04648	0.0973	2.28	0.98870	0.01130	0.0297
1.69	0.95449	0.04551	0.0957	2.29	0.98899	0.01101	0.0290
1.70	0.95543	0.04457	0.0940	2.30	0.98928	0.01072	0.0283
1.71	0.95637	0.04363	0.0925	2.31	0.98956	0.01044	0.0277
1.72	0.95728	0.04272	0.0909	2.32	0.98983	0.01017	0.0270
1.73	0.95818	0.04182	0.0893	2.33	0.99010	0.00990	0.0264
1.74	0.95907	0.04093	0.0878	2.34	0.99036	0.00964	0.0258
1.75	0.95994	0.04006	0.0863	2.35	0.99061	0.00939	0.0252
1.76	0.96080	0.03920	0.0848	2.36	0.99086	0.00914	0.0246
1.77	0.96164	0.03836	0.0833	2.37	0.99111	0.00889	0.0241
1.78	0.96246	0.03754	0.0818	2.38	0.99134	0.00866	0.0235
1.79	0.96327	0.03673	0.0804	2.39	0.99158	0.00842	0.0229
1.80	0.96407	0.03593	0.0790	2.40	0.99180	0.00820	0.0224
1.81	0.96485	0.03515	0.0775	2.41	0.99202	0.00798	0.0219
1.82	0.96562	0.03438	0.0761	2.42	0.99224	0.00776	0.0213
1.83	0.96638	0.03362	0.0748	2.43	0.99245	0.00755	0.0208
1.84	0.96712	0.03288	0.0734	2.44	0.99266	0.00734	0.0203
1.85	0.96784	0.03216	0.0721	2.45	0.99286	0.00714	0.0198
1.86	0.96856	0.03144	0.0707	2.46	0.99305	0.00695	0.0194
1.87	0.96926	0.03074	0.0694	2.47	0.99324	0.00676	0.0189
1.88	0.96995	0.03005	0.0681	2.48	0.99343	0.00657	0.0184
1.89	0.97062	0.02938	0.0669	2.49	0.99361	0.00639	0.0180
1.90	0.97128	0.02872	0.0656	2.50	0.99379	0.00621	0.0175
1.91	0.97193	0.02807	0.0644	2.51	0.99396	0.00604	0.0171
1.92	0.97257	0.02743	0.0632	2.52	0.99413	0.00587	0.0167
1.93	0.97320	0.02680	0.0620	2.53	0.99430	0.00570	0.0163
1.94	0.97381	0.02619	0.0608	2.54	0.99446	0.00554	0.0158
1.95	0.97441	0.02559	0.0596	2.55	0.99461	0.00539	0.0154
1.96	0.97500	0.02500	0.0584	2.56	0.99477	0.00523	0.0151
1.97	0.97558	0.02442	0.0573	2.57	0.99492	0.00508	0.0147
1.98	0.97615	0.02385	0.0562	2.58	0.99506	0.00494	0.0143
1.99	0.97670	0.02330	0.0551	2.59	0.99520	0.00480	0.0139
2.00	0.97725	0.02275	0.0540	2.60	0.99534	0.00466	0.0136
2.01	0.97778	0.02222	0.0529	2.61	0.99547	0.00453	0.0132
2.02	0.97831	0.02169	0.0519	2.62	0.99560	0.00440	0.0129
2.03	0.97882	0.02118	0.0508	2.63	0.99573	0.00427	0.0126
2.04	0.97932	0.02068	0.0498	2.64	0.99585	0.00415	0.0122
2.05	0.97982	0.02018	0.0488	2.65	0.99598	0.00402	0.0119
2.06	0.98030	0.01970	0.0478	2.66	0.99609	0.00391	0.0116
2.07	0.98077	0.01923	0.0468	2.67	0.99621	0.00379	0.0113
2.08	0.98124	0.01876	0.0459	2.68	0.99632	0.00368	0.0110
2.09	0.98169	0.01831	0.0449	2.69	0.99643	0.00357	0.0107
2.10	0.98214	0.01786	0.0440	2.70	0.99653	0.00347	0.0104
2.11	0.98257	0.01743	0.0431	2.71	0.99664	0.00336	0.0101
2.12	0.98300	0.01700	0.0422	2.72	0.99674	0.00326	0.0099
2.13	0.98341	0.01659	0.0413	2.73	0.99683	0.00317	0.0096
2.14	0.98382	0.01618	0.0404	2.74	0.99693	0.00307	0.0093
2.15	0.98422	0.01578	0.0396	2.75	0.99702	0.00298	0.0091
2.16	0.98461	0.01539	0.0387	2.76	0.99711	0.00289	0.0088
2.17	0.98500	0.01500	0.0379	2.77	0.99720	0.00280	0.0086

(Continua)

z	Larger Portion	Smaller Portion	y
2.78	0.99728	0.00272	0.0084
2.79	0.99736	0.00264	0.0081
2.80	0.99744	0.00256	0.0079
2.81	0.99752	0.00248	0.0077
2.82	0.99760	0.00240	0.0075
2.83	0.99767	0.00233	0.0073
2.84	0.99774	0.00226	0.0071
2.85	0.99781	0.00219	0.0069
2.86	0.99788	0.00212	0.0067
2.87	0.99795	0.00205	0.0065
2.88	0.99801	0.00199	0.0063
2.89	0.99807	0.00193	0.0061
2.90	0.99813	0.00187	0.0060
2.91	0.99819	0.00181	0.0058
2.92	0.99825	0.00175	0.0056

z	Larger Portion	Smaller Portion	y
2.93	0.99831	0.00169	0.0055
2.94	0.99836	0.00164	0.0053
2.95	0.99841	0.00159	0.0051
2.96	0.99846	0.00154	0.0050
2.97	0.99851	0.00149	0.0048
2.98	0.99856	0.00144	0.0047
2.99	0.99861	0.00139	0.0046
3.00	0.99865	0.00135	0.0044
⋮	⋮	⋮	⋮
3.25	0.99942	0.00058	0.0020
⋮	⋮	⋮	⋮
3.50	0.99977	0.00023	0.0009
⋮	⋮	⋮	⋮
4.00	0.99997	0.00003	0.0001

A.2 VALORES CRÍTICOS DA DISTRIBUIÇÃO T

<i>gl</i>	Two-Tailed Test (Teste Bilateral)		One-Tailed Test (Teste Unilateral)	
	<i>0.05</i>	<i>0.01</i>	<i>0.05</i>	<i>0.01</i>
1	12.71	63.66	6.31	31.82
2	4.30	9.92	2.92	6.96
3	3.18	5.84	2.35	4.54
4	2.78	4.60	2.13	3.75
5	2.57	4.03	2.02	3.36
6	2.45	3.71	1.94	3.14
7	2.36	3.50	1.89	3.00
8	2.31	3.36	1.86	2.90
9	2.26	3.25	1.83	2.82
10	2.23	3.17	1.81	2.76
11	2.20	3.11	1.80	2.72
12	2.18	3.05	1.78	2.68
13	2.16	3.01	1.77	2.65
14	2.14	2.98	1.76	2.62
15	2.13	2.95	1.75	2.60
16	2.12	2.92	1.75	2.58
17	2.11	2.90	1.74	2.57
18	2.10	2.88	1.73	2.55
19	2.09	2.86	1.73	2.54
20	2.09	2.85	1.72	2.53
21	2.08	2.83	1.72	2.52
22	2.07	2.82	1.72	2.51
23	2.07	2.81	1.71	2.50
24	2.06	2.80	1.71	2.49
25	2.06	2.79	1.71	2.49
26	2.06	2.78	1.71	2.48
27	2.05	2.77	1.70	2.47
28	2.05	2.76	1.70	2.47
29	2.05	2.76	1.70	2.46
30	2.04	2.75	1.70	2.46
35	2.03	2.72	1.69	2.44
40	2.02	2.70	1.68	2.42
45	2.01	2.69	1.68	2.41
50	2.01	2.68	1.68	2.40
60	2.00	2.66	1.67	2.39
70	1.99	2.65	1.67	2.38
80	1.99	2.64	1.66	2.37
90	1.99	2.63	1.66	2.37
100	1.98	2.63	1.66	2.36
$\infty(z)$	1.96	2.58	1.64	2.33

Todos os valores foram calculados pelo autor usando o SPSS 11.

A.3 VALORES CRÍTICOS DA DISTRIBUIÇÃO F

		gl (numerador)										
	p	1	2	3	4	5	6	7	8	9	10	
gl (denominador)	1	0.05 0.01	161.45 4052.18	199.50 4999.50	215.71 5403.35	224.58 5624.58	230.16 5763.65	233.99 5858.99	236.77 5928.36	238.88 5981.07	240.54 6022.47	241.88 6055.85
	2	0.05 0.01	18.51 98.50	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.35 99.36	19.37 99.37	19.38 99.39	19.40 99.40
	3	0.05 0.01	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.89 27.67	8.85 27.49	8.81 27.35	8.79 27.23
	4	0.05 0.01	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.55
	5	0.05 0.01	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.46	4.82 10.29	4.77 10.16	4.74 10.05
	6	0.05 0.01	5.99 13.75	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87
	7	0.05 0.01	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 6.99	3.73 6.84	3.68 6.72	3.64 6.62
	8	0.05 0.01	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.18	3.44 6.03	3.39 5.91	3.35 5.81
	9	0.05 0.01	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.61	3.23 5.47	3.18 5.35	3.14 5.26
	10	0.05 0.01	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.20	3.07 5.06	3.02 4.94	2.98 4.85
	11	0.05 0.01	4.84 9.65	3.98 7.21	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.89	2.95 4.74	2.90 4.63	2.85 4.54
	12	0.05 0.01	4.75 9.33	3.89 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.91 4.64	2.85 4.50	2.80 4.39	2.75 4.30
	13	0.05 0.01	4.67 9.07	3.81 6.70	3.41 5.74	3.18 5.21	3.03 4.86	2.92 4.62	2.83 4.44	2.77 4.30	2.71 4.19	2.67 4.10
	14	0.05 0.01	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.04	2.96 4.69	2.85 4.46	2.76 4.28	2.70 4.14	2.65 4.03	2.60 3.94
	15	0.05 0.01	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.71 4.14	2.64 4.00	2.59 3.89	2.54 3.80
	16	0.05 0.01	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69
	17	0.05 0.01	4.45 8.40	3.59 6.11	3.20 5.18	2.96 4.67	2.81 4.34	2.70 4.10	2.61 3.93	2.55 3.79	2.49 3.68	2.45 3.59
	18	0.05 0.01	4.41 8.29	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.84	2.51 3.71	2.46 3.60	2.41 3.51
	19	0.05 0.01	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.54 3.77	2.48 3.63	2.42 3.52	2.38 3.43
	20	0.05 0.01	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.51 3.70	2.45 3.56	2.39 3.46	2.35 3.37
	22	0.05 0.01	4.30 7.95	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.46 3.59	2.40 3.45	2.34 3.35	2.30 3.26

(Continua)

		gl (numerador)										
	p	1	2	3	4	5	6	7	8	9	10	
gl (denominador)	24	0.05 0.01	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.42 3.50	2.36 3.36	2.30 3.26	2.25 3.17
	26	0.05 0.01	4.23 7.72	3.37 5.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.42	2.32 3.29	2.27 3.18	2.22 3.09
	28	0.05 0.01	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.75	2.45 3.53	2.36 3.36	2.29 3.23	2.24 3.12	2.19 3.03
	30	0.05 0.01	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.33 3.30	2.27 3.17	2.21 3.07	2.16 2.98
	35	0.05 0.01	4.12 7.42	3.27 5.27	2.87 4.40	2.64 3.91	2.49 3.59	2.37 3.37	2.29 3.20	2.22 3.07	2.16 2.96	2.11 2.88
	40	0.05 0.01	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.89	2.08 2.80
	45	0.05 0.01	4.06 7.23	3.20 5.11	2.81 4.25	2.58 3.77	2.42 3.45	2.31 3.23	2.22 3.07	2.15 2.94	2.10 2.83	2.05 2.74
	50	0.05 0.01	4.03 7.17	3.18 5.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.19	2.20 3.02	2.13 2.89	2.07 2.78	2.03 2.70
	60	0.05 0.01	4.00 7.08	3.15 4.98	2.76 4.13	2.53 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63
	80	0.05 0.01	3.96 6.96	3.11 4.88	2.72 4.04	2.49 3.56	2.33 3.26	2.21 3.04	2.13 2.87	2.06 2.74	2.00 2.64	1.95 2.55
	100	0.05 0.01	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.31 3.21	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.93 2.50
	150	0.05 0.01	3.90 6.81	3.06 4.75	2.66 3.91	2.43 3.45	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.63	1.94 2.53	1.89 2.44
	300	0.05 0.01	3.87 6.72	3.03 4.68	2.63 3.85	2.40 3.38	2.24 3.08	2.13 2.86	2.04 2.70	1.97 2.57	1.91 2.47	1.86 2.38
	500	0.05 0.01	3.86 6.69	3.01 4.65	2.62 3.82	2.39 3.36	2.23 3.05	2.12 2.84	2.03 2.68	1.96 2.55	1.90 2.44	1.85 2.36
	1000	0.05 0.01	3.85 6.66	3.00 4.63	2.61 3.80	2.38 3.34	2.22 3.04	2.11 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34

		gl (numerador)							
		p	15	20	25	30	40	50	1000
gl (denominador)	1	0.05	245.95	248.01	249.26	250.10	251.14	251.77	254.19
		0.01	6157.31	6208.74	6239.83	6260.65	6286.79	6302.52	6362.70
	2	0.05	19.43	19.45	19.46	19.46	19.47	19.48	19.49
		0.01	99.43	99.45	99.46	99.47	99.47	99.48	99.50
	3	0.05	8.70	8.66	8.63	8.62	8.59	8.58	8.53
		0.01	26.87	26.69	26.58	26.50	26.41	26.35	26.14
	4	0.05	5.86	5.80	5.77	5.75	5.72	5.70	5.63
		0.01	14.20	14.02	13.91	13.84	13.75	13.69	13.47
	5	0.05	4.62	4.56	4.52	4.50	4.46	4.44	4.37
		0.01	9.72	9.55	9.45	9.38	9.29	9.24	9.03

(Continua)

		gl (numerador)							
	p	15	20	25	30	40	50	1000	
gl (denominador)	6	0.05 0.01	3.94 7.56	3.87 7.40	3.83 7.30	3.81 7.23	3.77 7.14	3.75 7.09	3.67 6.89
	7	0.05 0.01	3.51 6.31	3.44 6.16	3.40 6.06	3.38 5.99	3.34 5.91	3.32 5.86	3.23 5.66
	8	0.05 0.01	3.22 5.52	3.15 5.36	3.11 5.26	3.08 5.20	3.04 5.12	3.02 5.07	2.93 4.87
	9	0.05 0.01	3.01 4.96	2.94 4.81	2.89 4.71	2.86 4.65	2.83 4.57	2.80 4.52	2.71 4.32
	10	0.05 0.01	2.85 4.56	2.77 4.41	2.73 4.31	2.70 4.25	2.66 4.17	2.64 4.12	2.54 3.92
	11	0.05 0.01	2.72 4.25	2.65 4.10	2.60 4.01	2.57 3.94	2.53 3.86	2.51 3.81	2.41 3.61
	12	0.05 0.01	2.62 4.01	2.54 3.86	2.50 3.76	2.47 3.70	2.43 3.62	2.40 3.57	2.30 3.37
	13	0.05 0.01	2.53 3.82	2.46 3.66	2.41 3.57	2.38 3.51	2.34 3.43	2.31 3.38	2.21 3.18
	14	0.05 0.01	2.46 3.66	2.39 3.51	2.34 3.41	2.31 3.35	2.27 3.27	2.24 3.22	2.14 3.02
	15	0.05 0.01	2.40 3.52	2.33 3.37	2.28 3.28	2.25 3.21	2.20 3.13	2.18 3.08	2.07 2.88
	16	0.05 0.01	2.35 3.41	2.28 3.26	2.23 3.16	2.19 3.10	2.15 3.02	2.12 2.97	2.02 2.76
	17	0.05 0.01	2.31 3.31	2.23 3.16	2.18 3.07	2.15 3.00	2.10 2.92	2.08 2.87	1.97 2.66
	18	0.05 0.01	2.27 3.23	2.19 3.08	2.14 2.98	2.11 2.92	2.06 2.84	2.04 2.78	1.92 2.58
	19	0.05 0.01	2.23 3.15	2.16 3.00	2.11 2.91	2.07 2.84	2.03 2.76	2.00 2.71	1.88 2.50
	20	0.05 0.01	2.20 3.09	2.12 2.94	2.07 2.84	2.04 2.78	1.99 2.69	1.97 2.64	1.85 2.43
	22	0.05 0.01	2.15 2.98	2.07 2.83	2.02 2.73	1.98 2.67	1.94 2.58	1.91 2.53	1.79 2.32
	24	0.05 0.01	2.11 2.89	2.03 2.74	1.97 2.64	1.94 2.58	1.89 2.49	1.86 2.44	1.74 2.22
	26	0.05 0.01	2.07 2.81	1.99 2.66	1.94 2.57	1.90 2.50	1.85 2.42	1.82 2.36	1.70 2.14
	28	0.05 0.01	2.04 2.75	1.96 2.60	1.91 2.51	1.87 2.44	1.82 2.35	1.79 2.30	1.66 2.08
	30	0.05 0.01	2.01 2.70	1.93 2.55	1.88 2.45	1.84 2.39	1.79 2.30	1.76 2.25	1.63 2.02
	35	0.05 0.01	1.96 2.60	1.88 2.44	1.82 2.35	1.79 2.28	1.74 2.19	1.70 2.14	1.57 1.90
	40	0.05 0.01	1.92 2.52	1.84 2.37	1.78 2.27	1.74 2.20	1.69 2.11	1.66 2.06	1.52 1.82

(Continua)

		gl (numerador)							
	p	15	20	25	30	40	50	1000	
gl (denominador)	45	0.05 0.01	1.89 2.46	1.81 2.31	1.75 2.21	1.71 2.14	1.66 2.05	1.63 2.00	1.48 1.75
	50	0.05 0.01	1.87 2.42	1.78 2.27	1.73 2.17	1.69 2.10	1.63 2.01	1.60 1.95	1.45 1.70
	60	0.05 0.01	1.84 2.35	1.75 2.20	1.69 2.10	1.65 2.03	1.59 1.94	1.56 1.88	1.40 1.62
	80	0.05 0.01	1.79 2.27	1.70 2.12	1.64 2.01	1.60 1.94	1.54 1.85	1.51 1.79	1.34 1.51
	100	0.05 0.01	1.77 2.22	1.68 2.07	1.62 1.97	1.57 1.89	1.52 1.80	1.48 1.74	1.30 1.45
	150	0.05 0.01	1.73 2.16	1.64 2.00	1.58 1.90	1.54 1.83	1.48 1.73	1.44 1.66	1.24 1.35
	300	0.05 0.01	1.70 2.10	1.61 1.94	1.54 1.84	1.50 1.76	1.43 1.66	1.39 1.59	1.17 1.25
	500	0.05 0.01	1.69 2.07	1.59 1.92	1.53 1.81	1.48 1.74	1.42 1.63	1.38 1.57	1.14 1.20
	1000	0.05 0.01	1.68 2.06	1.58 1.90	1.52 1.79	1.47 1.72	1.41 1.61	1.36 1.54	1.11 1.16
	Todos os valores foram calculados pelo autor usando o SPSS 11.								

A.4 VALORES CRÍTICOS DA DISTRIBUIÇÃO QUI-QUADRADO

gl	p	
	0.05	0.01
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22
13	22.36	27.69
14	23.68	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.81
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98

gl	p	
	0.05	0.01
25	37.65	44.31
26	38.89	45.64
27	40.11	46.96
28	41.34	48.28
29	42.56	49.59
30	43.77	50.89
35	49.80	57.34
40	55.76	63.69
45	61.66	69.96
50	67.50	76.15
60	79.08	88.38
70	90.53	100.43
80	101.88	112.33
90	113.15	124.12
100	124.34	135.81
200	233.99	249.45
300	341.40	359.91
400	447.63	468.72
500	553.13	576.49
600	658.09	683.52
700	762.66	789.97
800	866.91	895.98
900	970.90	1001.63
1000	1074.68	1106.97

A.5 O TESTE F DE WELCH

Veja o arquivo **Welch F.pdf** no *site* www.artmed.com.br.

A.6 CALCULANDO EFEITOS SIMPLES

Veja o arquivo **Calculating Simple Effects.pdf** no *site* www.artmed.com.br.

A.7 TESTE DA TENDÊNCIA DE JONCKHEERE

Veja o arquivo **Jonckheere.pdf** no *site* www.artmed.com.br.

A.8 CAPÍTULO 14

Veja o arquivo **Appendix Chapter 14.pdf** no *site* www.artmed.com.br.

A.9 CÁLCULOS DOS COEFICIENTES DOS ESCORES DOS FATORES

Veja o arquivo **Factor Scores.pdf** no *site* www.artmed.com.br.

REFERÊNCIAS

- AGRESTI, A., FINLAY, B. *Statistical methods for the social sciences*. San Francisco (CA): Dellen, 1985. (2ª ed.).
- APA (American Psychological Association). *Publication manual of the American Psychological Association*. Washington (DC): APA Books, 2001 (5ª ed.).
- ARRINDELL, W. A., van der ENDE, J. An empirical test of the utility of the observer-to-variables ration in factor and components analysis. *Applied Psychological Measurement*. v. 9, p. 165-78, 1985.
- BARGMAN, R. E. Interpretation and use of a generalized discriminant function. In: R. C. Bose et al (eds.). *Essays in probability and statistics*. Chapel Hill: University of North Carolina Press, 1970.
- BARNETT, V., LEWIS, T. *Outliers in statistical data*. New York: Wiley, 1978.
- BELSEY, D. A., KUH, E., WELSCH, R. *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley, 1980.
- BERRY, W. D. *Understanding regression assumptions*. Sage university paper series on quantitative applications in the social sciences. Newbury Park (CA): Sage. p. 07-92, 1993.
- BERRY, W. D., FELDMAN, S. *Multiple regression in practice*. Sage university paper series on quantitative applications in the social sciences. Beverly Hills (CA): Sage. p. 07-50, 1985.
- BOCK, R. D. *Multivariate statistical methods in behavioural research*. New York: McGraw-Hill, 1975.
- BOIK, R. J. A priori tests in repeated-measures designs: effects of nonsphericity. *Psychometrika*. v. 46, n. 3, p. 241-55, 1981.
- BOWERMAN, B. L., O'CONNELL, R. T. *Linear statistical models: an applied approach*. Belmont (CA): Duxbury, 1990 (2ª ed.).
- BRAY, J. H., MAXWELL, S. E. *Multivariate analysis of variance*. Sage university paper series on quantitative applications in the social sciences. Newbury Park (CA): Sage. p. 07-54, 1985.
- BROWN, M. B. FORSYTHE, A. B. The small sample behaviour of some statistics which test the equality of several means. *Technometrics*. v. 16, p. 129-132, 1974.
- BROWN, W. Some experimental results in the correlation of mental abilities. *British Journal of Psychology*. v. 3, p. 296-322, 1910.
- CATTEL, R. B. *The scientific analysis of personality*. Chicago: Aldine, 1966.
- CATTELL, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*. v. 1, p. 245-76, 1966.

- CLIFF, N. *Analyzing multivariate data*. New York: Harcourt, Brace Javanovich, 1987.
- COHEN, J. Multiple regression as a general data-analytic system. *Psychological Bulletin*. v. 70, n. 6, p. 426-43, 1968.
- COHEN, J. *Statistical power analysis for the behavioural sciences*. New York: Academic Press, 1988. (2nd ed.).
- COHEN, J. Things I have learned (so far). *American Psychologist*. v. 45, n. 12, 1990. p. 1304-1312.
- COHEN, J. A power primer. *Psychological Bulletin*. v. 112, n. 1, p. 155-59, 1992.
- COHEN, J. The earth is round ($p < .05$). *American Psychologist*. v. 49, 1994. p. 997-1003.
- COLE, D. A., MAXWELL, S. E., ARVEY, R., SALAS, E. How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. *Psychological Bulletin*. v. 115, n.3, 1994. p. 465-474.
- COLIER, R. O., BAKER, F. B., MANDEVILLE, G.K., HAYES, T. F. Estimates of test size for several test procedures based on conventional variance ratios in the repeated measures design. *Psychometrika*. V. 32, n. 2, 1967. p. 339-352.
- COMREY, A. L., LEE, H. B. *A first course in factor analysis*.
- COOK, R. D., WEISBERG, S. *Residuals influence in regression*. New York: Chapman & Hall. 1982.
- COOPER, C. L., SLOAN, S. J. & WILLIAMS, S. *Occupational Stress Management Guide*. Windsor, Berks: NFER-Nelson. 1988.
- CORTINA, J. M. What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104. 1993.
- COX, D. R., SNELL, D. J. *The analysis of binary data*. London: Chapman & Hall. (2nd ed.). 1989.
- CRONBACH, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*. v. 16, p. 297-334. 1951.
- CRONBACH, L. J. The two disciplines of scientific psychology. *The American Psychologist*. v. 12, p. 671-684. 1957.
- DAVIDSON, M. L. Univariate versus multivariate tests in repeated-measures experiments. *Psychological Bulletin*. v. 77, p. 446-452. 1972.
- DOMJAN, M., BLESBIS, E., WILLIAMS, J. The adaptive significance of sexual conditioning: Pavlovian conditioning of sperm release. *Psychological Science*. v. 9, p. 411-415. 1998.
- DUNTEMAN, G. E. *Principal Components analysis*. Sage university paper series on quantitative applications in the social sciences, 07-069. Newbury Park, CA: Sage. 1989
- DURBIN, J., WATSON, G. S. Testing for serial correlation in least squares regression, II. *Biometrika*. v. 30, p. 159-178. 1951.
- ERIKSSON, S. G., BECKHAM, D., VASSEL, D. Why are the English so shit at penalties? A review. *Journal of Sporting Ineptitude*. v. 31, p. 231-1072. 2004.
- ERLEBACHER, A. Design and analysis of experiments contrasting the within-and between-subjects of the independent variable. *Psychological Bulletin*. v. 84, p. 212-219. 1977.
- EYSENCK, H. J. *The structure of human personality*. New York: Wiley. 1953.
- EINSPRUCH, E. L. *An Introductory guide to SPSS for Windows*. Thousand Oaks (CA): Sage, 1998.
- FIELD, A. P. A bluffer's guide to sphericity. *Newsletter of the Mathematical, Statistical and Computing section of the British Psychological Society*. v. 6, n. 1, p. 13-22. 1988a.
- FIELD, A. P. Review of nQuery Adviser Release 2.0. *British Journal of Mathematical and Statistical Psychology*. v. 52, n. 2, p. 368-369. 1998b.
- FIELD, A. P. *Discovering statistics using SPSS for Windows: advanced techniques for the beginner*. London: Sage. 2000.
- FIELD, A. P. Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychology Methods*. v. 6, p. 161-180. 2001.
- FIELD, A. P. *Clinical psychology*. Crucial: Exeter. 2001.
- FIELD, A. P., DAVEY, G. C. L. Reevaluating evaluative conditioning: a nonassociative explanation of conditioning in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology: Animal Processes*. v. 25, n.2, p. 211-224. 1999.
- FIELD, A. P. *Designing a questionnaire*. 2000.

- FIELD, A. P., HOLE, G. *How to design and report experiments*. London: Sage, 2003.
- FISHER, R. A. *Statistical methods, experimental design, and scientific inference*. Oxford: Oxford University Press. 1925/1991.
- FLANAGAN, J. C. A proposed procedure for increasing the efficiency of objective tests. *Journal of Educational Psychology*. v. 28, p. 17-21. 1937.
- FOSTER, J. J. *Data Analysis using SPSS for Windows: a beginner's guide*. London: Sage, 2001 (2ª ed.).
- FRIDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*. v. 32, p. 675-701. 1937.
- GIRDEN, E. R. *ANOVA: repeated measures*. Sage university paper series on quantitative applications in the social sciences. p. 07-084. Newbury Park (CA): Sage. 1992.
- GLASS, G. V., PECKHAM, P. D. & SANDERS, J. R. Consequences of failure to meet assumptions underlying analysis of variance and covariance. *Educational Research*. v. 42, p. 237-288. 1972.
- GOULD, S. J. *The Mis measure of Man*. London: Penguin. 1981.
- GRAHAM, J. M., GUTHRIE, A. C., THOMPSON, B. Consequences of not interpreting structure coefficients in published CFA research: a reminder. *Structural Equation Modeling*. v. 10, n. 1, p. 142-153. 2003.
- GRAYSON, D. Some myths and legends in quantitative psychology. *Understanding Statistics*. v. 3, n. 1, p. 101-134. 2004.
- GREEN, S. B. How many subjects does it take to do a regression analysis? *Multivariate Behavioural Research*. v. 26, p. 499-510. 1991.
- GREENHOUSE, S. W., GEISSER, S. On methods in the analysis of profile data. *Psychometrika*. v. 24, p. 95-112. 1959.
- GUADAGNOLI, E., VELICER, W. Relation of sample size to the stability of component patterns. *Psychological Bulletin*. v. 103, p. 265-275. 1988.
- HAKSTIAN, A. R., ROED, J. C., LIND, J. C. Two sample T^2 procedure and assumption of homogeneous covariance matrices. *Psychological Bulletin*. v. 86, p. 1255-1263. 1979.
- HARDY, M. A. *Regression with dummy variables*. Sage university paper series on quantitative applications in the Social Sciences. 07-093. Newbury Park (CA): Sage, 1993.
- HARMAN, B. H. *Modern factor analysis*. Chicago: University of Chicago Press. 1976.
- HARRIS, R. J. *A primer of multivariate statistics*. New York: Academic Press. 1975.
- HOAGLIN, D., WELSCH, R. The hat matrix in regression and ANOVA. *American Statistician*. v. 32, p. 17-22. 1978.
- HODDLE, G., BATTY, D., INCE, P. How not to take penalties in important soccer matches. *Journal of Cretinous Behaviour*. v. 1, p. 1-2. 1998.
- HOSMER, D. W., LEMESHOW, S. *Applied logistic regression*. New York: Wiley. 1997.
- HOWELL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 1997. (4ª ed)
- HOWEL, D. C. *Statistical methods for psychology*. Belmont (CA): Duxbury, 2002. (5ª ed)
- HUBERTY, C. J. & MORRIS, J. D. Multivariate analysis versus multiple univariate analysis. *Psychological Bulletin*. v. 105, p. 302-308. 1989.
- HUTCHESON, G., SOFRONIOU, N. *The multivariate social scientist*. London: Sage, 1999.
- HUYNH, H. & FELDT, L. S. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics*. v. 1, n. 1, p. 69-82. 1976.
- IVERSEN, G. R., NORPOTH, H. *ANOVA*. Newbury Park (CA): Sage. 1987. 2nd ed.
- JACKSON, S. & BRASHERS, D. E. *Random factors in ANOVA*. Sage university paper series on quantitative applications in the social sciences. Thousands Oaks (CA): Sage.. p. 07-098. 1994.
- JOLLIFFE, I. T. Discarding variables in a principal component analysis, I: artificial data. *Applied Statistics*. v. 21, p. 160-173.
- JOLLIFFE, I. T. *Principal component analysis*. New York: Springer-Verlag. 1986.
- JONCKHEERE, A. R. A distribution-free k -sample test against ordered alternatives. *Biometrika*. v. 41, p. 133-145. 1954.
- KAISER, H. F. The application of electronic computers to factor analysis. *Educational and*

- Psychological Measurement*. v. 20, p. 141-151. 1960.
- KAISER, H. F. A second-generation little jiffy. *Psychometrika*. v. 35, p. 401-415. 1970.
- KAISER, H. F. An index of factorial simplicity. *Psychometrika*. v. 39, p. 31-36. 1974.
- KASS, R. A., & TINSLEY, H. E. A. Factor analysis. *Journal of Leisure Research*. v. 11, p. 120-138.
- KESELMAN, H. J. & KESELMAN, J. C. Repeated measures multiple comparison procedures: effects of violating multisample sphericity in unbalanced designs. *Journal of Educational Statistics*. v. 13, n.3, p. 215-226. 1988.
- KINNEAR, P. R., GRAY, C. D. *SPSS for Windows made simple (Release 10)*. Hove: Psychology Press, 2000.
- KLINE, P. *The handbook of psychological testing*. London: Routledge. 1986. (2nd ed)
- KLOCKARS, A. J., SAX, G. *Multiple comparisons*. Sage university paper series on quantitative applications in the social sciences. Beverly Hills (CA): Sage. p. 07-61, 1986.
- KRUSKAL, W. H. & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*. v. 47, p. 583-621. 1952.
- LOFTUS, G. R. & MASSON, M. E. J. Using intervals in within-subject designs. *Psychonomic Bulletin and Review*. v. 1, n.4, p. 476-490. 1994.
- LUNNEY, G. H. Using analysis of variance with a dichotomous dependent variable: an empirical study. *Journal of Educational Measurement*. v. 7, n. 4, p. 263-269. 1970.
- MacCALLUM, R. C., WIDAMAN, K. F., ZHANG, S., HONG, S. Sample size in factor analysis. *Psychological Methods*. v. 4, n. 1, p. 84-99. 1999.
- MANN, H. B., WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*. v. 18, p. 50-60. 1947.
- MAXWELL, S. E. Pairwise multiple comparisons in repeated-measures designs. *Journal of Educational Statistics*. v. 5, n. 3, p. 269-287. 1980.
- MAXWELL, S. E., DELANEY, H. D. *Designing experiments and analyzing data*. Belmont (CA): Wadsworth. 1990.
- MENARD, S. *Applied logistic regression analysis. Sage university paper series on quantitative applications in the social sciences*. Thousand Oaks (CA): Sage. p. 07-106, 1995
- MENDOZA, J. L., TOOTHAKER, L. E. & CRAIN, B. R. Necessary and sufficient conditions for *F* ratios in the $L \times J \times K$ factorial design with two repeated factors. *Journal of the American Statistical Association*. v. 71, p. 992-993. 1976.
- MENDOZA, J. L., TOOTHAKER, L. E., NICEWANDER, W. A. A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated-measures design. *Multivariate Behavioural Research*. v. 9, p. 165-177. 1974.
- MILES, J., SHEVLIN, M. *Applying regression and correlation: a guide for students and researchers*. London: Sage, 2001.
- MITZEL, H. C. & GAMES, P. A. Circularity and multiple comparisons in repeated-measures designs. *British Journal of Mathematical and Statistical Psychology*. v. 34, p. 253-259. 1981.
- MYERS, R. *Classical and modern regression with applications*. Boston (MA): Duxbury. 1990.
- NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. *Biometrika*. v. 78, p. 691-692. 1991.
- NAMBOODIRI, K. *Matrix algebra: an introduction*. Sage university paper series on quantitative applications in the social sciences. Beverly Hills (CA): Sage. p. 07-38. 1984.
- NORUSIS, M. J. *SPSS for Windows Base system user's guide, release 7.5*. SPSS Inc. 1997.
- NUNNALLY, J. C. *Psychometric theory*. New York: McGraw-Hill. 1978.
- O'BRIEN, M. G. & KAISER, M. K. MANOVA method for analyzing repeated-measures designs: an extensive primer. *Psychological Bulletin*. v. 97, n. 2, p. 316-333. 1985.
- OLSON, C. L. Comparative robustness of six tests in multivariate analysis of variance. *Journal of the Statistical Association*. v. 69, p. 894-908. 1974.
- OLSON, C. L. On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*. v. 83, p. 579-586. 1976.

- OLSON, C. L. Practical considerations in choosing a MANOVA test statistic: a rejoinder to Stevens. *Psychological Bulletin*. v. 86, p. 1350-1352. 1979.
- PEDHAZUR, E., SCHMELKIN, L. *Measurement, design and analysis*. Hillsdale (NJ):Erlbaum. 1991.
- RAMSEY, P. H. Empirical power of procedures for comparing two groups on p variables. *Journal of Educational Statistics*. v. 7, p. 139-156. 1982.
- ROSENTHAL, R. *Meta-analytic procedures for social research* (revise). Newbury Park (CA): Sage. 1991.
- ROSENTHAL, R. ROSNOW, R. L., RUBIN, D. B. *Contrasts and effect sizes in behavioral research: a correlational approach*. Cambridge: Cambridge University Press, 2000.
- ROSNOW, R. L., ROSENTHAL, R. *Beginning behavioral research: a conceptual primer*. Englewood Cliffs (NJ): Pearson/Prentice Hall, 2005. 5ª ed.
- ROSNOW, R. L., ROSENTHAL, R., RUBIN, D. B. Contrasts and correlations in effect-size estimation. *Psychological Science*. v. 11, p. 446-453. 2000.
- ROUANET, H., LÉPINE, D. Comparison between treatments in a repeated-measurement design: ANOVA and multivariate methods. *The British Journal of Mathematical and Statistical Psychology*. v. 23, p. 147-163. 1970.
- ROWNTREE, D. *Statistics without tears: a primer for non-mathematicians*. London: Penguin, 1981.
- RULON, P. J. A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*. v. 9, p. 99-103. 1939.
- RUTHERFORD, A. *Introducing ANOVA and ANCOVA: A GLM Approach*. London: Sage, 2000.
- SCARIANO, S. M., DAVEMPORT, J. M. The effects of violations of independence in the one-way ANOVA. *The American Statistician*. v. 41, n. 2, p. 123-29. 1987.
- SIEGEL, S & Castellan, N. J. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill. 1988. (2ª ed)
- SPEARMAN, C. Correlation calculated with faulty data. *British Journal of Psychology*. v. 3, p. 271-295. 1910.
- SPSS Inc. *SPSS Base 7.5 syntax reference guide*. SPSS Inc. 1997.
- STEVENS, J. P. Comment on Olson: choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*. v.86, p. 355-360. 1979.
- STEVENS, J. P. Power of the multivariate analysis of variance tests. *Psychological Bulletin*. v. 88, p. 728-737. 1980.
- STEVENS, J. P. *Applied multivariate statistics for the social sciences*. Hillsdale (NJ): Erlbaum. 1992. (2ª ed.)
- STRANG, G. *Linear algebra and its applications*. New York: Academic Press. (2ª ed.)
- STUART, E. W., SHIMP, T. A., ENGLE, R. W. Classical conditioning of consumer attitudes: four experiments in an advertising context. *Journal of Consumer Research*. v. 14, p. 334-349. 1987.
- STUDENMUND, A. H., CASSIDY, H. J. *Using econometrics: a practical guide*. Boston: Little, Brown. 1987.
- TABACHNICK, B. G., FIDELL, L. S. *Using multivariate statistics*. Boston: Allyn & Bacon. 2001. (4ª ed.).
- TERPSTRA, T. J. The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*. v. 14, p. 327-333. 1952.
- TINSLEY, H. E. A. & TINSLEY, D. J. Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*. v. 34, p. 414-424. 1987.
- TOMARKEN, A. J. & SERLIN, R. C. Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin*. v. 99, p. 90-99. 1986.
- TOOTHAKER, L. E. *Multiple comparison procedure*. Sage university paper series on quantitative applications in the social sciences. Newbury Park (CA): Sage. p. 07-89. 1993.
- WELCH, B. L. On the comparison of several mean values: an alternative approach. *Biometrika* v. 38, p. 330-336. 1951.
- WICOXON, F. Individual comparisons by ranking methods. *Biometrics*. v. 1, p. 80-83. 1945.
- WILDT, A. R., AHTOLA, O. *Analysis of covariance*. Sage university paper series on quantitative applications in the social sciences. p. 07-12. Newbury Park (CA): Sage, 1978.

- WILLIAMS, J. M. G. *Suicide and attempted suicide*. London: Penguin. 2001.
- WRIGHT, D. B. *Understanding statistics: an introduction for the social sciences*. London: Sage. 1997.
- WRIGHT, D. B. *First steps in statistics*. London: Sage, 2002.
- WRIGHT, D. B. Making friends with your data: improving how statistics are conducted and reported. *British Journal of Educational Psychology*. v. 73, p. 123-136. 2003.
- ZWICK, R. Nonparametric one-way multivariate analysis of variance: a computational approach based on the Pillai-Bartlett trace. *Psychological Bulletin*. v. 97, p. 148-152. 1985.

ÍNDICE

-2LL, veja Estatística de Verossimilhança Log

A

α de Cronbach 593-602, 641-665
Abscissa 129-130

Aderência 31-33, 36, 158-160, 246-247, 353-354, 565-567, 628-629, 641-665

Aderência de Hosmer e Lemeshow's 232-233, 251, 254
Ajuste do gl 174, 176-178, 641-665, *veja também* DF ajuste padronizado

Ajuste *veja Aderência*

Aleatorização 268-270, 641-665
Amostra 641-665

Amplitude 93-94, 96, 99

Análise da Função Discriminante 514, 531-533, 541-549, 551, 537-538, 563, 641-665

Análise da tendência 320-326, 332-333

Análise de Confiabilidade 593-603, *veja também* α de Cronbach

Análise de Covariância, *veja* ANCOVA

Análise de Fatores 258-259, 553-604, 641-665, *veja também* Extração, Carga do Fator, Escores dos Fatores

Análise de Fatores Confirmatória 561-562, 641-665

Análise de Fatores de Máxima Verossimilhança 561-562, 565-567

Análise de Variância, *veja* ANOVA

Análise de Variância Multivariada 641-665, *veja* MANOVA

Análise dos Componentes Principais 553-604, 641-665, *veja também* Análise de Fatores

Análise dos Efeitos Simples 382-383, 422-425, 641-665

Análise dos Valores Omissos 191

Análise Loglinear 617-638, 641-665

ANCOVA 153, 343-363, 531-533, 641-665

ANOVA 165-166, 195-196, 215-216, 514-515, 518-521, 526, 531-533, 537-538, 550, 597-598, 607-608, 618-619, 622-623, 641-665

ANOVA de medidas repetidas 395-440

ANOVA de medidas repetidas de dois fatores 418-437

ANOVA de medidas repetidas de um fator 398-415, 503

ANOVA de três fatores 441-472

ANOVA Fatorial 364-393, 618-621, 641-665

ANOVA independente de dois fatores 364-394

ANOVA independente de um fator 298-340, 490-491

ANOVA Mista 441-472

Hipóteses 309-311

Método da razão das variâncias 299-300

veja também Razão F

ANOVA de dois Fatores *veja também* ANOVA

ANOVA de Friedman 503-512, 597-598, 641-665

ANOVA de Medidas Repetidas de dois Fatores *veja* ANOVA

ANOVA de Medidas Repetidas *veja* ANOVA

ANOVA de Três Fatores *veja* ANOVA

ANOVA de um Fator Independente *veja* ANOVA

ANOVA Fatorial *veja* ANOVA

ANOVA Mista *veja* ANOVA

Assimetria 37, 91-94, 100-101, 107-109, 210, 641-665

Assimetria Negativa 37-38, 100-101, 641-665

Assimetria Positiva 37-38, 90-92, 100-101, 641-665

Assimetria Negativa *veja* Assimetria

Assimetria Positiva *veja* Assimetria

Autocorrelação 178-179, 641-665

Autovalores 183-184, 187-188, 202-204, 257-258, 527-528, 545-546, 564-565, 580-582

Autovetor 202-204, 526-527, 545-547, 564, 580-581

B

b, *veja* Coeficientes de Regressão

β , *veja* Coeficientes de Regressão Padronizados

Barra de Erro 87-88, 119-123, 269-279, 641-665

Bimodal 108-114, 641-665

C

C de Dunnet 323-324

Carga do Fator 546-547, 555-559, 568-570, 576-577, 582-583, 587-591, 641-665

Casos Influentes 174-177, 243-247, *veja também* Influência, ponto de influência

Causalidade 141-146, 153, 265

Centroides 547-548

Circularidade 395-396

Circularidade Local 396-397

Código Indicador 230-231

Coefficiente de Determinação 143, 641-665, *veja* R^2

Coeficientes da Regressão Padronizados 199-200, 212, *veja também* Coeficientes da Regressão

Coeficientes de Regressão *veja* Regressão

Colinarietà, *veja* Multicolinarietà, Colinarietà Perfeita

Colinearidade Perfeita 182-183, 641-665

Comando *Descriptives* (Descritivas) *veja* SPSS

Comando *Frequencies* (Frequências) *veja* SPSS

Comando: *Compute* (Calcular), *veja* SPSS

Comando: *Explore* (Explorar) *veja* SPSS

Comando: *recode* (recodificar) *veja* SPSS

Comando: *Resumo de Casos* (*Case Summaries*), *veja* SPSS

Comando: *split file* (dividir arquivo) *veja* SPSS

Comando: *Weight Cases* (Ponderar Casos) *veja* SPSS

Combinação Linear (CL), *veja* Variate

Comparações par a par 322-323, 641-665

Comparações Planejadas *veja* Contrastes Planejados

Comunalidade 562-567, 581-583, 587, 641-665

Confiabilidade Meio a Meio 593-594, 641-665

Confiabilidade Teste-Retest 593, 641-665

Contrabalanceamento 268-269, 641-665

Contraste 230-231, 302-303, 345-346, 351-352, 372-373, 404-406, 412-415, 420-423, 432-437, 444-446, 452, 454-468, 514, 532-533, *veja também* Contraste Planejado

Contraste de Helmert Reverso 641-665

Contraste Desvio 230-231, 233-235, 320-321, 641-665

Contraste Diferença 230-231, 320-321, 420-421, 641-665

Contraste Ortogonal 316-318

Contraste Polinomial 224-225, 230-231, 320-325, 405, 641-665

Contraste Repetido 230-231, 320-321, 372-373, 405-406, 641-665

Contraste Simples 230-231, 233-235, 320-321, 345-346, 351-352, 372-373, 405, 420-423, 444-445, 532-533, 539-542, 641-665

Contrastes de Helmert 230-231, 320-321, 372-373, 378-380, 641-665

Contrastes Não-ortogonais 319-320

Contrastes Planejados 310-321, 324-328, 333-337, 356-358, 641-665

Correção de Bonferroni 322-324, 347-348, 380-381, 383, 406-408, 427-431, 496-498, 502-503, 508, 531, 550, 641-665

Correção de Continuidade de Yates 609-610, 614-616, 641-665

Correção de Greenhouse e Geisser 397-399, 409-412, 416-417, 427-429, 641-665

Correção de Huynh-Feldt 397-399, 409-410, 641-665, *veja também* Esfericidade

Correção de Sidak 347-348, 407-408, 641-665

Correlação 125, 129, 514-516, 615-616

Correlação Parcial 138-139, 148-153, 171-172, 182-183, 200-202, 224-225, 641-665

Correlação Ponto Bisserial 145-149, 153-154, 212-213, 258-259, 285-286, 641-665

Correlação Semiparcial 153, 170-172, 187, 641-665

ρ de Spearman 138-140, 144-146, 153-154, 641-665

r de Pearson 56-58, 127-129, 138-148, 151-154, 161-162, 165-166, 175-176, 182-183, 187, 191-194, 200-201, 212-213, 223-224, 238, 258, 284-286, 292-294, 309-310, 337-339, 360-362, 388, 415-416, 437-438, 467-469, 482-484, 490, 501-503, 509-511, 553-559, 569-570, 577-579, 641-665

Tau de Kendall 138-140, 145-146, 153-154, 612-613, 641-665

Correlação Bisserial 145-149, 641-665

Correlação Bivariada 641-665, *veja* Correlação

Correlação Canônica 546-547

Correlação de Ordem Zero 151-153, 187, 200-201

Correlação Parcial de Segunda Ordem 151-153

Correlação Parcial de Terceira Ordem 151-153

Correlação Parcial *veja* Correlação

Correlação Ponto Bisserial 641-665, *veja* Correlação

Correlação Semiparcial *veja* Correlação

Corridas de Wald-Wolfowitz 479-481, 641-665

Covariância 125-128, 538-540, 641-665
 Covariável 343-344, 352-353, 531-533, 641-665
 Critério de Kaiser 564-565, 574-575, 580-581, 583, 587, 641-665
 Curtose 36-39, 92-94, 107-109, 111-113, 115-116, 120, 641-665
 Leptocúrtica 37-39, 88-89, 641-665
 Platicúrtica 37-39, 531, 641-665

D

d de Cohen 56-57
 Dados de Razão 71-72, 641-665
 Dados Intervalares 71-72, 85-86, 140-141, 144-145, 178, 280, 309-310, 641-665
 Dados Nominais 71-72, 641-665
 Dados omissos (*missing data*) ou dados faltantes 190-191
 Dados Ordinais 71-72, 144-145, 641-665
 Dados Paramétricos 139-140
 Delineamento de Medidas Repetidas 266-268, 364-365, 395-440, 641-665
 Delineamento entre grupos 266-268, 641-665
 Delineamento entre Participantes, *veja* Delineamento entre Grupos
 Delineamento Independente *veja* Delineamento entre Grupos 641-665
 Delineamento Misto 364-365, 441-472, 641-665
 Delineamentos desbalanceados 322-323, 370-372, 406-407
 Delineamentos Relacionados 641-665, *veja* Delineamentos de Medidas Repetidas
 Desvio 33-34, 36-37, 159-160, 304-305, 606-607, 641-665
 Desvio Padrão 35-42, 44, 92-96, 99, 107-109, 111-113, 115-116, 120, 127-129, 187, 191, 199-200, 328, 328-330, 641-665
 Desvios dos Produtos Cruzados 126-127, 139-140, 518-519, 521-526, 538-540, 641-665
 Modelo (CP_M) 521-525
 Residual (CP_R) 521-525
 Total (CP_T) 521-523

Determinante 571-572, 577-579
 DFBeta 175-176, 206-208, 244, 246-247, 641-665, *veja também* DFBeta Padronizado
 DFBeta Padronizado 175-176, 189-191, 641-665, *veja também* DFBeta
 DFFit Padronizado 176-178, 189-191, 641-665, *veja também* DFFit
 Diagnósticos por casos 187-188, 204-208
 Diagrama de barras 87-88, 92-96, 99, 119-123, 269-270, 328, 376, 384-386
 Diagrama de Caixa e Bigodes (*Boxplot*) 87-92, 96, 99-97, 113-114, 496-497, 641-665
 Diagrama de Classificação 232-233, 240-242, 251, 253-254
 Diagrama de Declividade 564-565, 573-575, 583-584, 587, 641-665
 Diagrama de Dispersão 47-48, 89-91, 129-138, 163-165, 159-160, 168-169, 177, 188-189, 202-203, 210-211, 352-353, 357-358, 641-665
 Diagrama de Dispersão em 3D
 Diagrama de Dispersão Sobreposto 134-136
 Diagrama de *Pizza* 87-88, 92-93
 Diferença Honestamente Significativa (DHS) de Tukey 322-324, 327-328, 334-338, 347-348
 Diferença Minimamente Significativa (DHM) de Tukey 322-323, 407-408, 427-429
 Distância de Cook 175-176, 189-191, 205-208, 244, 246-247, 641-665
 Distância de Mahalanobis 174-175, 189-191, 205-208, 641-665
 Distribuição Amostral 42, 44, 280-281, 641-665
 Distribuição de Frequência/ Histograma 36-37, 86-93, 107-114, 188-189, 207-212, 232-233, 240-241, 281-282, 641-665
 Distribuição de Probabilidade 41-42, 641-665
 Distribuição Normal 36-42, 85-86, 140-141, 202-203, 641-665
 Verificação da Normalidade 93-95, 112-116, *veja também*

 Teste de Kolmogorov-Smirnov, teste de Shapiro-Wilk
 Distribuição Qui-Quadrado 224-226, 235-236, 238, 493, 506-508, 608-609, 641-665
 Divisão pela Mediana 290-291
 DMS (Diferença Menos Significativa) ou LSD *veja* Least Significant Difference

E

E *veja* SSCP
 E^{-1} *veja* SSCP
 Efeito álcool (*beer goggles*) 365-389, 641-665
 Efeito Prático 268-269, 641-665
 Efeito Principal 359-360, 367-369, 376-380, 383-384, 427-431, 447-448, 450-456, 469-470, 641-665
 Efeito Tédio 268-269, 641-665
 Efeitos Supressores 171-172, 227-228, 641-665
 Eixo x 129-130, 575-576
 Eixo y 129-130, 575-576
 Encolhimento 180-181, 194-195, 641-665
 Entrada Forçada 170-171, 227
 Entre participantes 641-665, *veja* Medidas Repetidas
 Equação de Wherry 180-181
 Erro do SSCP *veja* SSCP
 Erro do Tipo I 55-56, 298-299, 309-310, 322-324, 328-330, 405-407, 412, 483-484, 496-498, 514-515, 531, 537-538, 609, 641-665
 Erro Padrão da Média 42-45, 191, 197, 199-200, 212, 225-226, 246-247, 270-271, 278-282, 288, 328-330, 333-336, 641-665
 Erro Padrão das Diferenças 281-283, 641-665
 Erros Independentes (Suposição de) 178-179, 186-188, 195-197, 641-665
 Escores Discriminantes 544-548, 641-665
 Escores dos Fatores 558-561, 575-579, 590-593, 641-665
 Escores-z 41-42, 45-47, 93-94, 97, 99-100, 107-109, 174, 476-477, 481-484, 488-491, 498-503, 508-511, 628-629, 632-633, 641-665

- Esfericidade 395-399, 405-406, 408-411, 416-417, 426-429, 437, 641-665, *veja também* Teste de Bartlett, Correção de Greenhouse e Geisser, Correção de Huynh-Feldt, Estimativa Limite Inferior e Teste de Mauchly
- Espermatozoide
Codorna Japonesa 44-45, 269-271
Humano 340-341, 490-503
- Estatística de Verossimilhança Log 223-225, 235-238, 242-243, 250-251, 254-255, 641-665
- Estatística de Wald 225-227, 234-235, 242-243, 246-247, 251, 254, 641-665
- Estatística do Escore Eficiente de Rao 227, 234-236, 641-665
- Estatística do Escore *veja* Estatística do Escore Eficiente de Rao
- Estatística Passo (*Step*) 237-238
- Estatística Qui-Quadrado 234-237, 240-243, 509-510, 606-623, 628-629, 631-632, 635-637, 641-665
- Estatística Qui-Quadrado do Modelo 236-238, 247-248, 250-251, 254-255
- Estatística *Studentizados* 199-200
- Estatísticas Condicionais 227
- Estimador não viciado 179-180, *veja também* Estimador não tendencioso
- Estimativa da Variância Agrupada 288-289
- Estimativa de Máxima Verossimilhança 223-224, 608-609, 641-665
- Eta ao quadrado 337-338, 347-348, 360-361, 386-387, 641-665
- Exp *b* 226-227, 232-233, 239-256, 641-665
- Extração 564-567, 573-575, 580-583, 587, 641-665, *veja também* Critério de Kaiser
- F**
- F de Welch 328-333, 641-665
- Fator 364, 526, 553-559, 641-665, *veja também* Análise de Fatores, Variável Independente, Variável latente
- Fator de Inflação da Variância 183-184, 187-188, 201-203, 256-258, 641-665
- Fatoração Alfa de Kaiser 561-563, 641-665
- Fatoração da imagem 561-562
- Fatoração pelo Eixo Principal 561-562
- Fi ou *Phi* 612-613, 616, 641-665
- FIV (Fator de Inflação da Variância) 641-665
- Fórmula de Stein 180-181, 194-195
- G**
- λ de Goodman e Kruskal 612-613, 641-665
- Generalização 178, 641-665
- Gradiente 156-159, 172-173, 221-222, 293-294
- Gráfico de Barras de Erro 119-123
- Gráfico de Dispersão *versus* de Nível 347-349
- Gráfico de Linhas 87-88, 378-380, 382-385
- Gráficos de Interação 372-373, 382-387, 422-426, 641-665
- Gráficos de Probabilidade Normal 189-190, 207-212
- Gráficos no SPSS 86-92
- Gráficos Interativos 119-123
- Plotar a mediana 120
- Plotar a moda 120
- Plotar a Variância 120
- Plotar o Desvio Padrão 120
- Gráficos P-P *veja* Gráficos de Probabilidade Normal
- Gráficos Q-Q ou Diagrama Quantil-Quantil 114-116
- Graus de Liberdade 163-166, 196-197, 199-200, 236-237, 284, 286-289, 292, 305-310, 339, 350-351, 360-362, 366-370, 388-390, 397-398, 401-404, 409-411, 416-417, 437-438, 493, 607-609, 641-665
- GT2 de Hochberg 323-324, 337-338
- H**
- H* *veja* SSCP
- HE^{-1} *veja* SSCP
- Heterocedasticidade 178-179, 188-189, 207-212, 641-665
- Heterogeneidade da Variância *veja* Homogeneidade da Variância
- Hipótese de SSCP (*H*) *veja* SSCP
- Hipótese Experimental 49-50, 280, 641-665
- Hipótese Nula 49-50, 278-280, 287-288, 641-665
- Hipóteses dos testes paramétricos 85-87, 309-310, 474
- Histograma *veja* Distribuição das Frequências
- Homocedasticidade 641-665
- Homogeneidade da Inclinação da Regressão 357-361, 641-665
- Homogeneidade da Matriz de Covariância 529-531, 534-536, 641-665
- Homogeneidade da Variância 85-86, 116-120, 280, 290-292, 295, 309-310, 322-324, 328-333, 337-338, 347-350, 374-375, 395-396, 407-408, 447-448, 450, 478-479, 490-491, 493-495, 529-530, 641-665
- Homogeneidade Secundária 486-488
- I**
- Independência 85-87, 178-179, 280, 395-396, 529-530, 641-665
- Índice de Condição 183-184, 187-188, 202-204, 257-258
- Influência ou ponto de influência 174, 189-191, 205-208, 244, 246-247, 641-665
- Interação 228-230, 359-360, 364-365, 368-369, 376-380, 382-387, 430-437, 490, 454-471, 514, 641-665
- Interceptar 156-158, 172-173, 221-222, 232-233, 293-295
- Intervalo Interquartilico 96, 99, 641-665
- Intervalos de Confiança 44-47, 187, 197-198, 200-201, 232-233, 239-241, 251, 254-256, 269-271, 285-286, 327-330, 334-336, 352, 378-380, 540-542, 641-665
- Intervalos de Confiança Simultâneos de Roy-Bose 405-406
- Inversa (de uma matriz) 525-526

K

τ de Kendall *veja* Correlação
KMO *veja* Medida de adequação da Amostra de Kaiser-Meyer-Olkin

L

Lambda de Wilk 529-531, 536, 545-546, 641-665
Lei da Soma das Variâncias 288, 641-665
Leptocúrtica 641-665, *veja* Curtose
Lésbicas budistas 261
Limite Inferior da Estimativa 397-398, 409-410, 641-665
Linearidade (Suposição de) 178-179, 207-212, 222-223
Linha de Regressão *veja* Regressão
Linha do Melhor Ajuste 158-160, 163-165

M

Maior Raiz de Roy 529-531, 536, 641-665
MANOVA 324-325, 328, 398-399, 412, 514-552, 563-564, 641-665
Matriz 517-518, 641-665
Matriz anti-imagem 572-573, 579-581, 577-579
Matriz da Soma dos Quadrados e do Produto Cruzado *veja* SSCP
Matriz das Variâncias-Covariâncias 183-184, 187, 530-531, 538-540, 563-564, 593-595, 641-665
Matriz de Componentes 557-558, 641-665, *veja também* Matriz Padrão, Matriz Estrutural
Matriz de Correlação 141-142, 187, 191-193, 202-204, 232-233, 538-540, 553-554, 564, 571-574, 577-580, 583, 587
Matriz de identidade 518, 533-534, 539-540, 571-573, 580-581, 590-592, 641-665
Matriz de Transformação 412-414, 422-423, 426
Matriz de Transformação dos Fatores 567-568, 589, 641-665
Matriz do Diagrama de Dispersão 135-138

Matriz dos Fatores 557-558, 641-665, *veja também* Matriz Estrutural
Matriz Estrutural 557-559, 589, 591-592, 641-665, *veja* Matriz Padrão
Matriz Padrão 557-559, 589-591, 641-665, *veja também* Matriz Estrutural
Matriz R *veja* Matriz de Correlação
Matriz Reproduzida 572-573, 583, 587-586
Média 33-34, 92-93, 119-123, 160-161, 187, 223-225, 274-275, 328-330, 641-665
Média do Grupo Ajustada 347-348, 352, 355-356, 641-665
Média dos Quadrados 162, 165-166, 308-309, 330-333, 361-362, 369-370, 386-388, 403, 409-410, 641-665
Entre Grupos (MS_{BG}) 416
Modelo (MS_M) 162, 166-167, 196-197, 308-309, 403, 415-416, 520, 521
Residual (MS_R) 162, 166-167, 196-197, 308-309, 369, 403, 415-416, 520-521
Média dos Quadrados do Modelo (MS_M) *veja* Média dos Quadrados
Média Geral 274-276, 305-306, 367-368, 641-665
Média Harmônica 336-337, 641-665
Mediana 92-93, 96, 99, 482-484, 490-491, 641-665
Médias ao quadrado entre grupos, *veja* Médias ao quadrado
Médias dos Quadrados dos Resíduos (MS_R) *veja* Médias dos Quadrados
Medida de Adequação da Amostra de Kaiser-Meyer-Olkin 570-573, 577-581, 641-665
Método da Razão das Variâncias *veja* ANOVA
Método de Anderson-Rubin 560-561, 576-577, 590-592, 641-665
Método de Bartlett (dos escores dos fatores) 560-561
Método dos Mínimos Quadrados 156-160, 221-222

Método Monte Carlo 479-481, 486-488, 494-496, 505-506, 570-571, 641-665
Métodos de Regressão 169-173
MLG *veja* Modelo linear Geral
Moda 36-37, 92-93, 120, 184-185, 641-665
Modelo da Soma dos Quadrados *veja também* Soma dos Erros Quadrados
Modelo do Produto Cruzado (CP_M) *veja* Desvios do Produto Cruzado
Modelo Linear 47-49, 156-159, 526, 556-557, 641-665, *veja também* Modelo Linear Geral
Modelo Linear Geral (MLG) 293-295, 299-304, 343-344, 352-358, 388-393, 617-624
Modelo Saturado 621, 623-624, 628-629, 641-665
Modelos Customizados 370-372
Motörhead 116-117
Mudança no R^2 187, 193-197, 215-216
Multicolinearidade 178-179, 182-184, 186-188, 191-192, 197-198, 200-204, 254-261, 553-554, 560-561, 571-573, 577-579, 593, 641-665
Multimodal 107-109, 641-665

N

Níveis de Mensuração 71-72
Normalidade dos Erros 178-179, 210-211
Normalidade Multivariada 529-531, 641-665

O

O F de Brown-Forsythe 328-333, 641-665
Ômega ao Quadrado 338, 347-348, 360-361, 386-388, 415-416, 437, 641-665
Ordenada 129-130
Ortogonal 641-665

P

Padronização 135-136, 174, 281-283, 641-665
Parcializar 343-344

- Passo a Passo para Frente 170-171, 227-228, 233-235
- Platicúrtica 641-665, *veja* Curtose
- Poder 57-59, 180-181, 293-294, 322-323, 347-348, 395-399, 405-407, 412, 434-435, 483-501, 514-517, 530-531, 534-538, 609-610, 625, 641-665
- Ponderações *veja* Pesos 314-317, 641-665
- População 32-33, 641-665
- Postos 144-145, 475-477, 484-485, 491-493, 504-505
- Postos empatados 476-477
- Previsão de membro de grupo 231, 242-243
- Probabilidade 226-227, 239-240, 242-243, 641-665
- Probabilidades Previstas 231, 242-243
- Problema da Terceira Variável 143, 178-179
- Produto Cruzado Total (CP_T) *veja* Desvios do Produto Cruzado
- Proporção da Variância 117-118, 348-350, 641-665
- Q**
- Q* de Cochran 506-507, 597-598, 641-665
- Q* de Ryan, Einot, Gabriel e Welsh *veja* REGWQ
- Quartil 96, 99
- R**
- r* de Pearson 641-665, *veja* Correlação
- R* Múltiplo 169-170, 182-183, 187, 193-194, 224-225, 562-563, 641-665
- R^2 (na regressão logística) 223-225, 238-239, 246-247, 251, 254
- R^2 143-145, 147-148, 160-162, 165-166, 169-170, 187, 190-191, 193-197, 212, 215-216, 223-225, 238-239, 246-247, 251, 254, 337-338, 347-348, 353-354, 360-361, 386-387
- R^2 Ajustado 179-181, 187, 193-195, 641-665
- R^2 de Cox e Snell 225, 238-239, 247-248, 641-665
- R^2 de Hosmer e Lemeshow 225, 238, 247-248, 641-665
- R^2 de Nagelkerke 225, 239, 247-248, 641-665
- Razão da Verossimilhança 227, 234-235, 240-241, 608-609, 615-616, 624-625, 628-629, 631-633, 637-638, 641-665
- Razão de chances 616-618, 636-638, 641-665
- Razão de Covariâncias 176-178, 189-191, 206-208, 641-665
- Razão *F* 162, 165-166, 187, 194-196, 237-238, 299-301, 303-306, 308-311, 329, 332-333, 339, 361-362, 369-370, 388-390, 397-400, 403-406, 409-414, 437-438, 448, 450-452, 454-456, 455-462, 464-467, 469-471, 520-521, 526, 528-530, 536-538, 641-665
- Razões de Variâncias 187-188, 202, 258
- Reações Extremas de Moses 479-481, 641-665
- Regressão 131-132, 156-220, 607-608, 620-621, 641-665
- Coeficientes da Regressão 157-158, 166-168, 197-202, 212, 216-218, 225-226, 237-238, 242-243, 246-247, 258, 293-295, 347-358, 389-393, 526-527, 543-547, 557-559, 641-665, *veja também* Coeficientes da Regressão Padronizados
- Hipóteses 178-180
- Interações 389-393
- Linha de Regressão 158-159, 357-359
- Regressão Hierárquica 153, 170-171, 183-184, 193-194, 196-198, 201-202, 212, 237-238, 248-256, 343-344, 641-665
- Regressão Múltipla 168-219, 221-223, 246-247, 343-344, 388-389, 526, 617-618, 623-624, 641-665
- Regressão passo a passo 170-172, 201-202, 227-230, 641-665
- Regressão passo a passo para trás 171-172, 227-228
- Tamanho da Amostra 180-183
- Usando o Modelo 167-168
- Regressão Hierárquica *veja* Regressão
- Regressão Logística 221-264, 542-543, 606, 641-665, *veja também* Regressão
- Regressão Múltipla *veja* Regressão
- Regressão passo a passo (*Stepwise Backwards*), *veja* Regressão
- REGWQ 322-324, 327-328, 336-338, 380-381, 383
- Resíduos 158-159, 172-174, 188-189, 204-209, 231-233, 243-247, 532-533, 572-573, 583, 587-586, 641-665, *veja também* Resíduos Padronizados
- Resíduos eliminados 174-175, 188-191, 641-665
- Resíduos Eliminados *Studentizados* 174, 188-191, 641-665
- Resíduos não-padronizados 641-665, *veja* Resíduos
- Resíduos Padronizados 173-174, 187-191, 207-209, 232-233, 243-244, 246-247, 641-665, *veja também* Resíduos
- Resíduos Produto-Cruzado (CP_R) *veja* Desvios do produto-cruzado
- Resíduos *Studentizados* 174, 188-191, 243, 641-665
- Rotação 558-559, 565-570, 574-576, 587-592, 641-665
- Rotação Oblíqua 558-559, 566-568, 575-576, 589-592, 641-665, *veja também* Rotação Oblimin Direta, Rotação Pro-max
- Rotação Ortogonal 558-559, 566-568, 574-575, 587-589, 591-592, 641-665, *veja também* Rotação Equamax, Rotação Quartimax, Rotação Varimax
- Rotação do Fator *veja* Rotação
- Rotação Equamax 567-568, 641-665
- Rotação Oblimin Direta 567-569, 575-577, 589-592, 641-665
- Rotação Oblíqua *veja* Rotação
- Rotação Ortogonal *veja* Rotação
- Rotação Promax 567-568, 575-576, 641-665

Rotação *Quartimax* 567-569, 641-665

Rotação *Quartimax* Direta 568-569

Rotação *Varimax* 567-569, 575-576, 587-589, 641-665

S

ρ de Spearman *veja* Correlação
Separação Completa 261-262, 641-665

Significância 51-53

Simetria Composta 395-396, 641-665

Singularidade 571-572, 577-579

Soma dos Erros ao Quadrado (SS) 34-35, 158-160, 165-166, 169-170, 607-608, 641-665

Modelo da Soma dos Quadrados (SS_M) 160-161, 169-170, 196-197, 306-308, 310-312, 329-332, 343-344, 347-350, 353-354, 360-361, 365-367, 399-403, 409-410, 518-520, 523, 525-530, 537-538, 641-665

Soma dos Quadrados Residual (SS_R) 160-161, 169-170, 196-197, 223-224, 307-308, 310-312, 329, 332-333, 343-344, 347-350, 353-354, 360-361, 365-366, 368-369, 399-403, 409-410, 518-521, 523, 525-527, 529-530, 537-538, 641-665

Soma dos Quadrados Total (SS_T) 160-161, 169-170, 305-306, 311-312, 347-350, 353-354, 360-361, 365-367, 399-401, 518-520, 523, 528-529, 537-538, 641-665

Somas dos Quadrados Tipo II 370-372

Somas dos Quadrados Tipo III 370-372

Somas dos Quadrados Tipo IV 370-372

SS: Soma dos Quadrados entre Participantes (SS_w) 399-403

Soma dos Quadrados do Tipo II *veja* Soma dos Erros ao Quadrado

Soma dos Quadrados do Tipo III *veja* Soma dos Erros ao Quadrado

Soma dos Quadrados do Tipo IV *veja* Soma dos Erros ao Quadrado

Soma dos Quadrados Residual (SS_R) *veja* Soma dos Erros ao Quadrado

Soma Total dos Quadrados (CP_R) *veja* Soma dos Erros ao Quadrado
SPSS *veja também* Gráficos no SPSS

Abrir arquivos 81-83

Comando: *compute* (calcular) 97-98, 100-105, 274-278, 595-597

Comando: *descriptives* (descriptivas) 97, 99, 275-276

Comando: *explore* (explorar) 113-119

Comando: *frequencies* (frequências) 91-93, 107-110

Comando: *recode* (recodificar) 97-98, 214-216

Comando: *split file* (dividir arquivos) 106-114, 486-487

Comando: *weight cases* (ponderar casos) 611-612

Comando: *case summaries* (resumos de casos) 204-207, 242-243, 592

Editor de Dados 61-77

Entrar Dados 68-69

Imprimir 78-80

Iniciar 59-63

Janela de Sintaxe 80-81, 641-665

Menus 101-104

Salvar Arquivos 80-82

Visualizador 61-62, 76-80, 641-665

Visualizador de Dados 63-64

Visualizador de Variáveis 63-64, 69-72, 641-665

SS *veja* Soma dos Erros ao Quadrado

SSCP 407-408, 523, 533-534, 537-540, 563-564, 641-665

E^1 526

Erro do SSCP 518-519, 524-528, 533-534, 538-540, 641-665

HE^1 526-528, 545-546, 641-665

Hipóteses SSCP 518-519, 524-528, 538-540, 641-665

SSCP Total 518-519, 523, 525-528, 641-665

SS_M *veja* Soma dos Erros ao Quadrado

SS_R *veja* Soma dos Erros ao Quadrado

SS_T *veja* Soma dos Erros ao Quadrado

SS_w *veja* Soma dos Erros ao Quadrado

T

T *veja* SSCP

T^2 de Hotelling 398-399, 529, 531, 534-536, 598, 641-665

T^2 de Tamhane 323-324

T^3 de Dunnett 323-324

Tabela de Contingência 233-235, 606-607, 609-610, 612-613, 625-626, 641-665

Tabulação Cruzada 235-236, 260-261, 611-613, 625-626, *veja também* Tabela de Contingência

Tamanho de Efeito 56-58, 153-154, 180-181, 285-286, 292-293, 337-338, 347-348, 360-361, 386-388, 412, 415-416, 467-468, 482-484, 490, 501-503, 509-512, 515-516, 616-618, 636-638, 641-665

Taxa de Erro de Conjunto 298-299, 310-311, 322-323, 514-515, 641-665, *veja também* Taxa do Erro Experimental

Taxa de Erro Experimental 298-299, 641-665

Tendência Cúbica 320-322, 641-665

Tendência Linear 320-322, 330-333, 339

Tendência Quadrática 320-322, 330-333, 641-665

Tendência Quártica 320-322, 641-665

Tertium Quid (terceira coisa) *veja* Problema da Terceira Variável

Testando Hipóteses 47-51, 53-54, 141-142, 271

Teste bilateral 53-56, 138-141, 145, 285, 292, 310-311, 327-328, 333-336, 477-478, 641-665, *veja também* Teste unilateral

Teste da mediana 495-496, 506-507, 641-665

- Teste da Soma dos postos de Wilcoxon 295, 475-485, 491-492, 641-665
- Teste de Box 530-531, 534-536, 543-544, 641-665
- Teste de Dunn 322-323
- Teste de Dunnett 327-328, 334-338
- Teste de Durbin-Watson 178-179, 186-188, 193-197, 207-208, 641-665
- Teste de Esfericidade de Bartlett 533-536, 539-540, 571-573, 577-581, 641-665
- Teste de Jonckheere-Terpstra 494-496, 499-502, 641-665
- Teste de Kolmogorov-Smirnov 112-117, 210-211, 477-479, 493-495, 505-506, 641-665
- Teste de Kruskal-Wallis 339-340, 490-505, 641-665
- Teste de Levene 117-120, 280-281, 291-293, 328, 330-334, 337-338, 347-350, 374-376, 447-448, 450-451, 478-481, 493-495, 530-531, 536-538, 641-665, *veja também* Homogeneidade da Variância
- Teste de Mann-Whitney 295, 475-485, 491-492, 498-500, 641-665
- Teste de Mauchly 396-398, 408-410, 426-427, 438-439, 447-448, 450, 641-665
- Teste de McNemar 486-488, 641-665
- Teste de Newman-Keuls 322-323
- Teste de Scheffé 322-323
- Teste de Shapiro-Wilk 112-113, 477-479, 493-495, 505-506, 641-665
- Teste dos postos com sinais de Wilcoxon 295, 484-491, 641-665
- Teste Estatístico 51-54, 641-665
- Teste Multivariado 514, 641-665
- Teste *post hoc* de Gabriel 323-324, 327-328, 337-338
- Teste *post hoc* de Games-Howell 323-324, 327-328, 334-338, 380-381, 383
- Teste *t* 140-141, 146-148, 162-165, 167-168, 171-172, 197, 199-202, 217-218, 225-226, 268-271, 278-284, 298-300, 308-311, 316-317, 323-324, 328, 333-334, 338, 343-344, 350-351, 355-358, 360-361, 618-619, 641-665
- Teste *t* Dependente 280-287, 406-407, 484-485, 641-665
- Teste *t* Independente 287-293, 475, 641-665
- Teste *t* de amostras pareadas *veja teste t* Dependente
- Teste *t* dependente 641-665, *veja teste t*
- Teste *t* emparelhado *veja teste t* dependente
- Teste *t* independente 641-665, *veja teste t*
- Teste *t* Relacionado *veja Teste t* Dependente
- Teste Unilateral 53-56, 138-141, 145, 153-154, 284, 286-287, 292-293, 310-311, 327-328, 333-336, 641-665, *veja também* Teste Bilateral
- Teste univariado 514, 641-665
- Testes Não-paramétricos 106-107, 144-146, 295, 474-513, 641-665
- Testes Paramétricos 85, 641-665
- Testes *post hoc* 310-311, 313-314, 322-324, 327-328, 334-338, 345-348, 405-408, 352, 373-374, 379-381, 383, 413-415, 422-423, 426-434, 446-447, 496-500, 508-510, 533-534, 641-665
- Tolerância 183-184, 187-188, 201-203, 256-258, 641-665
- Traço de Pillai-Bartlett 528-531, 536, 641-665
- Transformação de Dados 98-107, 118-119, 330-332, 448, 450-451, 95-96, *veja também* Transformação Log, Transformação Recíproca, Transformação da Raiz Quadrada
- Transformação Inversa 100-101, 103-104
- Transformação Log 100-101, 103-104, 118-120, 222-223, 347-349, 619
- Transformação pela Raiz Quadrada 100-101, 103-104, 330-332, 376
- V**
- V de Cramer 612-613, 616, 641-665
- Validação Cruzada 179-181, 641-665
- Valor Atípico ou valor discrepante 88-89, 95-99, 129-134, 172-174, 176-178, 188-189, 207-208, 210-212, 641-665
- Lidando com valores atípicos 98-107
- Valor Chapéu 641-665 *veja* Influência ou ponto de influência
- Valores Esperados 607-610, 612-615, 621-625
- Valores omissos ou faltantes 75-76
- Valores previstos ajustados 174, 188-191, 641-665
- Valores Previstos Padronizados 188-191, 208-209
- Varição Amostral 42
- Varição não-sistemática 51, 53-54, 266-269, 281-283, 293-294, 308-309, 330-333, 395-396, 403, 525-526, 529, 641-665, *veja também* Varição Sistemática
- Varição Sistemática 51, 53-54, 266-269, 281-282, 308-309, 403, 525-526, 641-665, *veja também* Varição não-sistemática
- Variância 35-40, 92-93, 120, 125-126, 305-306, 308, 366-367, 538-540, 641-665
- Variância Aleatória 562-563, 641-665
- Variância Comum 562-563, 582-583, 641-665
- Variância Única 148-149, 151, 182-183, 562-563, 641-665
- Variates Função Discriminante 526-530, 543-548, 564, 641-665, *veja CL (Combinação Linear)*
- Variates *veja* Variates da função discriminante e Combinação Linear
- Variáveis Auxiliares (*Dummy*) 212-215, 230-231, 294-295, 300-304, 314-317, 343-344, 348-350, 356-358, 389-390, 618-623, 641-665
- Variável Categórica 606, 641-665
- Variável Código 71-74, 271
- Variável Contínua 145-148
- Variável de Confusão ou Interferente 269-270, 343-344, 348-350, 641-665

Variável Dependente 156-157,
184-185, 266, 641-665
Variável Dicotômica 145-146,
641-665
Variável Discreta 145-148
Variável Independente 156-157,
184-185, 266, 641-665, *veja tam-
bém* Variável Previsora
Variável Latente 526, 553-554,
561-562, 569-570, 641-665, *veja
também* Fator

Variável Numérica 74-75
Variável Previsora 156-157, 641-
665
Variável texto 74-75, 641-665
Vetor 202-203, *veja também* Vetor
Coluna, Vetor Linha
Vetor Coluna 518
Vetor Linha 518
Visualizador 641-665, *veja SPSS*
Visualizador de Variáveis 641-665,
veja SPSS

W

W de Kendall 506-507, 641-665
WSD de Tukey 405-407

Z

Z de Kolmogorov-Smirnov 479-
481, 641-665

