

ESTATÍSTICA

BÁSICA E APLICADA



ÍNDICE

1. APRESENTAÇÃO DO CURSO E INTRODUÇÃO À ESTATÍSTICA	1
2. O MÉTODO ESTATÍSTICO, SUAS CARACTERÍSTICAS E LIMITAÇÕES	3
3. TIPOS DE DADOS (QUALITATIVOS, QUANTITATIVOS, DISCRETOS, CONTÍNUOS, ETC.)	7
4. AMOSTRAGEM, COLETA E ORGANIZAÇÃO DOS DADOS	9
5. CONSTRUÇÃO E INTERPRETAÇÃO DE TABELAS E GRÁFICOS ESTATÍSTICOS	11
6. MEDIDAS DESCRITIVAS DE DADOS OBSERVADOS: TENDÊNCIA CENTRAL E DE POSIÇÃO	16
7. MEDIDAS DE POSIÇÃO, DISPERSÃO OU VARIABILIDADE E ASSIMETRIA	19
8. NOÇÕES DE PROBABILIDADE	21
9. DISTRIBUIÇÕES DE PROBABILIDADE	24
10. INTERVALOS DE CONFIANÇA	29
11. TESTE DE HIPÓTESES PARA UMA POPULAÇÃO	34
12. INFERÊNCIA PARA DUAS POPULAÇÕES	39
13. ANÁLISE DE VARIÂNCIA (ANOVA)	45
14. CORRELAÇÃO E REGRESSÃO	50
15. O TESTES QUI QUADRADO	56
16. NOÇÕES DE ESTATÍSTICA NÃO PARAMÉTRICA	61

ANEXOS

FÓRMULAS	1
DISTRIBUIÇÃO Z (-)	5
DISTRIBUIÇÃO Z (+)	6
DISTRIBUIÇÃO T	7
DISTRIBUIÇÃO F	9
DISTRIBUIÇÃO QUI-QUADRADO	10

1. APRESENTAÇÃO DO CURSO E INTRODUÇÃO À ESTATÍSTICA

Estatística Básica – algumas considerações:

- A estatística está associada na elaboração, análise e comunicação de todo tipo de pesquisa científica de qualquer área.
- Também é muito utilizada na divulgação destas políticas junto à sociedade.
- É utilizada de forma intencional ou não para justificar ou induzir a conclusões muitas vezes equivocadas.
- Uma sólida compreensão do método estatístico, suas características e principalmente suas limitações, é fundamental para uma análise crítica responsável da massa de informações dessa natureza existente no mundo atual.

b) Apresentação do Plano de Ensino

b.1. Objetivos do curso:

- b.1.1. Estatística como ferramenta de trabalho;
- b.1.2. Conhecimento básico das fases da análise estatística;
- b.1.3. Compreender o processo de construção de indicadores;
- b.1.4. Conhecer as ferramentas essenciais à análise estatística (pesquisa);

b.2. Estrutura do curso: aulas expositivas e aulas práticas (tutoriais e exercícios);

- b.2.1. Privilegiar conhecimento conceitual com o necessário de cálculos;
- b.2.2. Abordagem de interpretação de conceitos e uso de ferramentas matemáticas;

b.3. Bibliografia: indicada no plano de ensino e resumos no Infolabo.

c) Introdução à Estatística

c.1. Importância do estudo da Estatística na formação profissional:

- c.1.1. O raciocínio estatístico é muito utilizado no governo e na administração;
- c.1.2. O conhecimento estatístico serve para tomar decisões fundamentadas;
- c.1.3. Utilização em outras disciplinas do curso;
- c.1.4. Interpretação de resultados publicados em jornais, revistas profissionais e artigos científico.
- c.1.5. Usar a interpretação estatística nos artigos da imprensa e no cotidiano.
- c.1.6. Exemplos da “estatística” na mídia em geral
 - c.1.6.1. Aumento de 100% no consumo de frutas por crianças:
 - como foi feita a pesquisa – onde foi feita – quais são os números brutos
 - c.1.6.2. Pesquisa comprova efeitos benéficos do vinho – constestações
 - variáveis de confusão (o vinho, o estilo de vida, a alimentação, etc.)

c.2. Técnicas e métodos envolvendo planejamento, coleta de dados, inferência (previsões), processamento, análise e disseminação das informações obtidas;

c.3. Técnicas para abordarmos cientificamente situações de incerteza – busca evidências quantitativas de fenômenos em geral;

c.4. Cuidados: “se torturarmos apropriadamente os números, eles confessam qualquer coisa” – “Como dizer mentiras usando a Estatística”;

- c.4.1. Importância de conhecimento da Estatística para o correto entendimento do significado e limitações das “estatísticas”;

c.5. Desenvolvimento de raciocínio abstrato, avaliação crítica, associação, dedução e síntese; domínio de ferramentas básicas de computação (Excel / outros);

- c.5.1. Habilidade com os números (cálculo) não será enfatizada neste curso.

- c.5.1.1. Softwares apropriados fazem o trabalho pesado.
- c.5.2. O objetivo é chegar a uma compreensão suficiente para entender cada uma das etapas do método estatístico, o processo como um todo, suas limitações, e possibilitar a comunicação com um profissional da Estatística.
- c.6. A Estatística estuda fenômenos com um conjunto de muitos indivíduos, com pelo menos uma característica comum;
 - c.6.1. “Coleção de métodos para planejar, obter e organizar dados, resumi-los, interpretá-los e deles extrair conclusões”.
- c.7. Definição, segundo Antonio A. Crespo: “Estatística é uma parte da matemática aplicada que fornece métodos para a coleta, a organização, a descrição, a análise e a interpretação de dados quantitativos e qualitativos, e a utilização desses dados para a tomada de decisão.”
- c.8. Outras definições: Uma técnica para:
 - c.8.1. lidar com a coleta, análise, interpretação e apresentação de dados;
 - c.8.2. extrair significados a partir de dados;
 - c.8.3. lidar com as incertezas e planejar o futuro com segurança razoável;
 - c.8.4. para fazer inferências sobre o desconhecido;
 - c.8.5. É a arte de aprender a partir de dados (Ross, 2010).

d) Estatística Aplicada

- d.1. Termo “estatística” vem de “status”, em latim “Estado”, por ser feito pelos governantes para auxiliar a condução do Estado;
- d.2. Antiguidade: registro de número de habitantes, de nascimentos, de óbitos, estimativas de riqueza individual e social; distribuição de terras ao povo, cobrança de impostos, levantamentos em geral (censo – estatística);
- d.3. Idade média: registro de dados referentes à população (batizados, casamentos, funerais), ao comércio, produção de bens e alimentos (em 1085, Guilherme da Inglaterra realizou levantamento de informações sobre proprietários de terras, empregados e animais com objetivo de cobrança de impostos);
- d.4. Renascimento: desenvolvimento das ciências em geral, incluindo a Matemática. No século XVIII tornou-se campo específico do conhecimento sob o nome de Estatística (por Godofredo Achenwall) – desenvolvimento dos cálculos de probabilidades;
- d.5. Tecnologia: desenvolvimento de processos e técnicas avançadas e utilização de computadores e supercomputadores;
- d.6. Exemplos atuais: situação social de comunidades em geral, censo IBGE, pesquisa eleitoral, caracterização geográfica de epidemias, etc. Os resultados são utilizados como subsídios para elaboração de políticas e intervenções do Estado
- d.7. Pesquisa científica: investigação de fenômenos em geral, buscando tendências reveladas pelos dados obtidos.

e) Censo / Amostragem

- e.1. Censo: quando todos os indivíduos da população são estudados;
- e.2. Amostragem: quando parte da população é estudada.

2. O MÉTODO ESTATÍSTICO, SUAS CARACTERÍSTICAS E LIMITAÇÕES

1. O Método Estatístico

- 1.1. Refere-se a todo o processo de utilizar o método científico para responder questões e tomar decisões.
- 1.2. Envolve projetar estudos, coletar dados “bons”, descrever os dados com números e gráficos, analisar os dados, elaborar conclusões, eventualmente tomar decisões.
- 1.3. “Deixar os dados falarem por si próprios” – “Ouvir o que os dados relatam”
 - 1.3.1. Mas os dados só falam claramente quando são organizados, resumidos, apresentados, e os auxiliamos falar propondo questões apropriadas.

2. Etapas de um Estudo

- 2.1. Definição do problema (formulação, meta, revisão literatura, tipo de estudo)
- 2.2. Planejamento (o que, onde, como, quando, quem, quanto)
- 2.3. Coleta / tabulação dos dados
- 2.4. Descrição dos dados (resumo e apresentação – tabelas e gráficos)
- 2.5. Análise dos dados (técnicas de análise estatística)
- 2.6. Interpretação dos dados (extrair conclusões)

3. Tipos de Estudo

- 3.1. Observacional: os dados são coletados de indivíduos (pessoas ou coisas) de maneira tal que não os afeta. Por exemplo, questionários apresentados a indivíduos selecionados “de forma amostral” de uma população de interesse.
 - 3.1.1. Podem ser conduzidos de diversas formas: e-mail, web sites, telefone, etc.
 - 3.1.2. Se conduzidos não apropriadamente, podem resultar em informação “enviesada” (tendenciosa). Por exemplo, indução por escolha não apropriada das palavras ou formulação nas questões.
 - 3.1.3. Podem relatar relações entre as variáveis encontradas. Não podem estabelecer relações de causa e efeito. Por exemplo, pessoas que relatam tomar mais de duas latas de coca-cola no dia por relatam dormir menos horas por noite. Não se pode concluir que é a coca-cola que faz perder o sono. Outras variáveis não estudadas podem ser responsáveis pelo efeito (festas).
- 3.2. Experimento: impõe um ou mais tratamentos aos indivíduos (pessoas ou coisas), de tal modo que uma clara comparação pode ser feita (entre o antes e o depois do tratamento). Por exemplo, um estudo sobre o efeito da dosagem de um anti-hipertensivo na pressão sanguínea: um grupo recebe dose de 10 mg, outro recebe dose de 20 mg, outro 30 mg, etc. Um grupo “de controle” recebe placebo.
 - 3.2.1. Devem ser conduzidos em ambiente controlado e projetados para minimizar vieses que possam ocorrer.
 - 3.2.2. Possíveis problemas: pesquisadores sabem quem se submeteu a cada tratamento, alguma característica não levada em conta pode afetar o resultado (peso dos indivíduos quando estudando a dose do medicamento).
 - 3.2.3. Quando bem projetado e conduzido cuidadosamente, podem-se estabelecer relações de causa e efeito;
- 3.3. Dados publicados: estudos sobre dados já publicados, disponíveis em diversas fontes de dados, por exemplo, IBGE; INEP, etc.

4. Coleta de Dados

- 4.1. Tendo definido os objetivos do estudo, escolhido as variáveis a serem analisadas, definido o tipo de estudo (observacional ou experimental), chega-se à escolha da amostra – processo de amostragem.
- 4.2. Selecionando uma boa amostra: “Entra lixo, sai lixo” – se a amostra é tendenciosa, os resultados também serão tendenciosos. É necessário garantir a “aleatoriedade da seleção” dos indivíduos (pessoas ou coisas) que comporão a amostra.
 - 4.2.1. Pesquisas em web sites, em programas de TV (ligue e dê sua opinião) resultam de amostra “auto-selecionada” e não têm validade científica.
- 4.3. Diversos métodos são utilizados para selecionar amostras aleatórias: aleatória simples, sistemática, estratificada, por conglomerados (aula de amostragem).
- 4.4. Evitando viés (tendência) em seus dados: a amostra deve ser representativa da população que se quer estudar. Alguns possíveis equívocos:
 - 4.4.1. pesquisa de satisfação no trabalho, por telefonema em residência entre 09:00 e 17:00 – não atinge aqueles que trabalham durante o dia fora de casa, aqueles que não tem telefone, etc.
 - 4.4.2. questionários muito extensos: a partir de um certo ponto as respostas se tornam cada vez menos confiáveis – respostas sem maior reflexão.
- 4.5. A coleta de dados propriamente dita: momento crucial do processo. Deve-se garantir que os dados sejam cuidadosamente coletados, buscando-se minimizar o máximo possível as falhas (respostas incompletas, leituras apressadas, etc.).

5. Descrição dos Dados

- 5.1. Em geral, a coleta de dados resulta em grande volume de dados. A descrição desses dados objetiva visualizá-los de forma global e resumida. É comumente realizada de duas formas: visual (gráficos) e numérica.
- 5.2. Estatística Descritiva: números que descrevem importantes aspectos do conjunto de dados. Em geral são acompanhados de gráficos ou mapas.
 - 5.2.1. Se os dados são categóricos (grupos como sexo, partido político, etc.), são comumente resumidos pelo número de indivíduos em cada grupo (frequência), ou porcentagem de indivíduos em cada grupo (frequência relativa).
 - 5.2.2. Se os dados são numéricos (medidas: peso, altura, idade, etc.), além dos agrupamentos acima (faixa de peso, idade, etc.), mais aspectos podem representar os dados: tendência central (média, mediana), medidas de dispersão (variância, desvio padrão, extensão), e em certos casos, a relação entre duas variáveis (peso e altura).
 - 5.2.3. Obs: em alguns casos, números são utilizados para representar variáveis categóricas (sexo M=1, sexo F=2) – mas isso não significa que os dados são numéricos e não faz sentido calcular média desses dados.
- 5.3. Gráficos e mapas: resumem os dados em uma forma visual. Exemplos: distribuição por gênero, uso de celular por faixa de idade, pesquisas de opinião em geral.
 - 5.3.1. Dados categóricos: principalmente pizza e barras
 - 5.3.2. Dados numéricos: principalmente histograma e linha

6. Análise dos dados

- 6.1. Análise Estatística: existem diversos tipos; é fundamental escolher a apropriada.
 - 6.1.1. Escolha do método não apropriado leva a resultados incorretos

- 6.1.2. Cenários onde um número fixo de tentativas independentes resulta em sucesso ou fracasso utilizam a distribuição binomial;
- 6.1.3. Cenários onde a distribuição segue uma curva com formato de sino usam a distribuição normal;
- 6.2. Intervalos de Confiança: são intervalos em torno da média amostral onde se pode ter uma confiança determinada (95%, por exemplo) de se encontrar a média da população;
- 6.3. Teste de Hipóteses: testa se a diferença de médias entre duas amostras podem indicar diferença de média entre duas populações (se a amostra é pequena - $n < 30$ - e a distribuição é “normal”, utiliza-se o teste t);
- 6.4. Correlação: examina a relação entre duas variáveis numéricas (peso e altura) utilizando correlação e regressão linear simples; se as variáveis são categóricas (sexo, cidade onde nasceu, etc.) utiliza-se outro tipo de análise.

7. Interpretação dos dados (extrair conclusões)

- 7.1. O que os resultados significam: atribuição do pesquisador;
- 7.2. Equívocos comuns:
 - 7.2.1. Generalizar resultados que são válidos em condições particulares
 - 7.2.2. Resultados obtidos com amostras tendenciosas;
- 7.3. Para se chegar a conclusões sólidas, é crucial ter conhecimento de todas as etapas do método estatístico, seu significado, suas limitações e sua validade.

8. Conceitos Básicos

- 8.1. Ramos da Estatística:
 - 8.1.1. Estatística Descritiva: a observação de fenômenos de mesma natureza, a coleta de dados numéricos relativos a esses fenômenos, a organização e a classificação desses dados observados e a sua apresentação através de gráficos e tabelas, além da descrição desses dados através do cálculo de coeficientes;
 - 8.1.2. Estatística Inferencial ou Dedutiva: processo de elaboração de conclusões baseadas em informações contidas em uma amostra. A análise e interpretação de dados amostrais que possibilita conclusões sobre a população. Esse processo de generalização está associado a uma margem de incerteza.
 - 8.1.2.1. Inferência estatística: técnicas para chegar a conclusões sobre uma população baseando-se na informação contida na amostra;
- 8.2. População: conjunto de “indivíduos” que apresentam algumas características em comum, sobre os quais desejamos obter alguma informação particular;
 - 8.2.1. Indivíduo: são os objetos descritos por um conjunto de dados. Podem ser pessoas, animais ou coisas;
- 8.3. Amostra: subconjunto da população, definidos segundo algum critério estabelecido pelo observador, sobre o qual se conduz algum estudo com objetivo de se fazer inferências estatísticas sobre toda a população;
- 8.4. Parâmetro: uma medida numérica que descreve uma característica da população
- 8.5. Estatística: uma medida numérica que descreve uma característica da amostra
- 8.6. Variável: uma característica de interesse de elementos individuais de uma população ou amostra que será analisada (Indicadores definidos pelo observador, utilizados para se “medir” alguma característica da amostra); podem ser Quantitativas e Qualitativas

- 8.6.1. Observação: é o valor de uma variável para um elemento particular da amostra
- 8.6.2. Conjunto de dados: é o conjunto das observações de uma variável para todos os elementos da amostra.
- 8.7. Medidas Descritivas: “ouvir o que os dados falam”
 - 8.7.1. Medidas de tendência central: média, mediana, moda
 - 8.7.2. Medidas de dispersão: extensão, variância, desvio padrão
 - 8.7.3. Medidas de variabilidade: percentil, decil, quartil
 - 8.7.4. Medidas de assimetria: curtose, assimetria
- 8.8. Teste de hipóteses: teste t, teste z; erros tipo I e tipo II
- 8.9. Inferência: erro padrão, intervalo de confiança
- 8.10. Regressão e Correlação: medidas de associação entre dois parâmetros
- 8.11. Estatística não paramétrica: a variável ou os dados amostrais não são “normais”, ou são ordinais ou nominais.

3. TIPOS DE DADOS (QUALITATIVOS, QUANTITATIVOS, DISCRETOS, CONTÍNUOS, ETC.)**1. Variável, Observação e Data Set**

- 1.1. Variável: uma característica de interesse referente aos elementos individuais de uma população ou amostra.
- 1.2. Observação: o valor de uma variável de um elemento particular de uma amostra ou população.
- 1.3. Data Set: conjunto de observações de uma variável de elementos da amostra

2. Variável Qualitativa:

- 2.1. A descrição da característica de interesse resulta em valor não numérico;
- 2.2. Podem ser codificadas em números para análise, mas os números não podem sofrer operações entre si.
 - 2.2.1. Exemplos: cor dos olhos, marca de cerveja,
- 2.3. Escala Nominal: dados utilizados para identificação de categorias.
 - 2.3.1. Não podem ser ordenados em algum esquema.
 - 2.3.2. Operações aritméticas (+, -, *, /) não podem ser realizadas nesses dados.
 - 2.3.3. Exemplos: refrigerante preferido, bairro mais populoso, ...
- 2.4. Escala Ordinal: dados que podem ser arranjados em alguma ordem,
 - 2.4.1. As diferenças entre os valores não podem ser determinadas ou não tem sentido;
 - 2.4.2. Operações aritméticas não podem ser realizadas nesses dados;
 - 2.4.3. Exemplos: hierarquia de cargos na empresa, nível sócio-econômico, classe de renda familiar;

3. Variável Quantitativa:

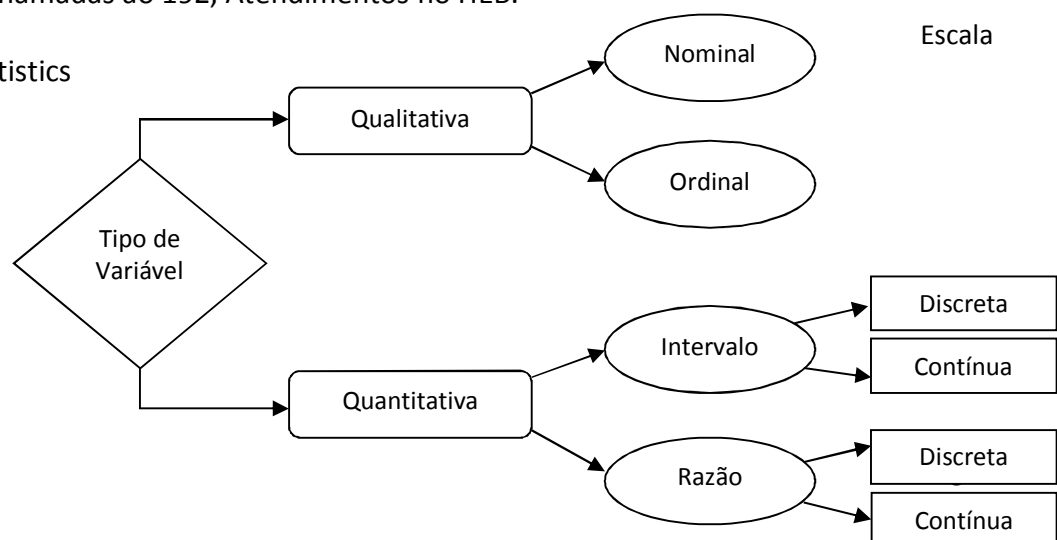
- 3.1. Quando a descrição da característica de interesse resulta em valor numérico;
 - 3.1.1. Exemplos: número de participantes de um evento, volume de chuva por dia; número de acidentes em determinada rodovia em determinado período;
- 3.2. Discreta: é uma variável quantitativa cujos valores são contáveis, isto é, resulta de um processo de contagem;
 - 3.2.1. Exemplos: n. de alunos na turma, n. veículos por hora no pedágio;
 - 3.2.2. Algumas variável discretas podem ser consideradas contínuas, embora rigorosamente sejam discretas: valores possíveis de nota de prova (0 a 10 com uma casa decimal: 101 valores) – é considerada contínua.
- 3.3. Contínua: é uma variável quantitativa que pode assumir qualquer valor numérico em um intervalo ou em diversos intervalos. Resulta de uma medida.
 - 3.3.1. Exemplos: peso, altura, renda familiar,
- 3.4. Escala de Intervalos: dados que podem ser arranjados em alguma ordem, e para os quais a diferença entre os valores é significativa.
 - 3.4.1. O nível de intervalo da medida resulta de uma contagem ou medição.
 - 3.4.2. Diferenças podem ser calculadas e interpretadas.
 - 3.4.3. O valor zero é arbitrariamente escolhido para um intervalo
 - 3.4.3.1. Não implica em ausência da característica sendo medida.
 - 3.4.4. Razões não são significativas para essa escala.
 - 3.4.5. Exemplos: temperatura, notas de prova,
- 3.5. Escala de Razão: dados que podem ser ordenados e para os quais as operações aritméticas podem ser realizadas;
 - 3.5.1. A divisão por zero é excluída;
 - 3.5.2. Resulta de contagem ou medição;

- 3.5.3. Diferenças e razões podem ser calculadas e interpretadas;
- 3.5.4. Tem um zero absoluto e o valor zero indica uma completa ausência da característica de interesse.
- 3.5.5. Exemplos: peso, quantidade de sal na alimentação, número de TVs em casa,

4. Exercícios:

- 4.1. Indivíduo: objeto descrito por um conjunto de dados – pessoas ou coisas
- 4.2. Variável: uma característica do indivíduo – valores diferentes para indivíduos
- 4.3. População e Amostra
 - 4.3.1. Todos os eleitores registrados / levantamento com 600 eleitores
 - 4.3.2. Proprietários de armas / entrevista com 1000 proprietários
 - 4.3.3. Famílias sustentadas por um único parceiro / questionários enviados
 - 4.3.4. Gerentes de empresas comerciais de RP / entrevista com 50 gerentes
 - 4.3.5. Produção diária de bolachas / amostras aleatórias de 30 em 30 minutos
- 4.4. Estudo Observacional / Estudo Experimental
 - 4.4.1. Entrevista com presidiários mostram que 85% sofreram algum tipo de violência doméstica na infância. Pode-se concluir que violência doméstica leva à criminalidade?
 - 4.4.2. Administração progressiva de álcool resulta em progressiva perda de reflexos – pode-se concluir que consumo de álcool leva a perda de reflexos?
- 4.5. Variável Quantitativa: Discreta e Contínua
 - 4.5.1. Número de agulhas defeituosas em uma caixa
 - 4.5.2. Número de indivíduos em determinado grupo com personalidade tipo A
 - 4.5.3. Número de questionários não respondidos
 - 4.5.4. Número de presidiários que não completaram o segundo grau
 - 4.5.5. Número de lances de dados até que surja o número 6
 - 4.5.6. Tempo de prisão atribuído ao crime de assassinato doloso
 - 4.5.7. Renda familiar na comunidade habitacional Parque dos Lagos
 - 4.5.8. Taxa de álcool no sangue 5 min. após beber duas latas de cerveja
- 4.6. Variável Qualitativa: Nominal e Ordinal
 - 4.6.1. Nominal: Estado civil, Gênero, Tipo de personalidade, Tipo sanguíneo, Cidade onde nasceu;
 - 4.6.2. Ordinal: Nível de dor, Classe de automóvel (tamanho), Avaliação de produtos (ruim a excelente), Classe de crime, classe socioeconômica, Categoria futebolística;
- 4.7. Variável quantitativa: Intervalar (diferenças têm sentido) e Razão
 - 4.7.1. Intervalar: Resultados de testes de QI, Temperatura ambiente, Resultados de provas ou testes;
 - 4.7.2. Razão: (operações aritméticas): Consumo de sal na alimentação, Número de chamadas ao 192, Atendimentos no HEB.

Beginning Statistics
pag. 16



4. AMOSTRAGEM, COLETA E ORGANIZAÇÃO DOS DADOS

1. **Regra Básica: evitar viés (desvios): Amostra deve ser absolutamente aleatória**
 - 1.1. Cada indivíduo da população deve ter a mesma chance de ser escolhido
 - 1.2. Amostras auto-selecionadas não têm validade científica
2. **Definição da População-Alvo: Todo o grupo de indivíduos que você quer estudar**
 - 2.1. Deve ser bem definida para que a amostra possa ser bem definida
 - 2.2. Opinião dos usuários de transporte coletivo em Ribeirão Preto: todos os usuários
3. **Definição da Amostra: deve ser representativa da população-alvo**
 - 3.1. Porcentagem da população, segundo erro amostral suportável 2% a 5%
 - 3.2. Erro amostral: fórmulas aproximadas e fórmulas mais precisas
 - 3.2.1. Fórmulas aproximadas: tamanho da população e erro suportável
 - 3.2.2. Fórmulas mais precisas: tamanho da população e variância na população
 - 3.2.2.1. Necessário estudo-piloto antes de definir a amostra

$$E = \frac{1}{\sqrt{n}}$$

$$n = \frac{1}{E^2}$$

$$n = \frac{N * \frac{1}{E^2}}{N + \frac{1}{E^2}}$$

n: tamanho da amostra
N: tamanho da população
E: erro amostral (decimal)

$$n = \left[\frac{Z_{\alpha/2} * \sigma}{E} \right]^2$$

n: tamanho da amostra
 $Z_{\alpha/2}$: Valor z para intervalo de confiança
 σ : desvio padrão populacional
E: erro amostral (decimal)

$$E = \sqrt{\frac{N - n}{N * n}}$$

- 3.3. Amostra deve ser escolhida aleatoriamente dentro da população-alvo
 - 3.3.1. Amostras não aleatórias trazem viés (distorção, tendências) nos resultados
 - 3.3.2. Amostragem por telefone celular: pessoas com mais de 1 celular (aceitável?)
 - 3.3.3. Pesquisas em websites: amostra acessa internet, aquele site, decidem responder
 - 3.3.4. Amostra deve ser analisada buscando-se possíveis vieses
 - 3.3.4.1. Estudo de hábitos de estudo pesquisando freqüentadores da biblioteca
 - 3.3.4.2. Estudo de um novo tipo de cirurgia de catarata: ética x voluntários
 - 3.3.4.3. Solução: grupo de tratamento e grupo de controle similares (duplo cego)
- 3.4. Tipos de Amostragem
 - 3.4.1. Amostragem aleatória simples: escolha aleatória dentro da população
 - 3.4.2. Amostragem aleatória estratificada (cluster): separação em estratos uniformes e dentro de cada estrato a escolha é aleatória (n pessoas em cada extrato)
 - 3.4.3. Amostragem por Conglomerados: semelhante à estratificada simples, mas proporcional ao número de elementos em cada conglomerado (bairros)
 - 3.4.4. Amostragem sistemática: cada 10º elemento em uma linha de produção
 - 3.4.5. Amostragem a esmo (parafusos na caixa),
 - 3.4.6. Amostragem intencional e amostragem por voluntários (auto-selecionada);
 - 3.4.7. Amostragem por Conveniência (não probabilística): sem rigor estatístico. Seleciona os elementos a que tem acesso, admitindo que estes possam representar a população.
- 3.5. Definida a amostra, as “não respostas” devem ser minimizadas (prêmios, brindes, etc.)
 - 3.5.1. Prêmios ou sorteios ou brindes podem causar viés – atraem um tipo de pessoas
 - 3.5.2. Importância do anonimato e confidencialidade das respostas
- 3.6. Avaliação crítica: Amostra planejada x Amostra efetiva
 - 3.6.1. Amostra efetivamente consultada deve ser, no mínimo, 70% da amostra planejada
 - 3.6.2. Composição da amostra efetiva deve ser similar à da amostra planejada
 - 3.6.3. Importância da busca de outros possíveis desvios entre planejamento e realidade
 - 3.6.4. Reportagem do New York Times afirma que consumo de maconha é muito mais efetivo no alívio dos efeitos da quimioterapia do que outras drogas:
 - 3.6.4.1. 29 pacientes, 15 com maconha, 14 com placebo (12 terminaram estudo)

4. Importância do Efeito-Placebo

- 4.1. Efeitos fisiológicos causados pela sugestão (psicológicos)
- 4.2. Doenças psicossomáticas: 80% das doenças
- 4.3. Situações éticas onde não se usa placebo, mas tratamentos convencionais

5. Questionários

- 5.1. Elaboração das questões e sua ordem no questionário são fundamentais
 - 5.1.1. Objetividade na elaboração do questionário: foco no objetivo do estudo
 - 5.1.2. Você concorda que o presidente tenha que viajar ao exterior representando o país?
 - 5.1.3. Você apóia a pena de prisão perpétua para seqüestradores (após reportagem)
- 5.2. Pessoas que aplicam o questionário devem ser bem treinadas
 - 5.2.1. Abordagem do entrevistado é fundamental
 - 5.2.2. Dúvidas sobre as perguntas devem ser respondidas da mesma forma por todos
 - 5.2.3. Em grandes levantamentos, estudos-piloto em pequena escala são importantes

6. Identificação de variáveis de confusão

- 6.1. São variáveis não incluídas no estudo, mas que influenciam fortemente os resultados
- 6.2. Ex: Pessoas que tomam vitamina C diariamente têm menos gripe:
 - 6.2.1. Estudo observacional, quem toma vitamina C pode ter mais preocupação com a saúde e se cuidar melhor, por isso ter menos gripe
 - 6.2.2. Solução: pessoas com mesmo nível de preocupação com a saúde, divididas aleatoriamente, consome vitamina C, consome Placebo, não consome nada

7. Qualidade dos Dados

- 7.1. Confiabilidade: resultados são reprodutíveis em medidas subseqüentes
 - 7.1.1. Medidas de pressão sanguínea, açúcar no sangue, temperatura corporal, etc.
- 7.2. Dados não enviesados: sem favoritismo sistemático em certas respostas individuais.
 - 7.2.1. Erros no instrumento de medida, amostras viciadas, respostas induzidas
 - 7.2.2. Vieses devem ser prevenidos – são de difícil correção depois de colhidos os dados
- 7.3. Validade: dados representam o que eles têm a intenção de medir
 - 7.3.1. Prevalência de crimes em um bairro/cidade não é o número absoluto de crimes no bairro/cidade, mas o número de crimes por habitante do bairro/cidade

8. Pontos a observar em um estudo:

- 8.1. Gráficos tendenciosos (equivocados)
- 8.2. Dados enviesados (tendenciosos)
- 8.3. Margem de erro não reportada (maior do que se pensa)
- 8.4. Tamanho da Amostra (pode ser insuficiente, levando a erro amostral elevado)
- 8.5. Amostra não aleatória (levam a tendências nos resultados)
- 8.6. Dados amostrais perdidos (nem toda a amostra planejada foi analisada)
- 8.7. Correlações equivocadas
 - 8.7.1. correlação: dados quantitativos (peso e altura) – não implica causa-efeito
 - 8.7.2. associação: dados qualitativos (votos femininos para o candidato)
 - 8.7.3. força de uma correlação – o R^2 não relatado
- 8.8. Variáveis de confusão (não previstas no estudo, mas com grande influência)
 - 8.8.1. Frequentadores de academia tem melhor saúde (estilo de vida: muitas variáveis)
- 8.9. Resultados reportados parcialmente
 - 8.9.1. Resultados estatisticamente significativos: em 1 dos 50 testes realizados?
 - 8.9.2. Resultados individuais – Artista X perde 10 kg em uma semana com a dieta Y

9. Organização dos Dados: tabulação em dois tipos

- 9.1. Cada linha é um registro (entrevistado), cada coluna uma pergunta (variável)
- 9.2. Cada linha é uma pergunta (variável), cada coluna um registro (entrevistado)

5. CONSTRUÇÃO E INTERPRETAÇÃO DE TABELAS E GRÁFICOS ESTATÍSTICOS

1. Organização dos Dados: Construção de Tabelas

- 1.1. Cada linha é um registro (entrevistado), cada coluna uma pergunta (variável)
- 1.2. Cada linha é uma pergunta (variável), cada coluna um registro (entrevistado)
- 1.3. Digitação de tabelas em Planilha Eletrônica (Excel):
 - 1.3.1. Pode-se digitar no formato 1.2 e copiar e colar transposto para o formato 1.1.

1.1. Tabela mais apropriada para análise estatística, particularmente em programas de estatística

Ficha	nome	idade	Altura	sexo	Cidade	Escolaridade	Pontuação
1	AAS	17	1,62	F	Rio Preto	1º grau	110
2	CMM	72	1,70	M	Barrinha	2º grau	81
3	CNS	26	1,55	F	Jardinópolis	3º grau	162

1.2. Tabela preferida por alguns para digitação das fichas (sequência com Enter)

Ficha	1	2	3	Tipo de Variável	Escala
nome	AAS	CMM	CNS	Qualitativa Nominal	
Idade (anos)	17	72	26	Quantitativa Razão	Discreta
Altura (m)	1,62	1,70	155	Quantitativa Razão	Contínua
sexo	F	M	F	Qualitativa Nominal	
Cidade	Rio Preto	Barrinha	Jardinópolis	Qualitativa nominal	
Escolaridade	1º grau	2º grau	3º grau	Qualitativa Ordinal	
Pontuação (0 a 200)	110	81	162	Quantitativa Intervalo	Contínua

- 1.4. Obs: Escala de Likert: 1 a 5 (discordo fortemente – concordo fortemente)
 - 1.4.1. Trata-se de uma escala Qualitativa Ordinal
 - 1.4.2. Prática comum: considerar como quantitativa - intervalar e até mesmo razão
 - 1.4.2.1. Avaliação docente: média para efeito de comparação
- 1.5. Importante:
 - 1.5.1. Numerar cada ficha e lançar os dados na linha/coluna correspondente
 - 1.5.1.1. Possibilita a revisão e verificação posterior da digitação
 - 1.5.2. Definir unidade e faixa de variação de cada medida
 - 1.5.3. Revisar tabela para localizar erros de digitação (dados fora da faixa de variação)
 - 1.5.4. Dados perdidos: deixar em branco ou inserir a média da variável (bem identificada)
- 1.6. Para análise em alguns softwares de Estatística:
 - 1.6.1. Alguns programas exigem que as variáveis categóricas sejam reduzidas a números
 - 1.6.1.1. Sexo M = 1; F=0 Cidade: Barrinha = 1; Rio Preto = 2; Jardinópolis = 3, etc.

2. Medidas Descritivas de Dados Observados (Estatística Descritiva)

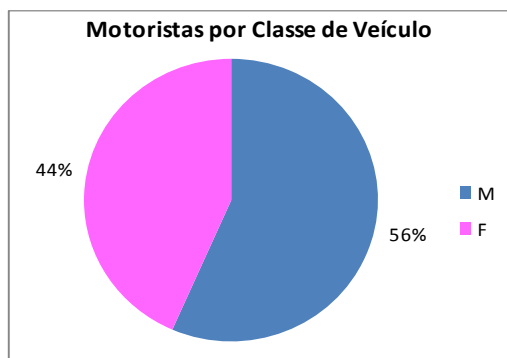
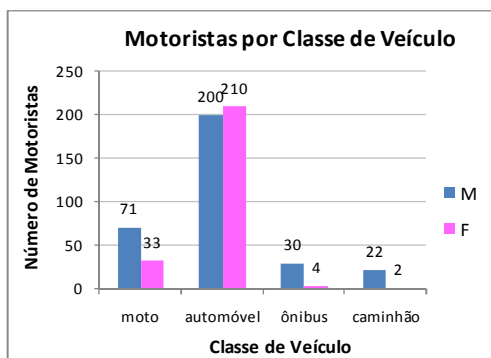
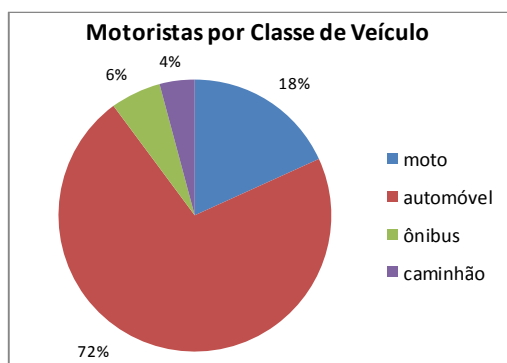
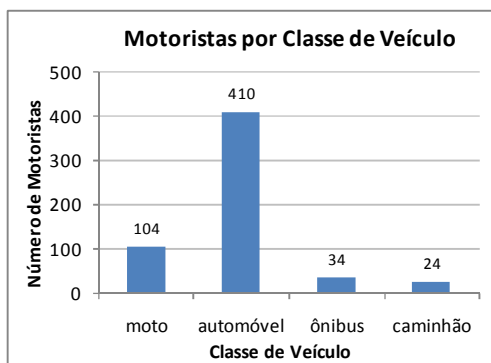
- 2.1. Estatística Descritiva: resume algumas características dos dados coletados
- 2.2. Dados Qualitativos: agrupa indivíduos em classes
 - 2.2.1. Frequência: nº de dados em cada classe (contagem simples)
 - 2.2.2. Frequência relativa: porcentagem de indivíduos em cada classe
 - 2.2.2.1. Total da classe dividido pelo total geral
 - 2.2.3. Tabelas cruzadas: duas variáveis correlacionadas, por exemplo, sexo e veículo
- 2.3. Quantitativos: frequência (grupos convencionados), frequência relativa (%) e também medidas centro (média, mediana, moda), dispersão (variância, desvio), correlação

Frequência Absoluta				
MOTORISTA	moto	automóvel	ônibus	caminhão
M	71	200	30	22
F	33	210	4	2
SubTotal	104	410	34	24
TOTAL	572			

Frequência Relativa					
MOTORISTA	moto	automóvel	ônibus	caminhão	SubTotal
M	12%	35%	5%	4%	56%
F	6%	37%	1%	0%	44%
SubTotal	18%	72%	6%	4%	100%

3. Gráficos

- 3.1. Obs: existem muitos tipos de gráfico, depende de o que se quer evidenciar
- 3.2. Frequência: Qualitativos: gráficos de barra
- 3.2.1. Eixo x: classe - eixo Y: contagem dentro da classe
- 3.2.1.1. No gráfico há espaço entre as colunas (barras)
- 3.3. Quantitativos: Histograma
- 3.3.1. Eixo x: classe - eixo Y: contagem dentro da classe
- 3.3.2. A classe é definida pelo pesquisador (intervalo de idade, de altura, etc.)
- 3.3.2.1. No gráfico não há espaço entre as colunas (barras)
- 3.4. Frequência relativa: gráficos de pizza
- 3.4.1. Porcentagem de elementos dentro da classe
- 3.4.1.1. Total da classe dividido pelo total geral, expresso em %
- 3.5. Correlação: gráficos xy

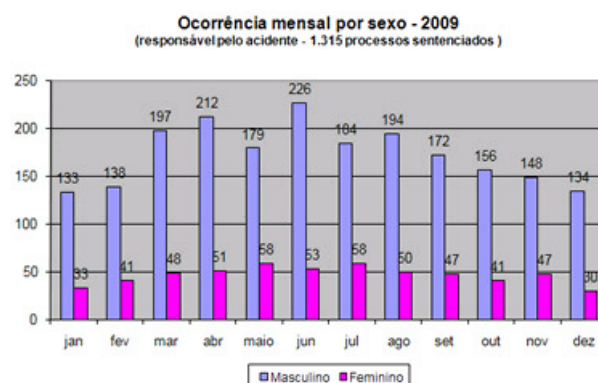
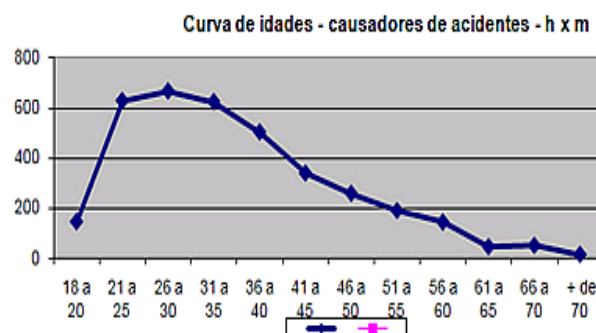


Obs: número absoluto, não informa número de motoristas em cada idade

3.6. Determinação do número de classes (i= número de classes; n = número de elementos)

3.6.1. Critério da Raiz: $i \cong \sqrt{n}$

3.6.2. Fórmula de Stuges: $i \cong 1 + 3,3 \log n$



Atenção: os dois gráficos acima indicam número absoluto de ocorrências, sem considerar o número de pessoas em cada classe (idosos que dirigem, homens e mulheres que dirigem). Seria importante incluir gráficos com a porcentagem dentro de cada classe.

Exemplo de gráfico mais complexo, que resume muitas informações

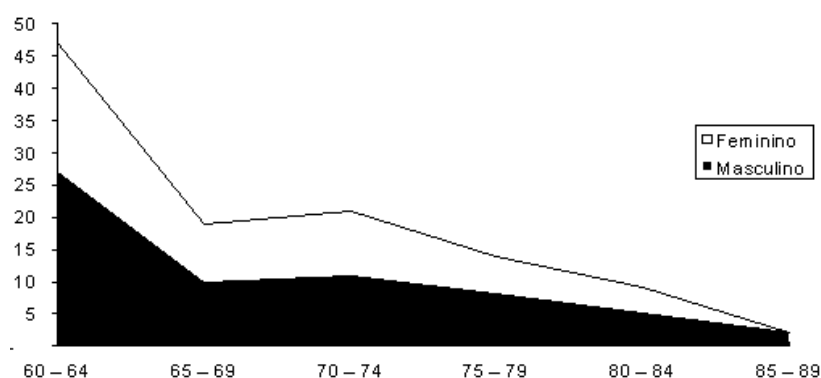
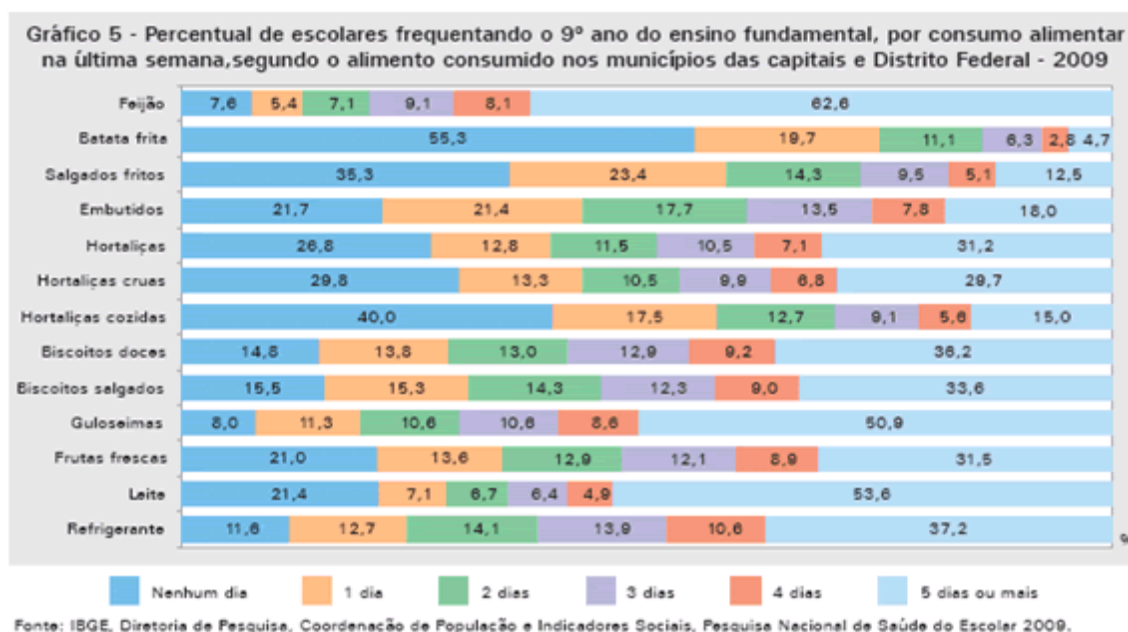


Figura 1 - Distribuição do número de idosos acidentados no trânsito, segundo sexo e idade, Ribeirão Preto-SP, 1998

O gráfico acima também não indica a frequência relativa, isto é, não considera o número absoluto de pessoas em cada classe. Isso não significa que o gráfico está incorreto e sim que o foco do gráfico está no número absoluto de pessoas acidentadas.

4. Histograma

Table 3-1 Colorado Live Births by Mother's Age									
Year	Total births	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1975	40,148	88	6,627	14,533	12,565	4,885	1,211	222	16
1980	49,716	57	6,530	16,642	16,081	8,349	1,842	198	12
1985	55,115	90	5,634	16,242	18,065	11,231	3,464	370	13
1990	53,491	91	5,975	13,118	16,352	12,444	4,772	717	15
1995	54,310	134	6,462	12,935	14,286	13,186	6,184	1,071	38
2000	65,429	117	7,546	15,865	17,408	15,275	7,546	1,545	93

* Note: The sum of births may not add up to the total number of births due to unknown or unusually high age (50 and over) of the mother.

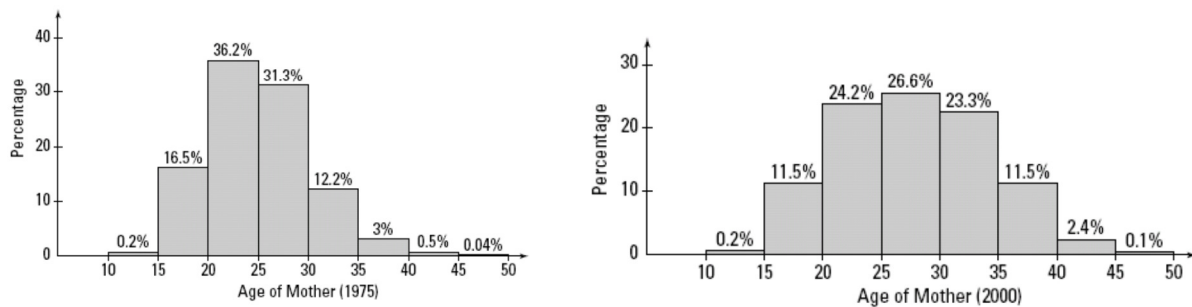
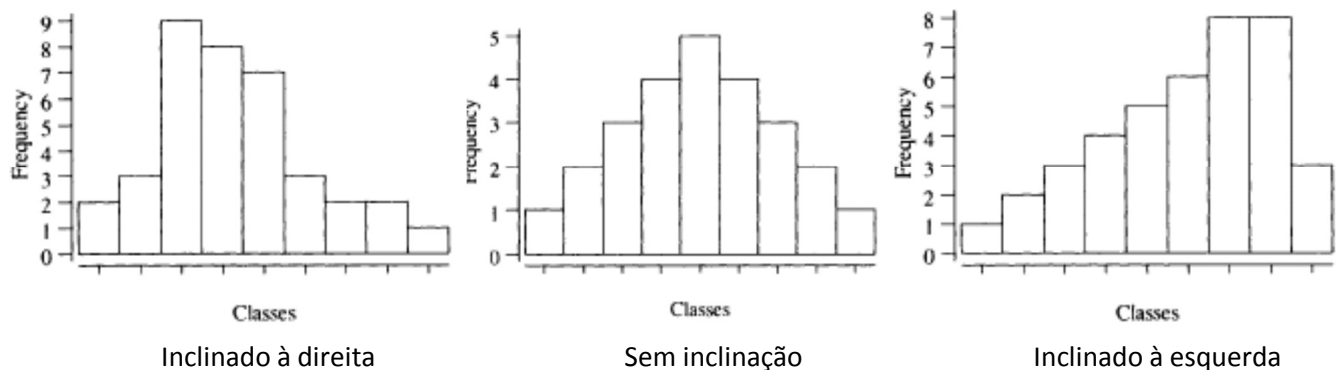


Figure 3-4: Colorado live births, by age of mother for 1975 and 2000.

4.1. Inclinação do Histograma (skewness)



4.2. Frequência Acumulada

Tabela 1: Horas assistindo TV por dia (pesquisa com 20.200 crianças nas férias)

Hours	Frequency	Upper limit	Cumulative frequency
0	0	0	0
0-1	4,300	1	4,300
1-3	6,900	3	4,300+6,900 = 11,200
3-5	4,900	5	4,300+6,900+4,900 = 16,100
5-10	2,000	10	4,300+6,900+4,900+2,000 = 18,100
10-24	2,100	24	4,300+6,900+4,900+2,000+2,100 = 20,200

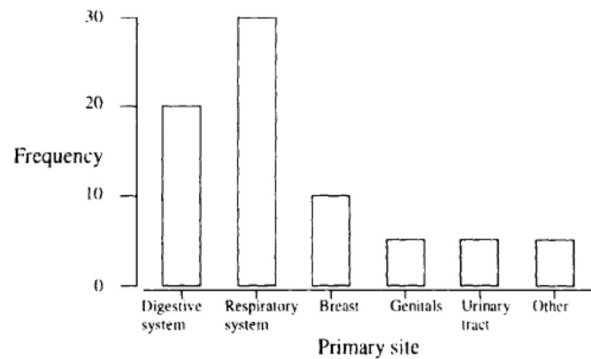


5. Exemplos:

5.1. Dados Qualitativos:

Table 2.3

Primary site	Frequency
Digestive system	20
Respiratory	30
Breast	10
Genitals	5
Urinary tract	5
Other	5



5.2. Dados Quantitativos

Table 2.5

IQ score	Frequency
80-94	8
95-109	14
110-124	24
125-139	16
140-154	13

Table 2.6

Class limits	Class boundaries	Class width	Class marks
80-94	79.5-94.5	15	87.0
95-109	94.5-109.5	15	102.0
110-124	109.5-124.5	15	117.0
125-139	124.5-139.5	15	132.0
140-154	139.5-154.5	15	147.0

EXAMPLE 2.7 Group the following weights into the classes 100 to under 125, 125 to under 150, and 150 to under 175.

111	120	127	129	130	145	145	150	153	155	160
161	165	167	170	171	174	175	177	179	180	180
185	185	190	195	195	201	210	220	224	225	230
245	248									

Table 2.7

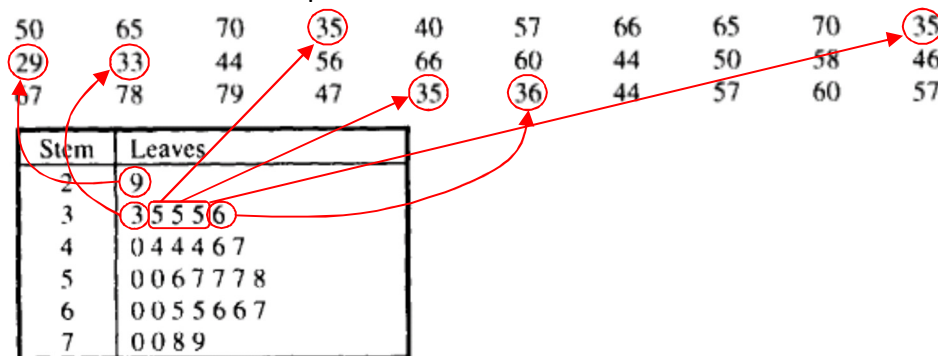
Weight	Frequency
100 to under 125	2
125 to under 150	5
150 to under 175	10
175 to under 200	10
200 to under 225	4
225 to under 250	4

6. Obs: Sistema de Caule e Folhas

1.1. Forma visual de apreender uma distribuição (pouco utilizada)

1.2. Exemplo: Pontuação de 30 estudantes num jogo qualquer (boliche, por exemplo)

1.2.1. Neste caso, registra-se a dezena na coluna da esquerda e cada instância da unidade na linha correspondente à dezena



Dezena Unidade

6. MEDIDAS DESCRITIVAS DE DADOS OBSERVADOS: TENDÊNCIA CENTRAL E DE POSIÇÃO

1. MÉDIA ARITMÉTICA (ou simplesmente Média)

1.1. Somatória de todos os valores (x_i) da amostra dividido pelo número de valores (n):

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

1.2. Medida afetada fortemente por “outliers” – valores discrepantes

1.2.1. Distância da cidade de origem (a maioria é regional, mas um é de Manaus)

1.3. Obs: quando a média se refere à população, utiliza-se a notação:

$$\mu = \frac{\sum_{i=1}^n x_i}{N}$$

1.4. Caso especial: acidentes com operadores utilizando ou não capacetes:

1.4.1. Encontre a média amostral da classificação de severidade de acidentes para motociclistas usando capacete e não usando capacete (obs: acidentes similares)

Classification of accident	Interpretation	Frequency of driver with helmet	Frequency of driver without helmet
0	No head injury	248	227
1	Minor head injury	58	135
2	Moderate head injury	11	33
3	Severe, not life-threatening	3	14
4	Severe and life-threatening	2	3
5	Critical, survival uncertain at time of ac	8	21
6	Fatal	1	6
		331	439

Solution

The sample mean for those wearing helmets is

$$\bar{x} = \frac{0.248 + 1.58 + 2.11 + 3.3 + 4.2 + 5.8 + 6.1}{331} = \frac{143}{331} = 0.432$$

The sample mean for those who did not wear a helmet is

$$\bar{x} = \frac{0.227 + 1.135 + 2.33 + 3.14 + 4.3 + 5.21 + 6.6}{439} = \frac{396}{439} = 0.902$$

Conclusão: a severidade é maior para aqueles que não utilizam capacete (...)

2. MEDIANA (ou valor do meio)

2.1. É o valor que está em posição intermediária na distribuição de valores, isto é, tem tantos elementos acima quanto abaixo de sua posição. Representada por **M** ou \tilde{x}

2.2. Para determinar-se a Mediana, primeiramente deve-se ordenar os **n** valores em ordem crescente (**n** é o número de elementos da amostra)

2.3. Quando **n** é ímpar, a Mediana é um valor real da amostra:

2.3.1. A Mediana é o valor que está na posição:

$$Vp = \frac{n + 1}{2}$$

2.3.2. Exemplo: 33 elementos na amostra ($n=33$) – a mediana está na posição 17

2.4. Quando ***n*** é par, a Mediana é um valor fictício entre dois valores reais da amostra

2.4.1. O valor da Mediana é dado pela média dos valores intermediários:

$$M = \frac{Vp \cdot \frac{n}{2} + Vp \cdot \frac{n}{2} + 1}{2}$$

2.4.2. Exemplo: 34 elementos na amostra ($n=34$) – a mediana é a média dos valores que estão nas posições 17 e 18.

3. MODA (ou valor mais frequente)

3.1. É o valor que aparece mais vezes nas observações da amostra.

3.2. Uma amostra pode ser:

3.2.1. Amodal: não ter moda (não há valores repetidos)

3.2.2. Unimodal: apenas um valor com repetições expressiva (homogênea)

3.2.3. Bimodal: dois valores têm repetições expressivas (heterogênea)

3.2.4. Trimodal: três valores têm repetições expressivas (heterogênea)

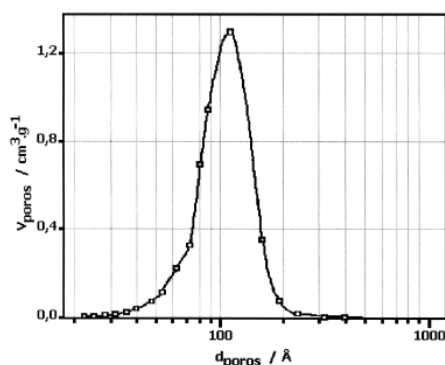


Figura 8: Curva BJH da solução sólida zircônia-ítria.
[Figure 8: BJH curve of the zirconia-yttria solid solution.]

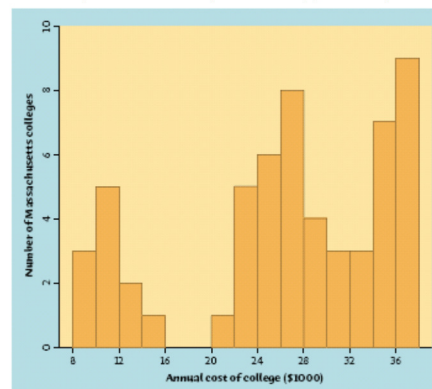


Figure 15 Histogram of the estimated costs (in thousands of dollars) for four-year colleges in Massachusetts. The two clusters distinguish public from private institutions.

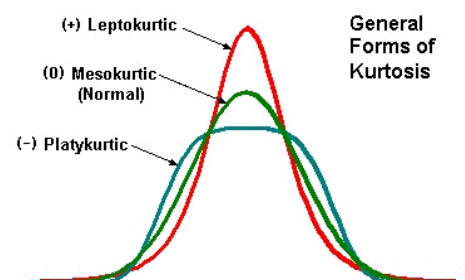
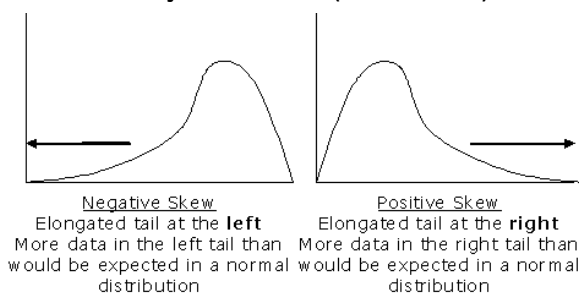
4. RELAÇÃO MÉDIA – MEDIANA

4.1. Distribuições simétricas têm média e mediana próximas

4.2. Distribuições assimétricas tem média e mediana mais distantes

4.3. Inclinação à esquerda (lado maior): média menor que mediana

4.4. Inclinação à direita (lado maior): média maior que mediana



5. Medidas de Posição: PERCENTIL, DECIL, QUARTIL

5.1. São utilizadas para descrever a localização de uma observação particular em relação ao restante do conjunto de dados

5.1.1. Percentil divide as observações (dados) em 100 partes iguais

5.1.2. Decil divide as observações (dados) em 10 partes iguais

5.1.3. Quartil divide as observações em 4 partes iguais

5.1.4. Mediana: divide as observações em 2 partes iguais

5.2. Inicialmente é necessário ordenar as ***n*** observações em ordem crescente (ranking)

6. PERCENTIL

6.1. O percentil para a observação x é encontrado dividindo-se o número de observações que são menores que x pelo número total de observações e multiplicando-se por 100

6.1.1. Se o valor encontrado for inteiro, este é o percentil

6.1.2. Se o valor encontrado não for inteiro, deve ser arredondado ao superior.

6.2. Significado do percentil de uma observação:

6.2.1. Percentil 25: 25% dos valores são menores e ~75% são maiores (grandes n s)

6.3. Exemplo: encontrar o percentil de 66 nos valores abaixo

43, 54, 56, 61, 62, 66, 68, 69, 69, 70, 71, 72, 77, 78, 88, 89, 93, 98, 99, 99.

6.3.1. Temos 20 valores nessa sequência – o valor 66 tem 5 valores abaixo dele:

$$P = \frac{n^{\circ} \text{ elementos} < \text{que o valor}}{n^{\circ} \text{ valores}} \times 100$$

$$P = \frac{5}{20} \times 100 = P25$$

6.4. Para encontrar o percentil P de uma distribuição, multiplica-se o valor do percentil pelo número de observações da amostra e divide-se por 100 (obs. percentil: $n \gg 100$)

6.4.1. Se o valor encontrado for inteiro, o percentil é a média da observação que está nessa posição do ranking com o valor do próximo no ranking

6.4.2. Se o valor encontrado não for inteiro, arredonda-se para o superior mais próximo e procura-se a observação que está nessa posição no ranking.

$$n^{\circ} \text{ elementos} < \text{que o valor} = \frac{P \times n^{\circ} \text{ valores}}{100}$$

$$n^{\circ} \text{ elementos} < \text{que o valor} = \frac{90 \times 20}{100} = 18 \text{ elementos menores} = \frac{98+99}{2} = 98,5$$

6.5. Interpretação do Percentil

6.5.1. 10% renda menor que 10 mil

6.5.2. 20% renda menor que 17 mil

6.5.3. 50% renda menor que 42 mil

6.5.4. 80% renda menor que 83 mil

6.5.5. 90% renda menor que 116 mil

6.5.6. 95% renda menor que 150 mil

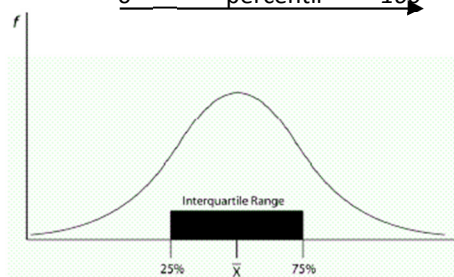
Table 2-1 U.S. Household Income for 2001	
Percentile	2001 Household Income
10th	\$ 10,913
20th	\$ 17,970
50th	\$ 42,228
80th	\$ 83,500
90th	\$ 116,105
95th	\$ 150,499

6.6. Cálculo do Percentil:

6.6.1. Percentil do valor = $100 \times \frac{n.\text{ordem}-1}{n-1}$

1 — n. de ordem — n

0 — percentil — 100



7. DECIL – QUARTIL – MEDIANA

7.1. Posições especiais de Quartil

7.2. Decil: 10% – 20% – 30% etc.

7.3. Quartil: 25% - 75%

7.4. Mediana: 50%

7.5. Cálculo do Quartil – regra prática

7.5.1. Calcular mediana – separar os valores inferiores e os superiores à mediana

7.5.2. Quartil 1 (25%): calcular a mediana dos valores inferiores

7.5.3. Quartil 3 (75%): calcular a mediana dos valores superiores

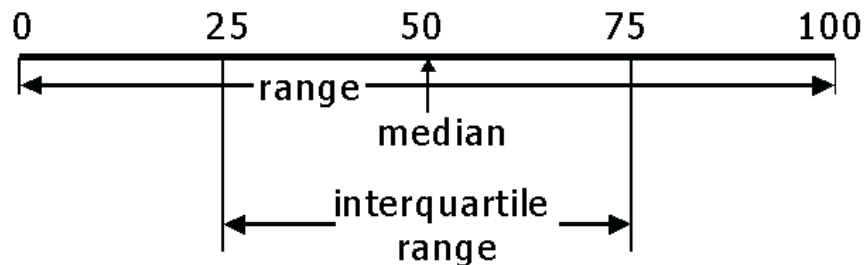
7. MEDIDAS DE POSIÇÃO, DISPERSÃO OU VARIABILIDADE E ASSIMETRIA

1. EXTENSÃO (Amplitude)

1.1. É a distância entre a menor e a maior observação: $A = x_{\max} - x_{\min}$

2. EXTENSÃO INTERQUARTIL

2.1. É a distância entre o quartil 25 e o quartil 75



3. DIAGRAMA DE CAIXA E HASTES (

3.1. Resume a Extensão, Extensão Interquartil e Mediana

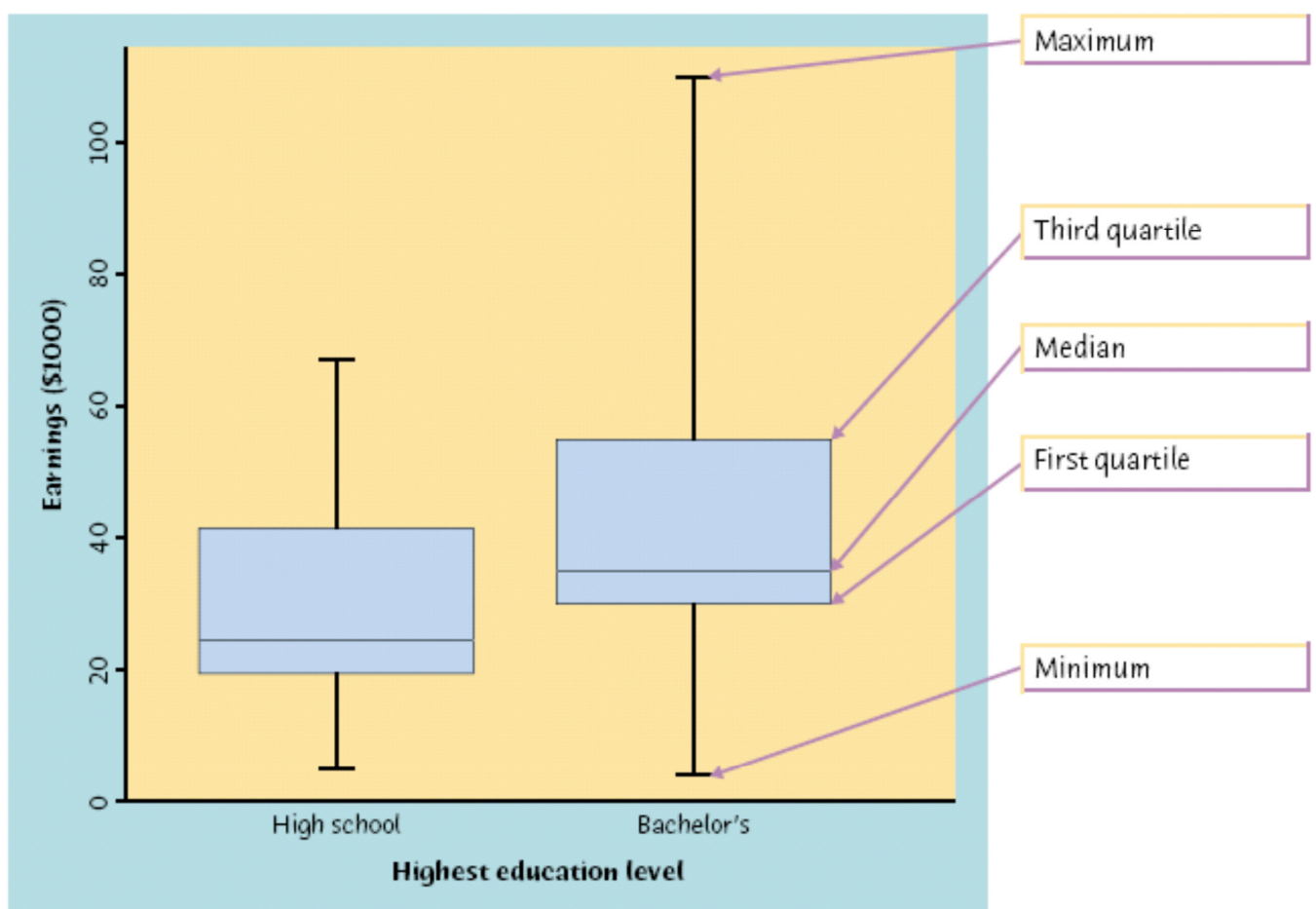


Figure 2.2 Side-by-side boxplots comparing the distributions of earnings for two levels of education.

4. VARIÂNCIA

4.1. Medida do espalhamento dos valores em torno da média



$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

σ^2 = variância

x_i = valor i

\bar{x} = valor médio

n = número elementos

4.2. Para amostras, utiliza-se o símbolo S^2 e $(n-1)$ no denominador

5. DESVIO PADRÃO

5.1. Medida do espalhamento dos valores em torno da média

População: $\sigma = \sqrt{\sigma^2}$

Amostras: $s = \sqrt{S^2}$

$$\sigma = \sqrt{\frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2}$$

5.2. Na prática utiliza-se o símbolo σ também para amostras

6. PADRONIZAÇÃO

6.1. Afastamento de um valor em relação à média em unidades de desvios-padrão

6.2. O valor padronizado = distância da média dividida pelo desvio padrão

$$x_p = \frac{x_i - \bar{x}}{\sigma}$$

σ : desvio padrão populacional ou amostral

7. DISTRIBUIÇÃO NORMAL (Curva Normal ou Gaussiana – em forma de sino)

7.1. Distribuição teórica desenvolvida por Abraham de Moivre em 1733

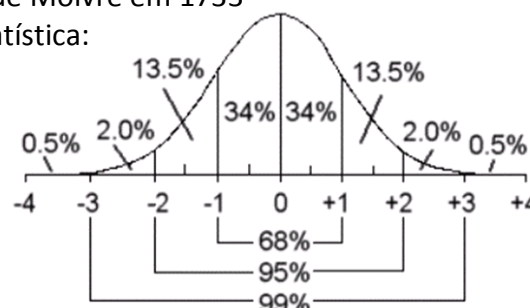
7.2. Apresenta propriedades importantes para a Estatística:

7.2.1. Média = 0 (zero)

7.2.2. Desvio Padrão: 1 (um)

7.2.3. Área total sob a curva = 1 (um)

7.2.4. Área parcial: densidade de probabilidade



8. TEOREMA DO LIMITE CENTRAL

8.1. A distribuição das médias amostrais tende a ser Normal quando o tamanho da amostra cresce, independentemente da distribuição da população;

8.1.1. A média da distribuição das médias amostrais é μ

8.1.2. O desvio padrão da distribuição das médias amostrais é $\frac{\sigma}{\sqrt{n}}$

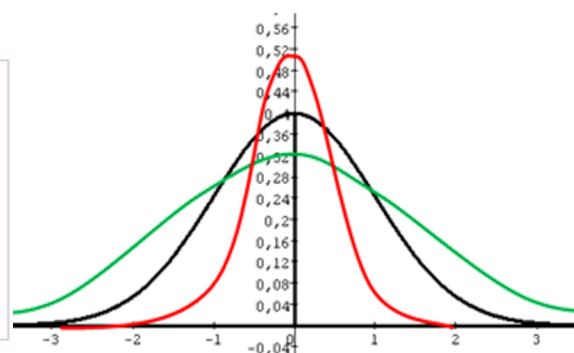
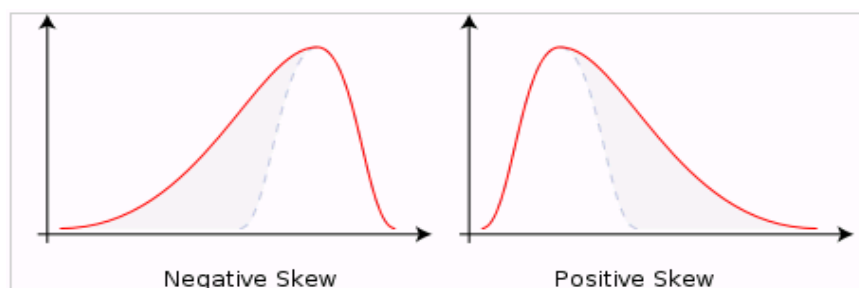
σ : desvio padrão

n : tamanho da amostra

9. ASSIMETRIA E CURTOSE

9.1. Assimetria (skewness) : Distribuição dos valores em torno da média

9.2. Curtose: Espalhamento dos valores em torno da média



8. NOÇÕES DE PROBABILIDADE

1. Fenômenos Determinísticos e Probabilísticos

- 1.1. Determinístico: seguem leis conhecidas, p/ex. leis físicas, químicas, etc.;
- 1.2. Probabilísticos: seguem leis da probabilidade, p/ex. acaso ao lançar um dado.

2. Definição Clássica de Probabilidade

- 2.1. $P(A)$ = número de resultados favoráveis / número de resultados possíveis;
- 2.2. Os eventos devem ser mutuamente exclusivos (se ocorre A não pode ocorrer B);
- 2.3. A probabilidade deve ser a mesma para cada um dos eventos;
- 2.4. A probabilidade sempre varia de zero a 1 (ou 0% a 100%);
 - 2.4.1. Obs: os extremos são “extremamente” improváveis: 0% e 100% são certezas;
- 2.5. Exemplos:
 - 2.5.1. Probabilidade de se tirar “Cara” em um lance de moeda = $1/2$;
 - 2.5.2. Probabilidade de se obter um “6” em um lance de dados = $1/6$.

3. Conceitos Básicos

- 3.1. Experimento Aleatório (E): as condições iniciais são sempre idênticas, mas os resultados finais são diferentes e não previsíveis;
 - 3.1.1. Exemplo: lance de dados, vida útil de uma bateria de celular, etc.
- 3.2. Frequência Relativa de um evento A: razão entre o número de vezes em que ocorreu A e o número total de eventos observados: $F(A) = nA / N$;
- 3.3. Espaço Amostral (S): conjunto de todos os resultados possíveis de um experimento;
 - 3.3.1. Exemplo: lance de dados $S = 1, 2, 3, 4, 5, 6$ – lance de moedas $S = \text{cara, coroa}$;
- 3.4. Evento: é um subconjunto do espaço amostral;
 - 3.4.1. Exemplo: números pares no lançamento de dados: $A = (2, 4, 6)$;
No lançamento de moeda 4 vezes, o número de caras possíveis é $(0, 1, 2, 3, 4)$;
Se nos interessa somente 2 caras, então o Evento $A = (2)$;
- 3.5. Eventos complementares: São todos os resultados do espaço amostral que não fazem parte do evento desejado: se p é a probabilidade de que um evento desejado ocorra, e q é a probabilidade de que não ocorra, então $p + q = 1$ e $q = 1 - p$;
- 3.6. Eventos Independentes: quando o resultado de um não influencia o resultado do outro, por exemplo, o lançamento de duas moedas, dois dados, retirada de uma bola em sorteio oficial não influencia a retirada da segunda bola (de um outro globo);
 - 3.6.1. A probabilidade de que eles ocorram simultaneamente é igual ao produto das probabilidades de cada um: $p = p_1 \times p_2$;
 - 3.6.2. Exemplo: a probabilidade de se tirar dois “6” no lançamento de dois dados é $1/6 \times 1/6 = 1/36$;
- 3.7. Eventos mutuamente exclusivos: a ocorrência de um deles exclui a ocorrência do outro (ou não tem elementos comuns);
 - 3.7.1. Ex: no lance de uma moeda, a ocorrência de Cara exclui a ocorrência de Coroa;
 - 3.7.2. A probabilidade de que um OU o outro ocorra é igual a soma das probabilidades: $p = p_1 + p_2$;
 - 3.7.3. Exemplo: ao lançar um dado, qual a probabilidade de se tirar um 1 ou um 6:
 $P(1) = 1/6$ e $P(2) = 1/6$, então a $P(1 \text{ ou } 2) = 1/6 + 1/6 = 1/3$;

4. Aplicações da Probabilidade (chance de ocorrência de um evento):

4.1. Tomada de decisão baseada em previsões (estimativas) sobre eventos futuros:

- 4.1.1. Estratégias de campanha política baseadas nas pesquisas de opinião;
- 4.1.2. Escolha de tratamento médico baseado em resultados experimentais;
- 4.1.3. Resistência do concreto ou aço baseado em ensaios de ruptura;
- 4.1.4. Levar (ou não) guarda-chuvas dependendo da chance de chuva.

4.2. Exemplo:

- 4.2.1. Quem compareceu à feira do livro neste ano?
 - a) Número de alunos que compareceram / número de alunos na classe;
 - b) Chance, de escolha ao acaso, de 1 aluno que compareceu: C / T ;
- 4.2.2. Número de alunos que foram ao cinema na última semana?
 - a) Número de alunos que foram / número de alunos na classe;
 - b) Chance, de escolha ao acaso, de 1 aluno que foi: F / T ;
- 4.2.3. Probabilidade de escolha ao acaso de quem foi à feira e ao cinema:
 - a) $(C/T) * (F/T)$
 - b) Verificar na classe

5. Terminologia:

5.1. Probabilidade: incerteza associada aos resultados de um experimento aleatório;

5.2. Resultado: um resultado particular de um experimento;

5.3. Espaço Amostral: todos os resultados possíveis do experimento;

5.3.1. Lançar uma moeda: $S = \{H, T\}$

5.4. Evento: uma combinação de resultados

5.4.1. Lançar uma moeda duas vezes: $S = \{HH, HT, TH, TT\}$

5.4.2. Lançar duas moedas: $S = \{HH, HT, TH, TT\}$

5.4.3. Lançar um dado: $S = \{1, 2, 3, 4, 5, 6\}$

5.4.4. Lançar dois dados: $S = \{11, 12, \dots, 21, 22, \dots, 65, 66\}$

6. Probabilidade:

6.1. Eventos igualmente prováveis: probabilidade de ocorrência de cada um é igual;

6.1.1. Jogar uma moeda e obter Cara: $1/2$;

6.1.2. Lançar um dado e obter 6: $1/6$;

6.2. Lei dos grandes números: a frequência observada de um evento tende (converge) à probabilidade P (calculada) quando o número de repetições cresce (tende a ∞);

6.3. Eventos não igualmente prováveis: dados viciados, moedas desbalanceadas, etc.

7. Eventos “E” e “OU”

7.1. Evento OU: quando o sucesso é definido pela obtenção do evento A ou do B:

7.1.1. Lançar um dado e obter um valor par (2 ou 4 ou 6): $1/2$

7.1.2. Jogar um dado e obter um número menor que 5: $4/6$

a) Soma da probabilidade dos eventos individuais (união dos eventos);

U

7.2. Evento E: quando o sucesso é definido pela obtenção do evento A e do B;

7.2.1. Jogar uma moeda duas vezes e obter Coroa e depois Cara: $1/4$

a) Produto da probabilidade dos eventos individuais (interseção dos eventos)

∩

7.3. Combinação do evento E com o evento OU

7.3.1. Jogar duas moedas e obter cara e coroa ou coroa e cara: $1/2$

8. Probabilidade Condicional

8.1. Probabilidade da ocorrência de um evento A dado que já ocorreu B:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

8.2. Exemplo: lançar um dado duas vezes:

8.2.1. O espaço amostral para cada lançamento é: $S = \{1, 2, 3, 4, 5, 6\}$

8.2.2. O evento A é definido como $S = \{2, 3\}$ (2/6)

8.2.3. O evento B é definido como $S = \{2, 4, 6\}$ (3/6)

8.2.4. Qual a probabilidade de ocorrer A dado que já ocorreu B?

a) $P(A) = 1/3$ $P(B) = 1/2$ $P(A \text{ e } B) = 1/6$ $P(A|B) = 1/3$

9. Exercícios

9.1. Uma determinada classe é composta por estudantes do sexo masculino e feminino.

Alguns estudantes tem cabelos curtos e outros cabelos longos. A simbologia dos eventos é indicada abaixo:

- a) F: evento em que o estudante é do sexo feminino
- b) M: evento em que o estudante é do sexo masculino
- c) C: evento em que o estudante tem cabelos curtos
- d) L: evento em que o estudante tem cabelos longos

9.1.1. Escreva os símbolos para a probabilidade de um estudante que:

- a) Não tenha cabelos longos
- b) Seja do sexo masculino ou tenha cabelos curtos
- c) Seja do sexo feminino e tenha cabelos longos
- d) Seja do sexo masculino, dado que tenha cabelos longos
- e) Tenha cabelos longos, dado que seja do sexo masculino
- f) Dentre os estudantes do sexo feminino, tenha cabelos curtos
- g) Dentre os estudantes com cabelos longos, seja do sexo feminino
- h) Seja do sexo feminino ou tenha cabelos longos
- i) Seja do sexo masculino, com cabelos curtos
- j) Seja do sexo feminino

9. DISTRIBUIÇÕES DE PROBABILIDADE

1. Distribuição de Variáveis Aleatórias Discretas

1.1. Variáveis discretas: resulta de contagem e tem seus valores limitados:

1.1.1. Lançamento de moedas (cara,coroa): binomial uniforme;

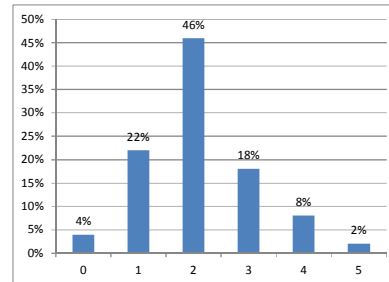
1.1.2. Lançamento de dados (1, 2, 3, 4, 5, 6): uniforme;

1.1.3. Número de golpes para enterrar uma estaca em um determinado terreno;

1.2. Exemplo: número de vezes que um recém-nascido acorda a mãe entre 24:00 e 06:00

1.2.1. Estuda-se 50 mães e observa-se os seguintes resultados:

x	$P(X = x)$
0	$P(X=0) = \frac{2}{50}$
1	$P(X=1) = \frac{11}{50}$
2	$P(X=2) = \frac{23}{50}$
3	$P(X=3) = \frac{9}{50}$
4	$P(X=4) = \frac{4}{50}$
5	$P(X=5) = \frac{1}{50}$

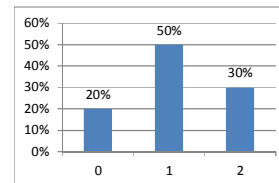


1.2.2. A Média dessa distribuição é calculada multiplicando-se o número de vezes que a mãe acorda pela probabilidade de ser acordada:

$$a) \mu = 0 \times 4\% + 1 \times 22\% + 2 \times 46\% + 3 \times 18\% + 4 \times 8\% + 5 \times 2\% = 2,1$$

1.3. Exercício: um time pode jogar 0, 1 ou 2 vezes por semana. As probabilidades de que cada evento desses ocorra é: 0x: 20%; 1x: 50%; 2x: 30%

x	$P(X=x)$
0	0.2
1	0.5
2	0.3



a) Em média, quantas vezes o time joga por semana?

$$N = 0 \times 20\% + 1 \times 50\% + 2 \times 30\% = 1,1 \text{ vezes por semana (média no longo prazo)}$$

2. Distribuição Binomial

2.1. Características:

2.1.1. Número fixo n de tentativas (repetições do experimento);

2.1.2. Apenas dois resultados possíveis em cada tentativa: sucesso e fracasso;

a) A letra p representa o sucesso e a letra q representa o fracasso;

2.1.3. As n tentativas são independentes e são repetidas em condições idênticas;

b) Independente significa que o resultado de uma tentativa não afeta o resultado de qualquer outra tentativa, isto é, a probabilidade de sucesso (p) e a probabilidade de fracasso (q) é a mesma para todas as tentativas;

2.2. Exemplos:

2.2.1. Ao jogar uma moeda, Cara representa sucesso, Coroa representa fracasso;

2.2.2. Para um determinado jogador, chute em Pênalti, o gol representa sucesso e não-gol representa fracasso – este jogador tem uma taxa de sucesso de 80% no longo prazo (obs: para o goleiro, sucesso seria a taxa de o “não-gol”);

2.2.3. Um determinado processo de fabricação apresenta taxa de 2% de produtos defeituosos (falhos);

2.2.4. Eventos em uma distribuição binomial são gerados por um processo de Bernoulli (sucesso-fracasso) e a distribuição binomial descreve o número de sucessos em n tentativas de um processo Bernoulli.

2.3. Distribuição de probabilidade binomial: resultados de um experimento binomial onde a variável aleatória X é o número de sucessos obtidos em n tentativas;

2.3.1. A média: $\mu = np$ a variância: $\sigma^2 = npq$ e o desvio padrão: $\sigma = \sqrt{npq}$

2.4. A fórmula para calcular a probabilidade de um número particular de sucessos em um número particular de tentativas é:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Onde $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ e $\binom{n}{k}$ é a combinação de n elementos k a k ;

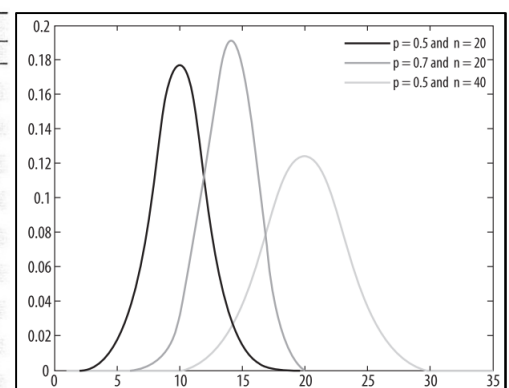
Sendo n o número de tentativas e k o número de sucessos nessas n tentativas;

2.4.1. Em suma, a fórmula binomial pode ser utilizada para calcular a probabilidade de se obter um número particular de sucessos, dada a probabilidade de sucessos por tentativa e um número fixo de tentativas.

2.5. Cada combinação de n (número de tentativas) e p (probabilidade de sucessos) resulta em uma distribuição de probabilidade diferente:

Tabela 2 Distribuição binomial: probabilidade de cada valor x em função de n e π (continuação)

n	x	π								
		0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95
1	0	0,4500	0,4000	0,3500	0,3000	0,2500	0,2000	0,1500	0,1000	0,0500
	1	0,5500	0,6000	0,6500	0,7000	0,7500	0,8000	0,8500	0,9000	0,9500
2	0	0,2025	0,1600	0,1225	0,0900	0,0625	0,0400	0,0225	0,0100	0,0025
	1	0,4950	0,4800	0,4550	0,4200	0,3750	0,3200	0,2550	0,1800	0,0950
	2	0,3025	0,3600	0,4225	0,4900	0,5625	0,6400	0,7225	0,8100	0,9025
3	0	0,0911	0,0640	0,0429	0,0270	0,0156	0,0080	0,0034	0,0010	0,0001
	1	0,3341	0,2880	0,2389	0,1890	0,1406	0,0960	0,0574	0,0270	0,0071
	2	0,4084	0,4320	0,4436	0,4410	0,4219	0,3840	0,3251	0,2430	0,1354
	3	0,1664	0,2160	0,2746	0,3430	0,4219	0,5120	0,6141	0,7290	0,8574
4	0	0,0410	0,0256	0,0150	0,0081	0,0039	0,0016	0,0005	0,0001	0,0000
	1	0,2005	0,1536	0,1115	0,0756	0,0469	0,0256	0,0115	0,0036	0,0005
	2	0,3675	0,3456	0,3105	0,2646	0,2109	0,1536	0,0975	0,0486	0,0135
	3	0,2995	0,3456	0,3845	0,4116	0,4219	0,4096	0,3685	0,2916	0,1715
	4	0,0915	0,1296	0,1785	0,2401	0,3164	0,4096	0,5220	0,6561	0,8145



2.6. Observações:

2.6.1. A probabilidade é calculada pela integral sob a curva até o ponto desejado;

2.6.2. Esse cálculo é resumido em tabelas ou efetuado em softwares específicos;

2.6.3. De maneira geral, pode-se assumir que quando o número de tentativas aumenta, a distribuição binomial assemelha-se à distribuição normal;

A aproximação da distribuição binomial pela normal é considerada boa quando: $np(1-p) \geq 3$

2.7. Exemplo: Ao se jogar uma moeda cinco vezes, qual a probabilidade de se obter exatamente apenas uma Cara? $p=0,5$ $k=5$ $n=1$

2.7.1. $P(n=1) = \binom{5}{1} 0,5^1 (1-0,5)^{5-1} = 0,156$

2.8. Exemplos práticos:

2.8.1. Uma máquina muito usada produz 1% de peças defeituosas. Qual a probabilidade de produto defeituoso nas próximas 25 peças?

2.8.2. Em um sistema de transmissão, 10% dos pacotes são recebidos com erro. Qual seria a taxa de erros na transmissão dos próximos 7 pacotes?

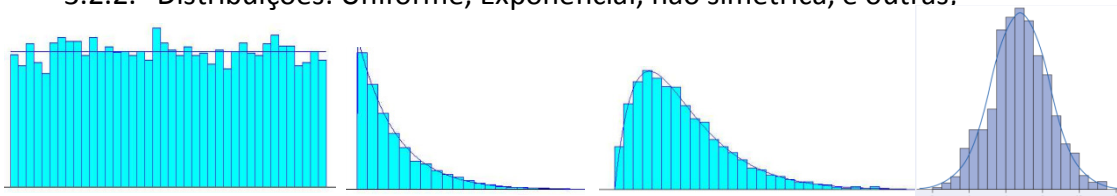
3. Funções de Probabilidade Contínuas – A Curva Normal Padrão (Gauss)

3.1. Variáveis contínuas: resulta de medição e tem seus valores ilimitados:

- 3.1.1. Resultados dos testes de ruptura de corpos de prova de concreto;
- 3.1.2. Resultados dos testes de tração de barras de aço;
- 3.1.3. Resultados do teste de inglês para os alunos da universidade;

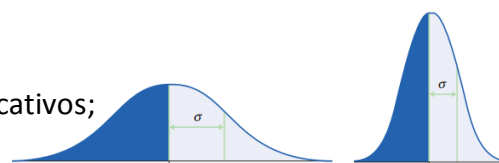
3.2. Curvas de Densidade

- 3.2.1. Histograma com classes tendendo a zero;
- 3.2.2. Distribuições: Uniforme, Exponencial, não simétrica, e outras;



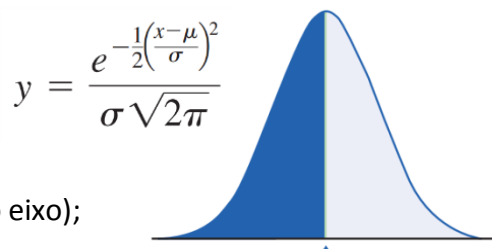
3.3. Distribuição Normal :

- 3.3.1. Simétrica, forma aproximada de “sino”;
- 3.3.2. Média, Mediana e Moda são próximas;
- 3.3.3. Não há dados discrepantes “outliers” significativos;
- 3.3.4. Não há “vazios” significativos;
- 3.3.5. Constituem “famílias” de curvas normais: médias iguais, desvios padrão diferentes;



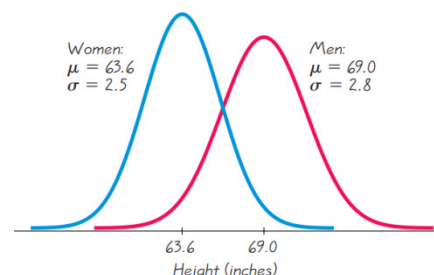
3.4. Curva Normal Padrão:

- 3.4.1. Curva precisamente simétrica;
- 3.4.2. Área da curva é igual a 1 (ou 100%);
- 3.4.3. Média “ μ ” é igual a zero;
- 3.4.4. Desvio Padrão “ σ ” é igual a 1;
- 3.4.5. Média, Mediana e Moda são idênticas;
- 3.4.6. Extremos tendem ao infinito (não tocam o eixo);



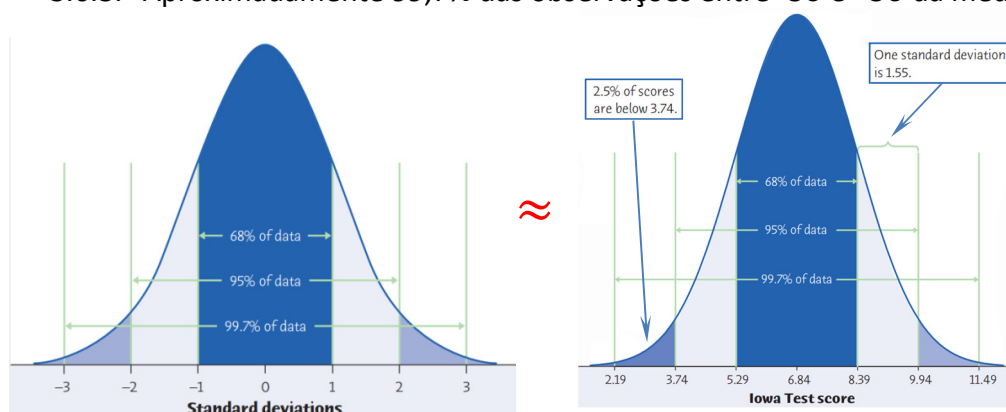
3.5. Importância da distribuição Normal:

- 3.5.1. Se aproximam muito de distribuições de dados reais, como tamanho de espigas de milho, diâmetro ou peso de frutas e legumes, dimensões de brita, resistência à compressão do concreto, resistência à tração do aço, etc.
- 3.5.2. São boas aproximações de diversos tipos de resultados aleatórios (em número elevado de tentativas) como lançamento de moedas;
- 3.5.3. Muito utilizada em testes estatísticos, por exemplo, comparação de médias;



3.6. Propriedades da curva Normal Padrão:

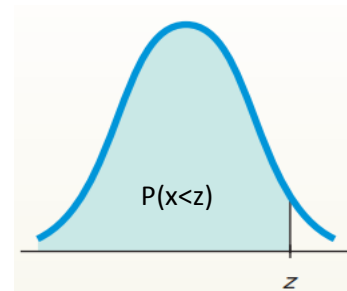
- 3.6.1. Aproximadamente 68% das observações entre -1σ e $+1\sigma$ da média μ ;
- 3.6.2. Aproximadamente 95% das observações entre -2σ and $+2\sigma$ da média μ ;
- 3.6.3. Aproximadamente 99,7% das observações entre -3σ e $+3\sigma$ da média μ .



3.7. Padronização: transformação do valor da observação em seu valor padronizado (escore z):

3.7.1. Integral sob a curva, até o valor z , representa a probabilidade de se encontrar um valor menor ou igual a z .

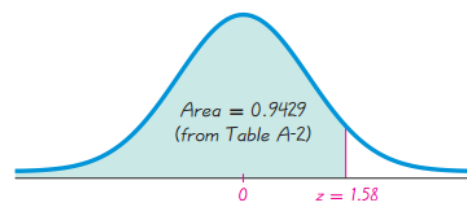
$$z = \frac{x - \mu}{\sigma}$$



3.8. Exemplos: Um teste com uma grande amostra de termômetros de precisão mostra que a temperatura média de congelamento da água é 0°C , com desvio padrão igual a 1°C (os termômetros não são tão precisos assim). Qual é a probabilidade de se escolher um termômetro ao acaso e a leitura ser menor que $1,58^{\circ}\text{C}$? Assumir que as leituras são Normalmente distribuídas.

3.8.1. A probabilidade da leitura ser menor que $1,58^{\circ}\text{C}$ é dada pela área sob a curva até o valor padronizado de 1,58: $(1,58-0)/1 = 1,58$

a) Consultando esse valor na tabela Z encontramos o valor 0,9429 (94,29%).



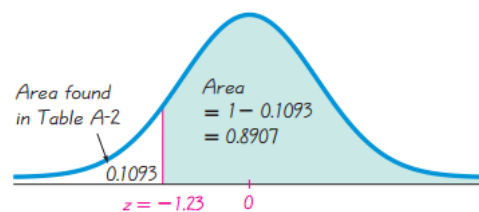
3.9. No caso do termômetro acima, qual é a probabilidade de, selecionando-se um termômetro ao acaso, encontrar leitura maior que $-1,23^{\circ}\text{C}$?

3.9.1. Calcula-se o valor padronizado de -1,23: $(-1,23-0)/1 = -1,23$

a) Consultando esse valor na tabela Z encontramos o valor 0,1093 (10,93%);

b) Esse valor indica a probabilidade de se encontrar uma temperatura menor que $-1,23^{\circ}\text{C}$.

c) Como a área total sob a curva é igual a 1, a probabilidade de se encontrar uma temperatura maior que -1,12 é dada por $(1-0,1093)$, ou seja, 0,8907 (89,07%).

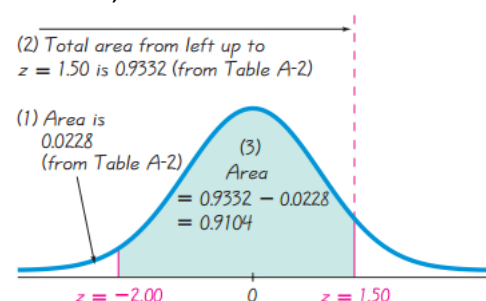


3.10. No caso do termômetro acima, qual é a probabilidade de, selecionando-se um termômetro ao acaso, encontrar leitura entre $-2,00^{\circ}\text{C}$ e $1,50^{\circ}\text{C}$?

3.10.1. Calcula-se o valor padronizado de -2,00 e de 1,50: $(-2,00 \text{ e } +1,50)$;

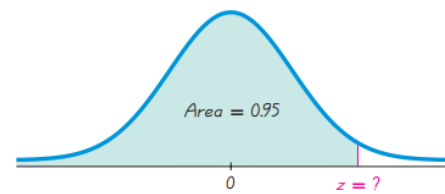
a) Consultando-se esses valores na tabela Z, encontramos os valores 0,0228 e 0,9332.

b) A área sob a curva entre os dois valores padronizados é dada por $0,9332 - 0,0228 = 0,9104$ ou 91,04%



4. Escore z de áreas (probabilidades) conhecidas

- 4.1. Supondo o mesmo termômetro do item anterior, qual é a temperatura correspondente ao P_{95} (percentil 95), isto é, separa os 95% termômetros com temperatura inferior e os 5% termômetros com temperatura superior?



- a) Consultando a tabela z encontramos o valor para área de 0,9495 e 0,9505. Como o valor 0,95 (95%) é muito utilizado, o valor z correspondente é muito conhecido e igual a 1,645;
b) Substituindo-se na equação de determinação do valor z, encontramos a temperatura de 1,645°C.

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

$$x = z\sigma + \bar{x}$$

- 4.2. Continuando com o mesmo termômetro, quais as temperaturas que separam os 2,5% inferiores e os 2,5% superiores?



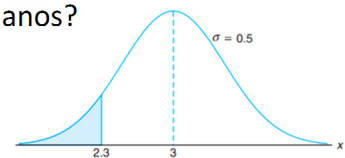
- a) Consultando a tabela z encontramos o valor para área de 0,025: -1,96 e para a área de 0,975 : +1,96.
b) Substituindo-se na equação, obtemos as temperaturas de -1,96°C e + 1,96°C.

5. Exercícios

- 5.1. Um determinado tipo de bateria dura, em média, 3,0 anos, com um desvio padrão de 0,5 anos. Assumindo que a vida da bateria é normalmente distribuída, qual é a probabilidade de uma dada bateria dure menos que 2,3 anos?

$$Z = (2,3 - 3) / 0,5 = -1,4$$

$$P(X < 2,3) = P(Z < -1,4) = 0,0808 = 8,08\%$$



- 5.2. Qual é a probabilidade dessa mesma bateria durar mais que quatro anos?

$$z = (4 - 3) / 0,5 = 2,00$$

$$P(X > 4,0) = 1 - P(Z < 2,0) = 1 - 0,9772 = 0,0228 = 2,28\%$$

- 5.3. As chuvas de verão em uma determinada localidade seguem aproximadamente uma distribuição normal com média de 852 mm e desvio padrão 82mm de chuva.

- 5.3.1. No verão de 2011, foi registrado 697mm de chuva. Qual a porcentagem de verões em que a chuva é de 697 ou menor?

$$z = (697 - 852) / 82 = -1,8902$$

$$P(X < 697) = P(Z < -1,892) = 0,0294 = 2,94\%$$

- 5.3.2. Chuva normal nessa localidade significa registro entre 683mm e 1022 mm. Qual a porcentagem de chuvas de verão que são consideradas normais ali?

$$z = (683 - 852) / 82 = -2,0610 \quad z = (1022 - 852) / 82 = 2,0731$$

$$P(z < -2,0610) = 0,0197 \quad P(z < 2,0731) = 0,9808$$

$$P(-2,0610 < z < 2,0731) = 0,9808 - 0,0197 = 0,9611 = 96,11\%$$

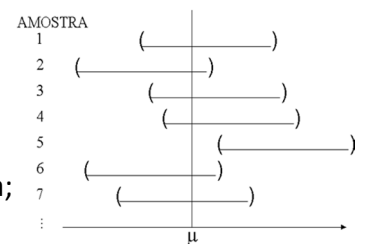
10. INTERVALOS DE CONFIANÇA

1. Conceito de Intervalo de Confiança

- 1.1. A Estatística fornece ferramentas para se estimar um parâmetro da população utilizando-se dados de uma amostra (estatísticas).
 - 1.1.1. Mas se os testes são repetidos com diferentes amostras, os resultados (as estatísticas) serão diferentes para cada uma delas;
 - 1.1.2. A Margem de Erro é uma medida de variabilidade que busca representar a variação dos resultados de diferentes amostras;
 - 1.1.3. A média somada e subtraída do valor da Margem de Erro fornece um intervalo onde o parâmetro da população deveria estar em X% das amostras;
 - 1.1.4. Esse X% é o grau de confiança que desejamos no resultado;
- 1.2. Um Intervalo de Confiança de 95% para a média amostral – representa um intervalo onde, se repetíssemos o experimento 100 vezes, em 95 vezes a média amostral deveria estar localizada dentro desse intervalo, e em 5% estaria localizada fora desse intervalo;
- 1.3. O Intervalo de confiança mede a variação dos resultados das amostras em virtude do acaso na seleção da amostra;

2. Para estimar um parâmetro com um intervalo de confiança:

- 2.1. Escolha o nível de confiança desejado e o tamanho da amostra;
- 2.2. Selecione uma amostra aleatória de indivíduos da população;
- 2.3. Colete dados confiáveis e relevantes de cada indivíduo da amostra;
- 2.4. Calcule as estatísticas desejada (por exemplo, média);
- 2.5. Calcule a margem de erro (item 3);
- 2.6. Relate a média \pm a margem de erro como a estimativa final do parâmetro
 - 2.6.1. O intervalo de confiança é o intervalo média-erro e média+erro



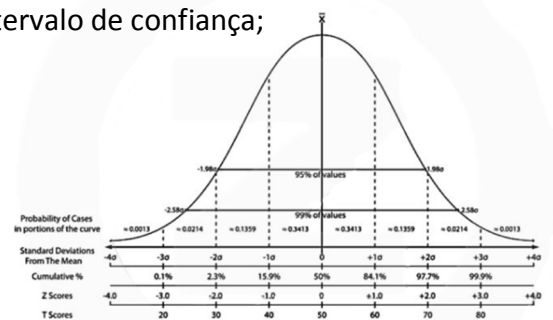
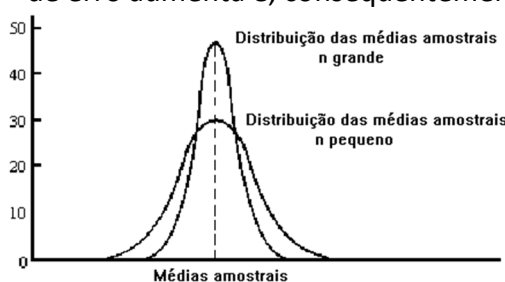
3. Cálculo do Intervalo de Confiança

- 3.1. A fórmula do Intervalo de Confiança para a média é:

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

\bar{x} : média da amostra
 z : valor do Z correspondente ao grau de confiança desejado
 σ : desvio padrão populacional
 n : número de elementos da amostra

- 3.2. A meta é sempre obter a menor margem de erro, resultando em um intervalo de confiança menor e, conseqüentemente, um resultado mais preciso;
- 3.3. Fatores que afetam o tamanho da margem de erro, (e conseqüentemente do intervalo de confiança):
 - 3.3.1. O nível de confiança: quando o nível de confiança aumenta, O valor Z aumenta, resultando em um intervalo de confiança maior
 - 3.3.2. O tamanho da amostra: quanto maior a amostra, menor a margem de erro e, conseqüentemente, o intervalo de confiança;
 - 3.3.3. A variabilidade na população: quando a variabilidade aumenta (σ), a margem de erro aumenta e, conseqüentemente, o intervalo de confiança;



- 3.4. Exemplo: estimar a percentagem de veículos de carga que trafegam em uma rodovia entre 00:00hs e 06:00 horas (madrugada), com 95% de confiança. Supondo que essa percentagem é 50% mais ou menos 4%, o intervalo de confiança de 95% é entre 46% e 54%.

4. Escolhendo um Nível de Confiança

- 4.1. A variabilidade em uma estatística da amostra é medida em Erros-Padrão, que é uma medida de espalhamento similar ao Desvio Padrão;

- 4.1.1. O Erro Padrão mede a variação entre os possíveis valores de uma estatística (por exemplo, da média da amostra);

- 4.1.2. O Desvio Padrão mede a variação entre todos os valores do parâmetro da população ou amostra (por exemplo, da altura);

- 4.2. O Nível de Confiança corresponde à percentagem de vezes que o resultado amostral estará correto se você repetir o teste com muitas amostras;

- 4.3. Os níveis típicos de Nível de Confiança são 80%, 90%, 95%, 98%, 99%;

- 4.4. Quando as condições apropriadas são encontradas ($N > 30$, distribuição aproximadamente normal), o número de erros-padrão a ser subtraído e acrescentado à média é dado pela distribuição Z (padrão normal: Z^*);

- 4.4.1. Quanto maior o nível de confiança, mais erros-padrão são acrescentados e subtraídos para estabelecer o intervalo de confiança.

- 4.4.2. O valor de z^* é obtido na tabela da distribuição Z. Os valores típicos são:

$$Ep = \frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Table 7-1 z^* -values for Selected (Percentage) Confidence Levels

Percentage Confidence	z^* -value
80	1.28
90	1.64
95	1.96
98	2.33
99	2.58

Obs:
 Z^* ou $Z_{\alpha/2}$

5. Nível de Confiança e o valor alfa (α)

- 5.1. O nível de confiança representa a probabilidade de se tomar a decisão correta quando a Hipótese Nula é verdadeira;

- 5.2. O nível de confiança pode ser expresso em termos do valor alfa: $1 - \alpha$

- 5.2.1. Alfa representa a percentagem de intervalos de confiança que são incorretos, isto é, não contém o parâmetro da população devido à amostragem ao acaso;

- 5.2.2. O valor alfa representa a chance de se cometer um erro Tipo I:

A hipótese nula é verdadeira e foi rejeitada:

TABELA DE ERROS		REALIDADE	
		H_0 verdadeira não culpado	H_0 falsa culpado
DECISÃO	Aceitar H_0 absolver	Decisão correta ($1 - \alpha$)	Erro tipo II (β)
	Rejeitar H_0 condenar	Erro tipo I (α)	Decisão Correta ($1 - \beta$)

6. Margem de Erro e Tamanho da Amostra

6.1. A Margem de Erro é dada pela equação:

(não confundir com Erro Padrão);

6.1.1. Quanto maior o nível de confiança (**z**) maior a margem de erro;

6.1.2. Quanto maior o desvio padrão (σ) maior a margem de erro;

6.1.3. Quando menor o tamanho da amostra (**n**) maior a margem de erro

6.2. O tamanho da amostra para se obter uma estatística com uma margem de erro determinada é dado pela equação (arredondar para cima):

6.2.1. É necessário conhecer o desvio padrão da população;

6.2.2. Se este não é conhecido, é necessário realizar um

Estudo Piloto e utilizar o desvio padrão da amostra substituindo o da população e substituir o valor **z** (distribuição normal) pelo valor **t** (distribuição *t* de *student*) com **n-1** graus de liberdade, onde **n** é o tamanho da amostra utilizada;

6.3. Para estimar o tamanho da amostra em proporções, utiliza-se $1 / \sqrt{n}$;

$$ME = z * \frac{\sigma}{\sqrt{n}}$$

$$n = \left[\frac{Z_{\alpha/2} * \sigma}{ME} \right]^2$$

7. Intervalo de Confiança para a Média da População

7.1. A média é uma medida de tendência central para dados quantitativos:

7.1.1. Renda, QI, altura, quantidade, peso, etc.

7.2. A média da população é estimada utilizando-se a média de uma amostra mais ou menos a margem de erro.

7.2.1. Esse intervalo é o intervalo de confiança para a média da população;

7.3. Exemplo: estimar com 95% de confiança a média do comprimento dos peixes na piracema do rio X:

7.3.1. O desvio padrão da população é conhecido e igual a 5,8 cm;

7.3.2. O intervalo de confiança de 95% resulta em valor **z*** igual a 1,96;

7.3.3. Tomando-se uma amostra de 100 peixes obtém-se uma média de 17,3 cm;

7.3.4. A margem de erro é encontrada utilizando-se a equação:

$$ME = z * \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad ME = 1,96 \frac{5,8}{\sqrt{100}} \quad \Rightarrow \quad ME = 1,14$$

7.3.5. O intervalo de 95% de confiança é : $17,3 \pm 1,14$ ou 16,16 a 18,44

O intervalo de valores prováveis para a média de comprimento dos peixes é entre 16,16 e 18,55 cm, baseado na amostra utilizada, com nível de confiança de 95%.

7.3.6. Se a amostra é pequena (**n** menor que 30) deve-se utilizar o valor **t** da distribuição *t* de *student*, com **n-1** graus de liberdade, e não o valor **z***.

8. Dados Emparelhados

8.1. Esse mesmo processo de cálculo pode ser utilizado para estimar a diferença média de dados emparelhados de uma população;

8.1.1. Dados emparelhados: análise de efeito de uma determinada droga sobre a redução da pressão arterial – mede-se a pressão antes e depois da aplicação da droga em um mesmo paciente e calcula-se a média da redução da pressão.

8.1.2. Essas diferenças representam uma única amostra de uma única população, assim o intervalo de confiança para a média de uma população pode ser utilizado para se estimar a diferença média na pressão sanguínea devido à utilização da droga específica.

9. Intervalo de Confiança para Proporção de uma População

9.1. Quando a característica de interesse é categórica (p/ex.: usa ou não cinto de segurança ao dirigir): o resultado é uma porcentagem ou uma proporção;

$$\hat{p} \pm z * \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

9.2. Exemplo: estimar a porcentagem de tempo que se encontra um farol vermelho em um determinado cruzamento de vias, com um intervalo de confiança de 95%:

9.2.1. Valor z^* para 95% = 1,96

9.2.2. Toma-se 100 diferentes amostras aleatórias nesse cruzamento ($n=100$);

➤ Encontra-se o farol vermelho em 53% delas

9.2.3. O intervalo de confiança é calculado segundo a fórmula acima:

$0,53 \pm 1,96 \sqrt{\frac{0,53(1 - 0,53)}{100}}$	$0,53 \pm 1,96 * 0,0499$
$53\% \pm 9,8\%$	

10. Intervalo de Confiança para Diferença entre Duas Médias

10.1. Quando a característica de interesse é numérica (p/ex.: altura, peso, renda);

$$(\bar{x}_1 - \bar{x}_2) \pm z * \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

10.2. Exemplo: estimar, com intervalo de confiança de 95%, a diferença entre as médias do PH de resíduos de dois processos diferentes. Toma-se 100 amostras do resíduo 1 e 110 amostras do resíduo 2. O resíduo 1 apresentou média de 8,5 e desvio padrão de 2,3. O resíduo 2 apresentou média de 7,5 com desvio padrão de 2,8.

$n_1 = 100$ $\bar{x}_1 = 8,5$ $\sigma_1 = 2,3$	$n_2 = 110$ $\bar{x}_2 = 7,5$ $\sigma_2 = 2,8$	$(8,5 - 7,5) \pm 1,96 \sqrt{\frac{2,3^2}{100} + \frac{2,8^2}{110}}$	$1,0 \pm 0,69$
--	--	---	----------------

1.1.1. O intervalo de confiança de 95% para a diferença entre as médias do PH é de 0,31 a 1,69. Conclui-se que o PH do resíduo 1 é, em média e com grau de confiança de 95%, maior que o PH do resíduo 2.

✓ Obs: quando o número de amostras é menor que 30, utiliza-se a distribuição *t de student* com n_1+n_2-2 graus de liberdade.

11. Intervalo de Confiança para duas Proporções

11.1. Quando se compara as proporções da variável de interesse em duas populações diferentes (p/ex.: proporção de fumantes entre homens e mulheres);

$$(\hat{p}_1 - \hat{p}_2) \pm z * \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

11.2. Exemplo: estimar, com grau de confiança de 95%, a diferença entre a proporção de homens e mulheres que frequentam cinema mais de 1 vez por mês;

* atenção ao plano de amostragem *

- 11.2.1. Foram entrevistados 100 mulheres e 100 homens aleatoriamente em um determinado shopping center (amostragem???). Constatou-se que 53 mulheres e 34 homens afirmaram que frequentam cinema mais de 1 vez por mês. O intervalo de 95% de confiança é calculado abaixo:

$$(0,53 - 0,34) \pm 1,96 \sqrt{\frac{0,53(1 - 0,53)}{100} + \frac{0,34(1 - 0,34)}{100}}$$

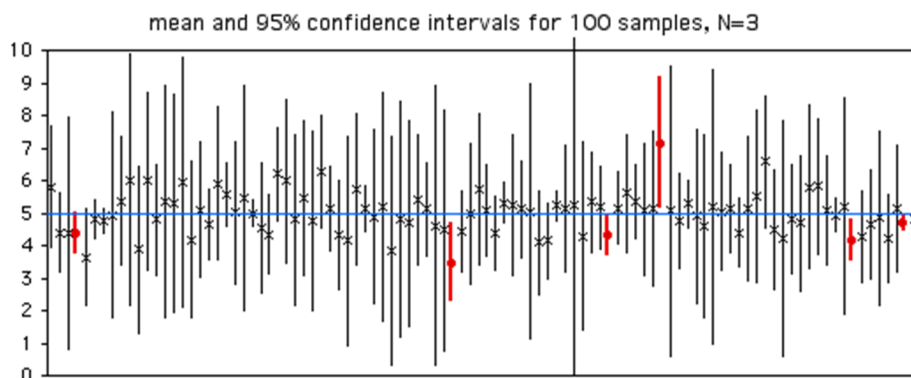
$$0,19 \pm 1,96 \sqrt{0,002491 + 0,002244}$$

$$0,19 \pm 0,1348 \quad 19\% \pm 13,5 \quad 5,5 \text{ a } 32,5$$

12. Interpretação do Intervalo de Confiança

- 12.1. O intervalo de confiança apresenta uma gama de valores prováveis para o parâmetro da população, baseado em uma amostra aleatória, com um certo grau de confiança.
- 12.2. Exemplo: uma pesquisa revela que 52% dos eleitores da cidade aprovam a gestão do atual prefeito, com uma margem de erro de 3% com grau de confiança de 95%.
- 12.2.1. Não é válido afirmar que a maioria dos eleitores aprovam a gestão, porque o intervalo de confiança de 95% varia de 49% a 55%.
- 12.2.2. É incorreto afirmar que “estamos 95% confiantes de que a média de aprovação está entre 49% e 55%.
- 12.2.3. A interpretação precisa do intervalo de confiança é a seguinte: nível de confiança é a porcentagem de todas as possíveis amostras de tamanho n cujos intervalos de confiança contém o parâmetro da população, isto é, em 95% dos casos o parâmetro da população estará dentro do respectivo intervalo de confiança.
- 12.2.4. O grau de confiança não se aplica a um único intervalo de confiança. Se o experimento for repetido 100 vezes, em 95 vezes o intervalo de confiança gerado conterá o parâmetro da população e em 5 vezes o intervalo de confiança gerado não conterá o parâmetro da população devido ao erro amostral.
- 12.3. Uma maneira correta de expressar o intervalo de confiança é “a faixa de valores prováveis para a média da população é de X a Y com um grau de confiança de Z%;
- 12.4. Uma maneira alternativa é “podemos dizer com Z% de confiança que a margem de erro é de mais ou menos W pontos absolutos (ou w% pontos percentuais).

O Intervalo de Confiança refere-se ao processo de amostragem, e não aos resultados de uma única amostra.



11. TESTE DE HIPÓTESES PARA UMA POPULAÇÃO

1. Conceitos Básicos

1.1. É um modo de confirmar ou negar uma afirmação sobre uma população utilizando dados de uma amostra;

1.1.1. Exemplo 1: 65% da população é a favor da proibição de fumar em lugares públicos: pesquisa-se uma amostra e compara-se com a afirmação acima;

1.1.2. Exemplo 2: pesquisa-se uma amostra e afirma-se que o resultado é válido para a população, com uma certa margem de erro (pesquisas eleitorais);

2. Estabelecendo as Hipóteses:

2.1. Estabelecer hipótese que será testada (p/exemplo: médias iguais ou diferentes?);

2.1.1. Todo teste de hipóteses tem duas hipóteses: H_0 e H_1 ;

2.2. Hipótese nula (H_0): a diferença entre os valores **não é** estatisticamente significativa;

2.2.1. Afirma que nada de diferente está ocorrendo com a amostra (igual a população);

2.2.2. “O réu é inocente até que se prove o contrário”;

2.3. Hipótese alternativa (H_1): a diferença entre os valores **é** estatisticamente significativa;

2.3.1. O valor da amostra é diferente do valor da população;

2.3.2. O valor da amostra é maior que o valor da população;

2.3.3. O valor da amostra é menor que o valor da população

2.4. O tipo de diferença é escolhido em função do que se deseja verificar:

2.4.1. O tempo de espera em um ponto de ônibus é diferente de 10 minutos;

2.4.2. O tempo de espera em um ponto de ônibus é menor que 10 minutos;

2.4.3. O tempo de espera em um ponto de ônibus é maior que 10 minutos;

2.5. IMPORTANTE: Se a hipótese nula for descartada, pode-se afirmar apenas que

2.5.1. “a diferença encontrada é estatisticamente significativa”,
não se pode afirmar que “os valores são diferentes”;

2.5.2. Exemplo: “*estatisticamente falando*”, afirma-se que a diferença entre porcentagem de fumantes em duas cidades é estatisticamente significativa, e não que o número de fumantes nessas cidades é diferente (pesquisa amostral);

2.5.3. O “júri” decide se HÁ ou NÃO HÁ evidências para o acusado ser considerado culpado (ou não culpado), mas nunca pode afirmar que ele é inocente!

2.6. Chance de tomar uma decisão errada: alfa (α)

2.6.1. Em geral, aceita-se que em 5% (0,05) das vezes tomamos a decisão errada;

3. Tipos de Erros (decisão errada)

TABELA DE ERROS		REALIDADE	
		H_0 verdadeira não culpado	H_0 falsa culpado
DECISÃO	Aceitar H_0 absolver	Decisão correta ($1 - \alpha$)	Erro tipo II (β)
	Rejeitar H_0 condenar	Erro tipo I (α)	Decisão Correta ($1 - \beta$)

Hipótese Nula: não há evidências para o réu ser considerado culpado
(não afirma que ele é INOCENTE)

4. Padronização de Valores:

4.1. Valor Padronizado: subtrai-se do valor a média e divide-se pelo desvio padrão;

4.1.1. Indica quantas unidades de desvio padrão esse valor afasta-se da média;

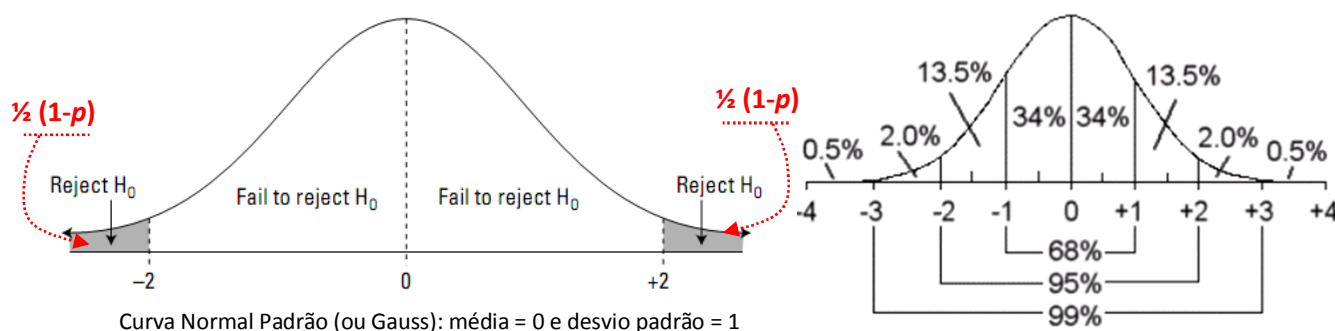
$$\text{Valor Padronizado} = \frac{\text{Valor} - \text{Média}}{\text{Desvio Padrão}}$$

4.2. Desvio Padrão é uma medida de espalhamento dos valores em torno da média:

$$\sigma = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}}$$

5. Avaliando as Evidências

5.1. Localizar o valor da média populacional em relação à distribuição amostral:



➡ 5.2. Se a média amostral padronizada está próxima da média populacional, não se rejeita a hipótese nula (as diferenças não são estatisticamente significativas);

➡ 5.3. Se a média amostral está afastada da média populacional, rejeita-se a hipótese nula (as diferenças são estatisticamente significativas);

5.4. A distância limite para rejeitar ou não rejeitar a hipótese nula determina o nível de confiança do teste: α (alfa), em geral igual a 5% (ou 1,96 desvios-padrão);

5.4.1. “Rigorosamente estatisticamente falando”, nunca se “aceita” a Hipótese nula;

5.5. A área sob a curva determina o valor de alfa;

5.5.1. A área total sob a curva padrão é igual a 1 (ou 100%)

6. Encontrando o valor p

6.1. O valor p indica a probabilidade da ocorrência do valor encontrado se a hipótese nula fosse verdadeira;

6.1.1. Quanto mais afastado da média for o valor, menor será o p ;

6.2. Localiza-se o valor encontrado na tabela de distribuição normal;

6.3. Lê-se a porcentagem (em forma decimal) de chance de se estar além desse valor;

6.3.1. Se a hipótese for “menor que”, a tabela apresenta o valor diretamente;

6.3.2. Se a hipótese for “maior que”, subtraia esse valor de 1;

6.3.3. Se a hipótese for “diferente de”, duplique esse valor;

6.4. Transforma-se esse valor decimal em porcentagem, multiplicando-se por 100.

p

$1-p$

$\frac{1}{2}(1-p)$

7. Interpretando o valor p

7.1. Define-se a probabilidade do erro tipo I que se está disposto a aceitar: alfa (α);

7.1.1. Erro I: Rejeitar a Hipótese Nula quando ela é verdadeira;

7.1.2. Em geral, assume-se valor alfa de 0,05 (5%), 0,01 (1%), 0,001 (0,1%);

7.2. Se o valor p é maior que alfa (α), falha-se em rejeitar a hipótese nula (H_0);

7.3. Se o valor p é menor que alfa (α), rejeita-se a hipótese nula (H_0);

8. Significância Estatística

- 8.1. Se o valor p é menor que 0,01 o resultado é altamente estatisticamente significativo;
- 8.2. Se o valor p é entre 0,05 e 0,01 o resultado é estatisticamente significativo;
- 8.3. Se o valor p é muito próximo ou igual a 0,05, o resultado é marginalmente estatisticamente significativo e a decisão pode ser aceitar ou rejeitar a hipótese nula;
- 8.4. Se o valor p é maior que 0,05 o resultado é estatisticamente não significativo e não se rejeita a hipótese nula.
- 8.5. Obs: Quando se relata que um resultado é estatisticamente significativo, é importante citar o valor p para se ter a dimensão dessa significância.

9. Passos gerais para um teste de hipótese

- 9.1. Estabeleça a hipótese nula e a hipótese alternativa (H_0 e H_a);
- 9.2. Escolha uma amostra aleatória da população e calcule as estatísticas da amostra:
 - 9.2.1. Média e desvio padrão;
- 9.3. Converta a estatística da amostra para estatística de teste: $Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$; onde:
 - 9.3.1. \bar{x} é a média da amostra e μ_0 é a média da população
 σ é o desvio padrão e n é o número de elementos na amostra
onde o erro padrão da média é dado pela fórmula σ / \sqrt{n}
- 9.4. Encontre o valor p para sua estatística de teste;
- 9.5. Avalie o valor p (significância estatística) e tome sua decisão;

10. Testando a média de uma população (exemplo):

- 10.1. Afirma-se que mulheres que trabalham passam apenas, em média, 11 minutos por dia conversando com seus filhos. Verifique se essa afirmação é verdadeira.
 - 10.1.1. A variável Tempo de conversa, é Quantitativa, Razão, Contínua;
- 10.2. A hipótese nula: a média de tempo de conversa é igual a 11 minutos;
- 10.3. A hipótese alternativa: a média de tempo de conversa é maior que 11 minutos;
- 10.4. O valor Z é calculado pela fórmula: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ ou $\bar{x} = \mu_0 + z * \sigma / \sqrt{n}$
- 10.5. Então, suponha que você pesquisou uma amostra de 100 mulheres que trabalham e a média de tempo de conversa foi 11,5 minutos por dia, com desvio padrão de 2,3 minutos:
 - 10.5.1. O erro padrão será igual a σ / \sqrt{n} , isto é: $2,3 / \sqrt{100} = 0,23$
 - 10.5.2. O valor Z será igual a $(11,5 - 11,0) / 0,23 \rightarrow 0,5 / 0,23 = 2,17$
- 10.6. A estatística de teste Z é então igual a 2,17, isto é, está afastada (acima) da média padronizada 2,17 desvios-padrão.
- 10.7. Para decidir se esse valor suporta ou não a hipótese nula, calculamos o valor p dessa estatística, buscando esse valor p na tabela da distribuição normal padrão (distribuição Z); esse valor é 0,985.
- 10.8. Como estamos buscando a probabilidade da cauda superior (hipótese alternativa – tempo médio de conversa > 11 minutos), subtraímos esse valor de 1,00, isto é, $1,00 - 0,985 = 0,015$
- 10.9. Esse valor p de 0,015 (1,5%) é menor que o alfa padrão de 5%, portanto rejeitamos a hipótese nula:
 - 10.9.1. Significado do valor p : probabilidade de que a média 11,5 minutos tenha sido encontrado por acaso devido à amostra escolhida é 1,5%.
- 10.10. OBS: se o tamanho da amostra fosse menor que 30 ou o desvio padrão da população fosse desconhecido, a distribuição a ser utilizada seria a T e não a Z ;

11. EXERCÍCIO

Uma entidade não oficial, afirma que, em Jurucê, o gasto médio (μ) de um casal com lazer em um final de semana é 200 reais. Você acredita que esse valor é muito baixo e resolve fazer uma pesquisa.

- a) Qual é a hipótese nula? $H_0: \mu = 200$ reais
 b) Qual é a hipótese alternativa? $H_a: \mu > 200$ reais

Você pesquisou 36 famílias de duas pessoas escolhidas aleatoriamente na cidade. Os valores gastos.

190	185	240	181	233	163
201	230	195	225	231	260
169	240	250	211	189	233
245	184	220	245	210	240
238	155	180	201	250	255
220	205	205	185	174	153

- c) Qual foi o valor médio gasto? $\bar{x} = 210$ reais
 d) Qual foi o desvio padrão do gasto? $S = 30$ reais
 e) Qual é o erro padrão da média?

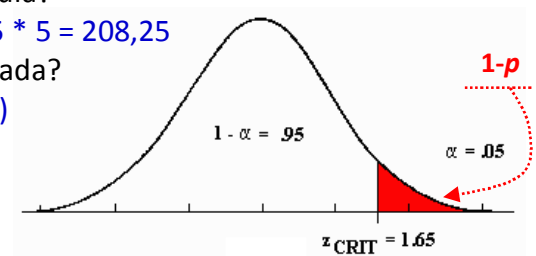
$$e = \frac{S}{\sqrt{n}} = \frac{30}{\sqrt{36}} = 5$$

Você decide que o erro aceitável (α) é de 5%:

- f) O que significa esse erro? Rejeitar a hipótese nula quando ela é verdadeira
 g) Que tipo de erro é esse? Erro Tipo I

Você busca na tabela da distribuição Z o valor de 5% de rejeição, que resulta em 1,65.

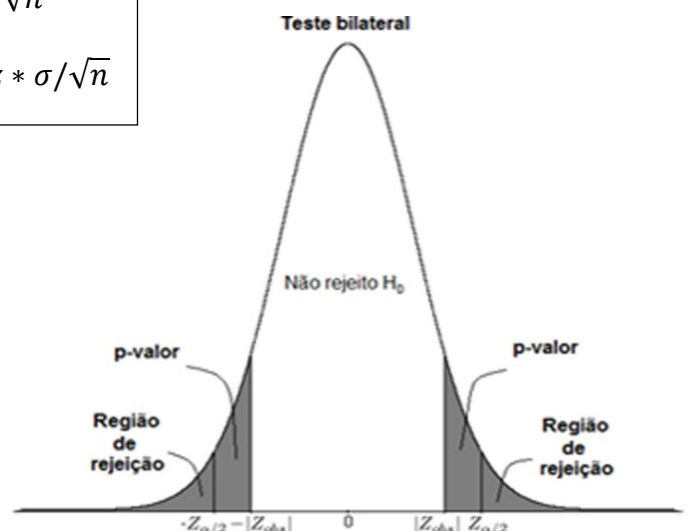
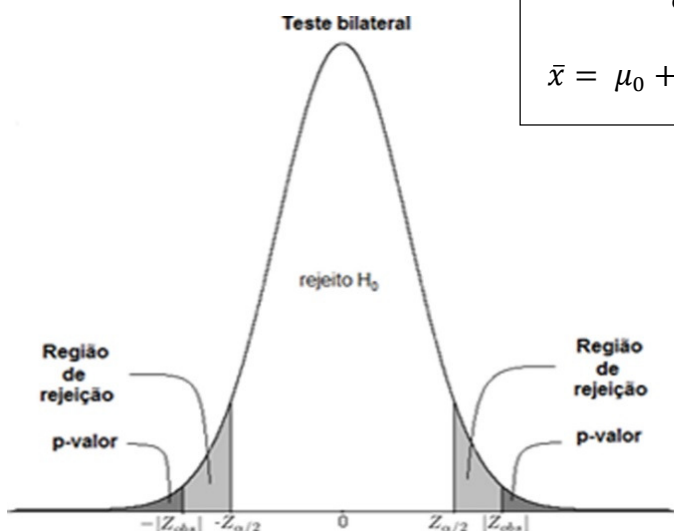
- h) Qual o valor limite para aceitação da hipótese nula?
 Valor limite: $\bar{x} + 1,65 * \text{erro padrão} = 200 + 1,65 * 5 = 208,25$
 i) Nesse caso, a hipótese nula será aceita ou rejeitada?
 Como $\bar{x} > \text{valor limite}$, será rejeitada ($210 > 208$)
 j) Como interpretar o erro tipo I nesse contexto?
 A probabilidade de se encontrar a média 210 por acaso (amostragem) é menor que 5%.



Obs: para testes bicaudais:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

$$\bar{x} = \mu_0 + z * \sigma / \sqrt{n}$$



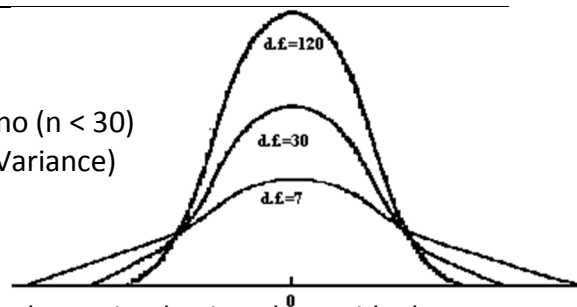
12. Teste t de Student

Empregado quando o tamanho da amostra é pequeno ($n < 30$)

Pode ser substituído pelo teste ANOVA (Analysis of Variance)

A distribuição da população deve ser normal

(ou aproximadamente normal)



Exemplo 1: a média histórica de tempo de volta em uma determinada pista de corrida de motos em uma determinada categoria é de 78 segundos. Em um determinado treino, um competidor, em 6 voltas alcançou a média de 79 segundos com um desvio padrão de 0,75 segundos. Pode-se afirmar, com $\alpha = 0,05$, que a média de tempo desse competidor é maior (mais lenta) que a média da categoria? E com $\alpha = 0,01$

- Hipótese nula H_0 : as médias são iguais
Hipótese alternativa H_1 : a média desse competidor é maior;
- Consultando-se a tabela t para 5 graus de liberdade ($n-1$) e $\alpha = 0,05$ encontramos o valor $t = 2,015$ e para $\alpha = 0,01$ encontramos o valor $t = 3,365$
- Se o valor t calculado (t_c) for maior ou igual ao valor t limite (t_l), rejeitamos a hipótese nula, se o t calculado for menor que o t limite não se rejeita a hipótese nula.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{79 - 78}{\frac{0,75}{\sqrt{6}}} = 3,26$$

Para $\alpha = 0,05$ $t_c > t_l$ rejeita-se H_0
Para $\alpha = 0,01$ $t_c < t_l$ não se rejeita H_0

Exemplo 2: a concentração de nitrogênio em bolhas de ar de resinas de ambar medidas em 9 amostras de 80 milhões de anos (indicam a atmosfera da época) é:

63,4	65,0	64,4	63,3	54,8	64,5	60,8	49,1	51,1
------	------	------	------	------	------	------	------	------

$$\begin{aligned}\bar{x} &= 59,6 \\ \sigma &= 6,24 \\ t_{0,05} &= 1,86\end{aligned}$$

Estime o intervalo de confiança de 90% para a média de nitrogênio na atmosfera da época.

$$IC_{0.90} = \bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}}$$

$$IC_{0.90} = 59,6 \pm 1,86 \frac{6,24}{\sqrt{9}}$$

$$IC_{0.90} = 59,6 \pm 3,87$$

$$55,73 \leq \mu \leq 63,47$$

13. Intervalo de Confiança para amostras pequenas

Método utilizado quando não se conhece a média da população, e a quantidade de amostras é pequena devido a algum tipo de limitação importante.

Exemplo: teste de blindagem em veículos. O teste é realizado alvejando-se a blindagem e medindo-se a profundidade da cavidade criada. Efetuou-se 10 disparos a determinada distância. Obteve-se uma média de 2,4 cm e desvio padrão de 0,411 cm. Determinar o intervalo de confiança de 95% para essa média.

Consultando-se a tabela t com 9 graus de liberdade e $P = 0,025$ encontramos $t = 2,262$

$$IC_{0.95} = \bar{x} \pm t_{0.025} \frac{s}{\sqrt{n}}$$

$$IC_{0.95} = 2,4 \pm 2,262 \frac{0,411}{\sqrt{10}}$$

$$IC_{0.95} = 2,4 \pm 0,294$$

$$2,16 \leq \mu \leq 2,694$$

12. INFERÊNCIA PARA DUAS POPULAÇÕES

1. A forma padrão de se testar a média de duas populações (para resultados numéricos):

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

\bar{x} : média da amostra x
 s_x : desvio padrão amostra x
 n_1 : número de amostras x

\bar{y} : média da amostra y
 s_y : desvio padrão da amostra y
 n_2 : número de amostras y

- Calcule a média \bar{x} e \bar{y} e o desvio padrão s_x e s_y das amostras de tamanho n_1 e n_2 ;
- Calcule a diferença entre as médias amostrais: $(\bar{x} - \bar{y})$;
- Calcule o erro padrão da média: $\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}$
- A estatística Z_a é dada por: diferença entre as médias / erro padrão da média
- Encontre o valor $P(Z \leq z)$ correspondente à estatística Z_a na tabela de distribuição t com $(n_1 + n_2 - 2)$ graus de liberdade (*Tabela A1 e A2 na página 27 e 28*);
- Calcule o valor $p = P$;
 - Se p for maior ou igual a α , aceita-se a hipótese nula;
 - Se p for menor que α rejeita-se a hipótese nula;

a) Exemplo: comparar o grau de absorvência de duas marcas de papel-toalha.

Hipótese nula: o grau de absorvência é igual (diferença zero);

Hipótese alternativa: o grau de absorvência é diferente;

São selecionadas aleatoriamente 50 amostras de cada marca e a absorvência é testada.

Amostra A: absorvência de 3,0 ml com desvio padrão de 0,9 ml;

Amostra B: absorvência de 3,5 ml com desvio padrão de 1,2 ml;

O erro padrão da média é dado pela equação:

$$Z = \frac{(3,0 - 3,5)}{\sqrt{\frac{0,9^2}{50} + \frac{1,2^2}{50}}} = \frac{(-0,5)}{\sqrt{\frac{0,81}{50} + \frac{1,44}{50}}} = \frac{-0,05}{\sqrt{\frac{2,25}{50}}} = \frac{-0,05}{\sqrt{0,045}} = \frac{-0,5}{0,2121} = -2,36$$

Na tabela Z, o valor correspondente a -2,36 é 0,0091 ou 0,91%. Esse valor corresponde à área sob a curva, menor que o valor $Z = -2,36$ e indica a probabilidade do resultado da marca A ser igual ao da marca B, e termos encontrado essa diferença (-0,5) devido ao acaso da amostra. Como a hipótese alternativa é que os resultados são diferentes (poderia ser maior ou menor) dobramos esse valor ($2 \times 0,0091 = 0,0182 = 1,82\%$).

* Isso significa que temos evidências para rejeitar a hipótese nula ($1,82\% < 5\%$);

* A conclusão é, com base nas amostras estudadas,

- a diferença entre as médias é estatisticamente significativa, então
- assumimos que as médias são diferentes

OBSERVAÇÃO: Se o número de amostras fosse menor que 30, utilizaríamos a distribuição t e não a distribuição Z – os procedimentos de cálculo são os mesmos da distribuição z:

Intervalo de Confiança:

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Valor t:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

2. Exercício:

- a) Um estudo sobre regulagem de motores (mesma marca e modelo e condições) mostrou que os 57 veículos que não passaram pela regulagem consumiram, em média, 105.32 litros de combustível, com um desvio padrão de 14.68 litros. Os 17 veículos que passaram pela regulagem consumiram 96.82 litros de combustível com um desvio padrão de 14,26 litros. As distribuições de consumo são aproximadamente normais. Pode-se afirmar que a regulagem reduziu o consumo de combustível (utilizar alfa de 0,05)?

Valor t:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{105.32 - 96.82}{\sqrt{\frac{14.68^2}{57} + \frac{14.26^2}{17}}} = \frac{8.5}{3.9677} = 2.142$$

O menor grau de liberdade é
57-1 = 56 ou 17-1 = 16

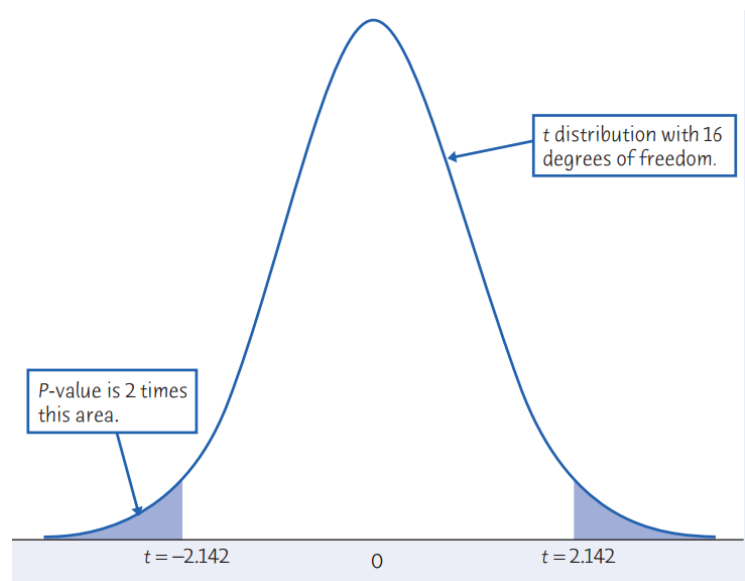
Na tabela t encontramos:

P(2.142 / 16) = 0.02 a 0.025

Teste bicaudal: 0.04 a 0.05

Conclusão: $P < 0.05$

Rejeita-se a hipótese nula



- b) Afirma-se que a compactação do solo úmido reduz mais a penetrabilidade de raízes do que a compactação do solo seco. Foram feitos dois experimentos de medida de penetrabilidade, mostrados abaixo. Pode-se afirmar, com significância estatística de 5%, que os resultados confirmam essa afirmação:

Compactação com solo úmido (amostra 1)

2.86	2.68	2.92	2.82	2.76	2.81	2.78	3.08	2.94	2.86
3.08	2.82	2.78	2.98	3.00	2.78	2.96	2.90	3.18	3.16

Compactação com solo seco (amostra 2)

3.14	3.38	3.10	3.40	3.38	3.14	3.18	3.26	2.96	3.02
3.54	3.36	3.18	3.12	3.86	2.92	3.46	3.44	3.62	4.26

Solução

$$\begin{aligned}\bar{x}_1 &= 2,91 \\ S_1 &= 0,139 \\ n_1 &= 20\end{aligned}$$

$$\begin{aligned}\bar{x}_2 &= 3,34 \\ S_2 &= 0,319 \\ N_2 &= 20\end{aligned}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{2,91 - 3,34}{\sqrt{\frac{0,139^2}{20} + \frac{0,319^2}{20}}} = -5,53$$

Consultando a tabela t para 19 graus de liberdade e alfa = 0,05 observamos o t crítico igual a 1,729 que é menor que 5,53 levando à rejeição da hipótese nula

3. Comparação entre resultados de dois procedimentos diferentes (proporção).

Um teste em 156 amostras iguais (por exemplo tratamento contra fogo) resultou que o procedimento A aplicado a 73 amostras foi bem sucedido em 67 delas ($p_1 \cong 92\%$) enquanto o procedimento B aplicado a 83 amostras foi bem sucedido em 60 delas ($p_2 \cong 72\%$). Pode-se afirmar que o tratamento A é melhor que o tratamento B?

- Amostras são independentes?;
- Para cada amostra, o número de sucessos e de fracassos é no mínimo 5?;
 - Hipótese Nula (H_0): $p_1 = p_2$; nível de significância $\alpha = 0,05$.
 - Hipótese alternativa (H_1): $p_1 > p_2$

p_1 : proporção na população
 n_1 : tamanho da amostra
 x_1 : número de sucessos na amostra

$$\hat{p}_1 = \frac{x_1}{n_1} \quad (\text{proporção na amostra})$$

$$\hat{q}_1 = 1 - \hat{p}_1 \quad (\text{complemento da proporção})$$

Proporção Combinada:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

Hipótese Nula:
 $p_1 = p_2$

Procura-se o valor **P** na tabela Z e compara-se com o valor **α** adotado.
 Se **$P > \alpha$** aceita-se a hipótese nula - se **$P < \alpha$** rejeita-se a hipótese nula.
 Se **$P = \alpha$** ...

Intervalo de Confiança para a diferença entre duas proporções

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

Supondo que os resultados seguem uma distribuição normal:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{67 + 60}{73 + 83} = 0.81410256$$

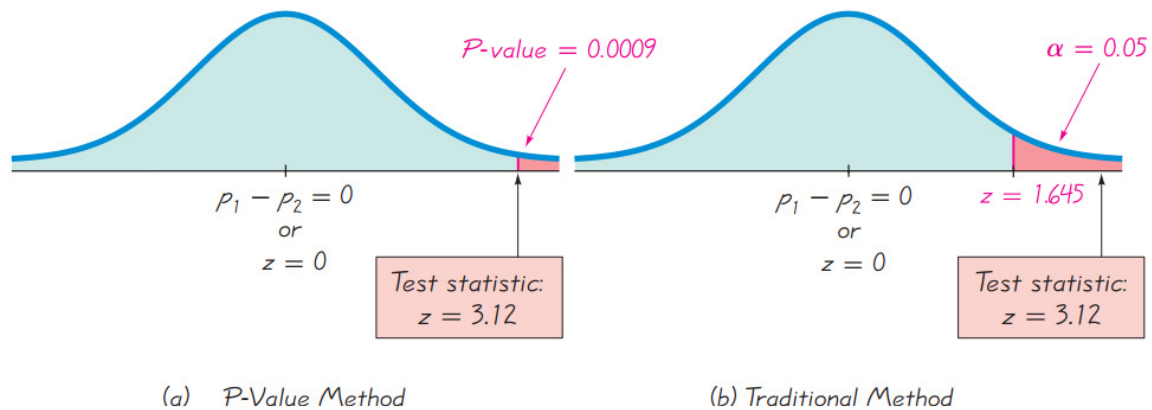
$$\bar{p} = 0.81410256$$

$$\bar{q} = 1 - \bar{p} = 0.18589744$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = Z = \frac{(0.92 - 0.72) - 0}{\sqrt{\frac{0.81410256 \cdot 0.18589744}{73} + \frac{0.81410256 \cdot 0.18589744}{83}}} = 3,12$$

Consultando a Tabela Z encontramos para $z=3,12$ o valor $P=0,9991$, que indica a probabilidade de do resultado ser menor ou igual a z . Assim, a probabilidade do resultado ser maior que z é igual a $1 - P$, ou seja, $1 - 0,9991 = 0,0009$.

Como P é menor que α ($0.0009 < 0.05$) rejeitamos a hipótese nula e aceitamos a hipótese alternativa, isto é, $p_1 > p_2$, concluindo que temos evidências estatísticas de que o tratamento A é melhor do que o tratamento B.



Pelo método tradicional, comparamos o valor z encontrado ($z=3,12$) com o valor z correspondente ao valor α adotado (para $\alpha = 0,05$ o valor $z = 1,645$):

- Se o valor encontrado for maior que o valor limite, rejeitamos a hipótese nula:
 - as diferenças são estatisticamente significativas;
- Se o valor encontrado for menor que o valor limite, aceitamos a hipótese nula:
 - as diferenças não são estatisticamente significativas.

Pelo método do Intervalo de confiança, construímos um intervalo de confiança para a diferença entre as proporções:

- Se o intervalo contiver o Zero, aceitamos a hipótese nula:
 - não temos evidências estatísticas de que as proporções são diferentes.
- Se o intervalo não contiver o Zero, rejeitamos a hipótese nula:
 - temos evidências estatísticas de que as proporções são diferentes

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = 1.645 \sqrt{\frac{\left(\frac{67}{73}\right)\left(\frac{6}{73}\right)}{73} + \frac{\left(\frac{60}{83}\right)\left(\frac{23}{83}\right)}{83}} = 0.0966$$

$$\hat{p}_1 = 67/73 = 0.9178 \quad \hat{p}_2 = 60/83 = 0.7299 \quad E = 0.0966$$

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.9178 - 0.7229) - 0.0966 < (p_1 - p_2) < (0.9178 - 0.7229) + 0.0966$$

$$0.0983 < (p_1 - p_2) < 0.292$$

OBS: Teste pelo valor P considera o desvio padrão das duas POPULAÇÕES;

Considera que não há diferença entre as duas proporções;

Teste pelo intervalo de confiança considera o desvio padrão das duas AMOSTRAS

Considera que há uma diferença entre as duas proporções

ATENÇÃO: teste monocaual.: z_α
 teste bicaual.....: $z_{\alpha/2}$

4. ROTEIRO

A forma padrão de se testar a média de duas populações (para resultados numéricos):

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

\bar{x} : média da amostra x
 s_x : desvio padrão amostra x
 n_1 : número de amostras x

\bar{y} : média da amostra y
 s_y : desvio padrão da amostra y
 n_2 : número de amostras y

- Calcule a média \bar{x} e \bar{y} e o desvio padrão s_x e s_y das amostras de tamanho n_1 e n_2 ;
- Calcule a diferença entre as médias amostrais: $(\bar{x} - \bar{y})$;
- Calcule o erro padrão da média: $\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}$
- A estatística Z_a é dada por: diferença entre as médias / erro padrão da média:
- Encontre o valor $P(Z_a \leq z)$ correspondente à estatística Z_a na tabela de distribuição Z (ou, se $n < 30$, na *tabela t*, com $(n_1 + n_2 - 2)$ graus de liberdade);
- Compare o valor P encontrado na tabela com o valor α (alfa) adotado:
 - Se P for maior ou igual a α , não se rejeita a hipótese nula;
 - Se P for menor que α , rejeita-se a hipótese nula.

$$Z = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

5. EXEMPLO

- Para testar os efeitos colaterais de uma nova droga para combate à enxaqueca, um pesquisador ministrou o medicamento a 500 pacientes e um placebo a 400 pacientes selecionados aleatoriamente de um grupo com as mesmas características de interesse ao estudo. No grupo que recebeu o medicamento, 100 pacientes apresentaram efeitos colaterais. No grupo de placebo, 50 pacientes apresentaram efeitos colaterais semelhantes. Pode-se afirmar que há diferença estatisticamente significativa entre os efeitos nesses dois grupos? Utilizar o método do valor P e o método do Intervalo de Confiança ($\alpha = 5\%$, *unicaudal*).

Teste pelo Intervalo de Confiança para a diferença entre duas proporções

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = \left(\frac{100}{500} - \frac{50}{400} \right) \pm 1.645 \sqrt{\frac{100 \cdot 400}{500 \cdot 500} + \frac{50 \cdot 50}{400 \cdot 400}}$$

$$(0.2 - 0.125) \pm 1.645 \sqrt{\frac{0.16}{500} + \frac{0.0156}{400}} = 0.075 \pm 0.0006 \quad 7.44\% \text{ a } 7.56\%$$

O intervalo não contém o zero (está relativamente distante), portanto concluímos que temos evidências estatísticas de que as proporções são diferentes.

Teste pelo valor P:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \bar{p} = \frac{100 + 50}{500 + 400} = 0.1667 \quad \bar{q} = 1 - 0.1667 = 0.8333$$

$$Z = \frac{(0.16 - 0.0156) - (0)}{\sqrt{\frac{0.1667 \cdot 0.8333}{500} + \frac{0.1667 \cdot 0.8333}{400}}} = \frac{0.0750}{\sqrt{\frac{0.1667 \cdot 0.8333}{500} + \frac{0.1667 \cdot 0.8333}{400}}} = 3.0 \quad P(3.0) = 0.9987$$

$P(z > 3) = 1 - 0.9987 = 0.0013 = \mathbf{0,13\%} \lll 0.05$ portanto rejeita-se a hipótese nula

6. EXEMPLO

Comparar o grau de absorvência de duas marcas de papel-toalha.

Hipótese nula: o grau de absorvência é igual (diferença zero);

Hipótese alternativa: o grau de absorvência é diferente (sem indicar qual é a maior);

São selecionadas aleatoriamente 50 amostras de cada marca e a absorvência é testada.

Amostra A: absorvência de 3,0 ml com desvio padrão de 0,9 ml;

Amostra B: absorvência de 3,5 ml com desvio padrão de 1,2 ml;

O erro padrão da média é dado pela equação:

$$Z = \frac{(3,0-3,5)}{\sqrt{\frac{0,9^2}{50} + \frac{1,2^2}{50}}} = \frac{(-0,5)}{\sqrt{\frac{0,81}{50} + \frac{1,44}{50}}} = \frac{-0,05}{\sqrt{\frac{2,25}{50}}} = \frac{-0,05}{\sqrt{0,045}} = \frac{-0,5}{0,2121} = -2,36$$

Na tabela Z, o valor correspondente a -2,36 é 0,0091 ou 0,91%. Esse valor corresponde à área sob a curva, menor que o valor Z = -2,36 e indica a probabilidade do resultado da marca A ser igual ao da marca B, e termos encontrado essa diferença (-0,5) devido ao acaso da amostra. Como a hipótese alternativa é que os resultados são diferentes (poderia ser maior ou menor) dobramos esse valor ($2 \times 0,0091 = 0,0182 = 1,82\%$).

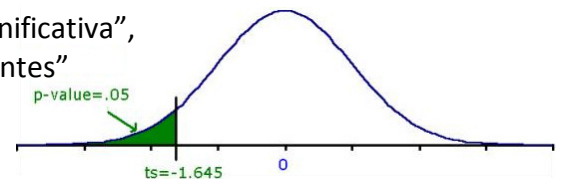
Isso significa que temos evidências para rejeitar a hipótese nula ($1,82\% < 5\%$);

A conclusão é, com base nas amostras estudadas:

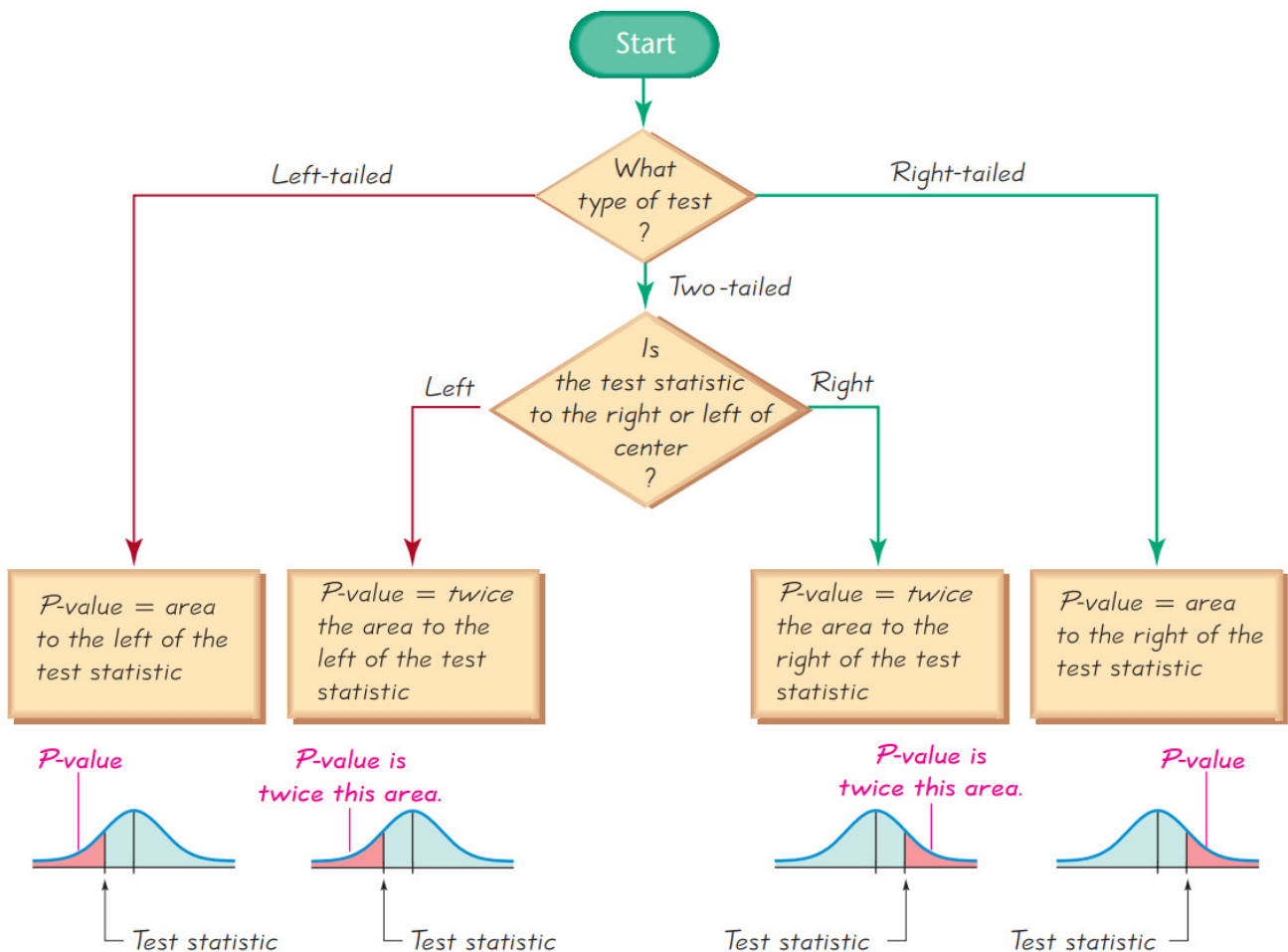
“A diferença entre as médias é estatisticamente significativa”,

Obs1: não podemos afirmar que “as médias são diferentes”

Obs.2: Se o número de amostras fosse menor que 30, utilizaríamos a distribuição t e não a distribuição Z.



7. Procedimento para encontrar o valor P:



13. ANÁLISE DE VARIÂNCIA (ANOVA)

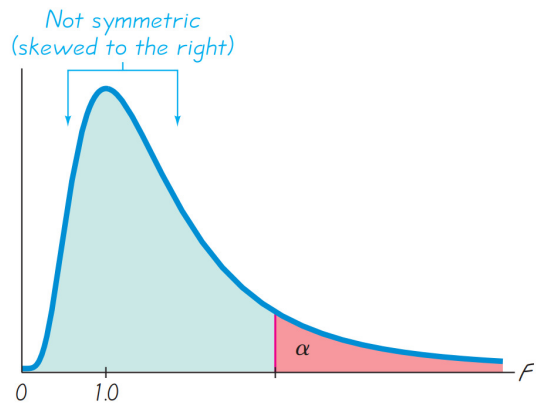
1. DEFINIÇÃO:

Análise de Variância (ANOVA) é um método de teste para a igualdade de três ou mais médias de população analisando a variância das amostras.

Não é válido comparar as médias duas a duas porque a chance de um erro tipo I (encontrar diferença onde ela não existe) aumenta na medida em que a quantidade de médias comparadas aumenta.

O método ANOVA utiliza a distribuição **F**, que apresenta as seguintes propriedades:

- Não é simétrica: é inclinada à direita;
- Os valores de F podem ser zero ou positivos, não podem ser negativos;
- Existe uma distribuição F para cada par de graus de liberdade para o numerador e o denominador.



Baseia-se na comparação de duas estimativas diferentes da variância comum a diferentes populações: a variância entre as amostras e a variância dentro de cada amostra.

2. ONE WAY ANOVA (ANOVA Fator Único)

O termo “One Way” é utilizado quando as amostras são separadas em grupos de acordo com uma característica ou fator. O termo “Two Way” é utilizado quando se estuda uma determinada característica, mas as amostras são categorizadas utilizando dois fatores, por exemplo sexo (masculino/feminino) e estado civil (solteiro, casado, viuvo, etc.).

O termo “Three Way”, é utilizado quando se categoriza as amostras utilizando três fatores.

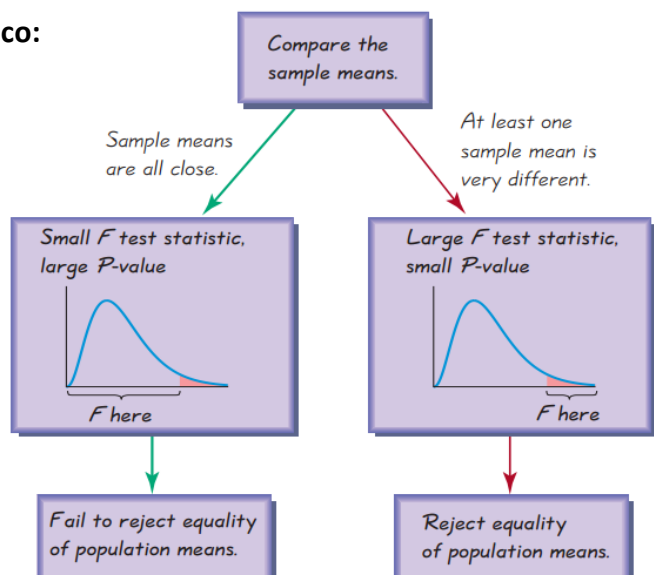
Requisitos:

- As populações devem ter distribuições aproximadamente normais;
 - Se a distribuição não é normal, utiliza-se o teste Kruskal-Wallis;
- As populações devem ter variâncias semelhantes (mas o teste funciona bem se as variâncias estiverem em uma razão de até 9:1);
- As amostras devem ser do tipo aleatória simples;
- As amostras devem ser independentes;
- As amostras são de populações categorizadas de apenas um modo;

3. Estatística de teste para ANOVA Fator Único:

$$F = \frac{\text{variância entre as amostras}}{\text{variância dentro das amostras}}$$

Obs: todas as amostras devem ter o mesmo tamanho (n).

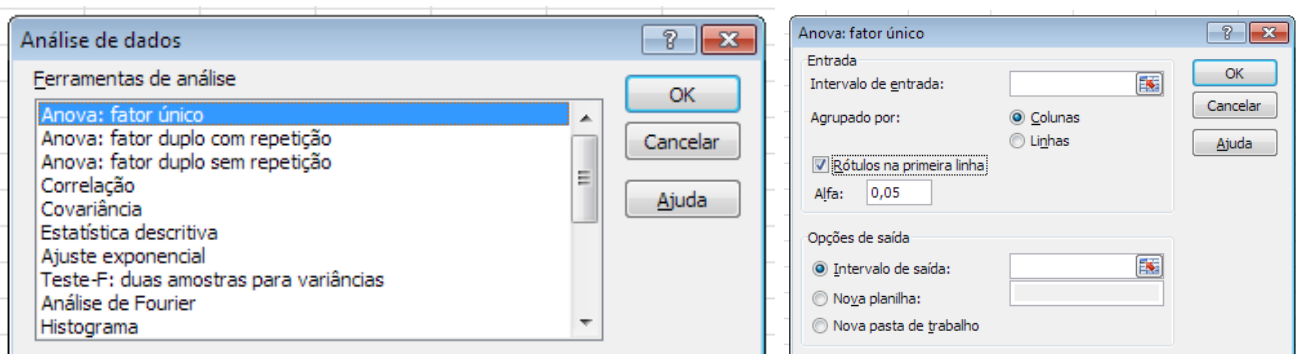


4. ANOVA fator único: Exemplo de cálculo

Table 12-2			Effect of a Mean on the F Test Statistic		
A add 10			B		
Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
7	6	4	17	6	4
3	5	7	13	5	7
6	5	6	16	5	6
6	8	7	16	8	7
$n_1 = 4$	$n_2 = 4$	$n_3 = 4$	$n_1 = 4$	$n_2 = 4$	$n_3 = 4$
$\bar{x}_1 = 5.5$	$\bar{x}_2 = 6.0$	$\bar{x}_3 = 6.0$	$\bar{x}_1 = 15.5$	$\bar{x}_2 = 6.0$	$\bar{x}_3 = 6.0$
$s_1^2 = 3.0$	$s_2^2 = 2.0$	$s_3^2 = 2.0$	$s_1^2 = 3.0$	$s_2^2 = 2.0$	$s_3^2 = 2.0$
Variance between samples	$ns_{\bar{x}}^2 = 4 (0.0833) = 0.3332$ tamanho da amostra (n) Variância das Médias amostrais (\bar{x}) Média das Variância amostrais (s^2) $s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$		Variance between samples	$ns_{\bar{x}}^2 = 4 (30.0833) = 120.3332$ tamanho da amostra (n) Variância das Médias amostrais (\bar{x}) Média das Variância amostrais (s^2) $s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$	
Variance within samples			Variance within samples		
F test statistic	$F = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{0.3332}{2.3333} = 0.1428$		F test statistic	$F = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{120.3332}{2.3333} = 51.5721$	
P-value (found from Excel)	P-value = 0.8688		P-value (found from Excel)	P-value = 0.0000118	
			Degrees of Freedom: (k = number of samples and n = sample size) numerator degrees of freedom = k - 1 denominator degrees of freedom = k (n - 1)		

5. ANOVA fator único no Microsoft Excel

- Habilitar Ferramentas de Análise em:
Arquivo => Opções => Suplementos => Ferramentas de Análise
- Acessar ferramentas de Análise em:
Dados => Análise de Dados



CASO 1			CASO 2		
A	B	C	A	B	C
7	6	4	17	6	4
3	5	7	13	5	7
6	5	6	16	5	6
6	8	7	16	8	7

Anova: fator único

\$B\$2:\$D\$6

CASO 1

			Anova: fator único				
A	B	C	RESUMO				
7	6	4	Grupo	Contagem	Soma	Média	Variância
3	5	7	A	4	22	5,5	3
6	5	6	B	4	24	6	2
6	8	7	C	4	24	6	2
			ANOVA				
			Fonte da variação	SQ	gl	MQ	F
			Entre grupos	0,66667	2	0,33333	0,14286
			Dentro dos grupos	21	9	2,33333	0,86881
			Total	21,6667	11		4,25649

CASO 2

			Anova: fator único				
A	B	C	RESUMO				
17	6	4	Grupo	Contagem	Soma	Média	Variância
13	5	7	A	4	62	15,5	3
16	5	6	B	4	24	6	2
16	8	7	C	4	24	6	2
			ANOVA				
			Fonte da variação	SQ	gl	MQ	F
			Entre grupos	240,6667	2	120,3333	51,57143
			Dentro dos grupos	21	9	2,333333	1,18E-05
			Total	261,6667	11		4,256495

6. ANOVA fator único com amostras de tamanhos diferentes (n diferentes):

Os cálculos se tornam realmente complicados quando o tamanho das amostras não é o mesmo (n_i diferentes). Normalmente o cálculo é feito com auxílio de programas de estatística.

$$F = \frac{\text{variance between samples}}{\text{variance within samples}} = \frac{\left[\frac{\sum n_i (\bar{x}_i - \bar{\bar{x}})^2}{k - 1} \right]}{\left[\frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)} \right]}$$

$\bar{\bar{x}}$ = mean of all sample values combined
 k = number of population means being compared
 n_i = number of values in the i th sample
 \bar{x}_i = mean of values in the i th sample
 s_i^2 = variance of values in the i th sample

7. Testes de comparação múltipla

- Compara as amostras duas a duas e indica se há diferença estatisticamente significativa entre cada par de amostras. São efetuados utilizando-se programas de estatística (SPSS, SAS, GraphPad Prism, Statistica, etc.)
- Não há consenso sobre qual teste é o melhor. Os testes mais comuns são
 - Teste de Duncan
 - Teste de Student-Newman-Keuls (SNK)
 - Teste de Tukey
 - Teste de Scheffé
 - Teste de Bonferroni

$$t = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{MS(\text{error}) \cdot \left(\frac{1}{n_1} + \frac{1}{n_4} \right)}}$$

8. ANOVA fator duplo:

Quando os dados são categorizados em dois fatores. Por exemplo, comparar o peso de uma determinada árvore (certo tempo após o plantio da semente) considerando o tipo de solo e o tratamento dado ao solo:

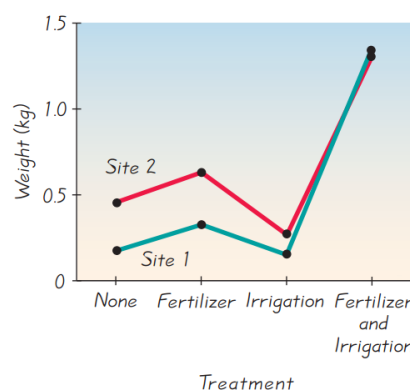
ANOVA fator duplo sem repetição: apenas uma medida para cada categoria (média);

ANOVA fator duplo com repetição: várias medidas para cada categoria (caso abaixo);

Table 12-4 Poplar Tree Weights (kg)				
	No Treatment	Fertilizer	Irrigation	Fertilizer and Irrigation
Site 1 (rich, moist)	0.15	1.34	0.23	2.03
	0.02	0.14	0.04	0.27
	0.16	0.02	0.34	0.92
	0.37	0.08	0.16	1.07
	0.22	0.08	0.05	2.38
Site 2 (sandy, dry)	0.60	1.16	0.65	0.22
	1.11	0.93	0.08	2.13
	0.07	0.30	0.62	2.33
	0.07	0.59	0.01	1.74
	0.44	0.17	0.03	0.12

Table 12-5 Means (kg) of Cells from Table 12-4

	Treatment			
	None	Fertilizer	Irrigation	Fertilizer and Irrigation
Site 1 (rich, moist)	0.184	0.332	0.164	1.334
Site 2 (sandy, dry)	0.458	0.630	0.278	1.308



A análise é feita utilizando-se o software Microsoft Excel ou algum programa de estatística específico (SPSS, SAS, GraphPad Prism, Statistica, etc).

					Anova: fator duplo com repetição					
	no treatment	fertilizer	irrigation	fertilizer & irrigation	RESUMO	no treatment	fertilizer	irrigation	fertilizer & irrigation	Total
Site 1	0,15	1,34	0,23	2,03	Site 1					
rich	0,02	0,14	0,04	0,27	Contagem	5	5	5	5	20
moist	0,16	0,02	0,34	0,92	Soma	0,92	1,66	0,82	6,67	10,07
	0,37	0,08	0,16	1,07	Média	0,184	0,332	0,164	1,334	0,5035
	0,22	0,08	0,05	2,38	Variância	0,01613	0,31932	0,01593	0,73793	0,4757713
Site 2	0,6	1,16	0,65	0,22	Site 2					
sandy	1,11	0,93	0,08	2,13	Contagem	5	5	5	5	20
dry	0,07	0,3	0,62	2,33	Soma	2,29	3,15	1,39	6,54	13,37
	0,07	0,59	0,01	1,74	Média	0,458	0,63	0,278	1,308	0,6685
	0,44	0,17	0,03	0,12	Variância	0,18667	0,17325	0,10697	1,12547	0,4950345
					Total					
					Contagem	10	10	10	10	
					Soma	3,21	4,81	2,21	13,21	
					Média	0,321	0,481	0,221	1,321	
					Variância	0,1109878	0,2435878	0,0582322	0,8283656	
					ANOVA					
					Fonte da variação	SQ	gl	MQ	F	valor-P
					Amostra	0,27225	1	0,27225	0,8121805	0,3742095
					Colunas	7,547	3	2,5156667	7,5047763	0,0006136
					Interações	0,17163	3	0,05721	0,1706698	0,915411
					Dentro	10,72668	32	0,3352088		
					Total	18,71756	39			

Os resultados indicam valor P para as amostras (linhas = sites) igual a 0,37 levando a não rejeitar a hipótese nula (as diferenças entre os sites não são estatisticamente significativas) e o valor P para os tratamentos (colunas = tratamentos) igual a 0,0006 levando a rejeitar a hipótese nula (as diferenças são estatisticamente significativas quanto aos tratamentos). Finalmente o valor P para a interação é de 0,91 levando a não rejeição da hipótese nula quanto à interação entre os dois fatores (site e tratamento) indicando que a diferença devido à interação entre o site e o tratamento não é estatisticamente significativa.

Obs: Existem críticas e objeções à utilização de planilhas eletrônicas para cálculos estatísticos. Quando uma maior precisão nos resultados é necessária ou desejável, é recomendável a utilização de softwares específicos para essa finalidade.

14. CORRELAÇÃO E REGRESSÃO

1. Definição

- Uma correlação é uma relação entre duas variáveis. Os dados podem ser representados por pares ordenados (x, y) onde x é a variável independente (ou explanatória) e y é a variável dependente (ou resposta).

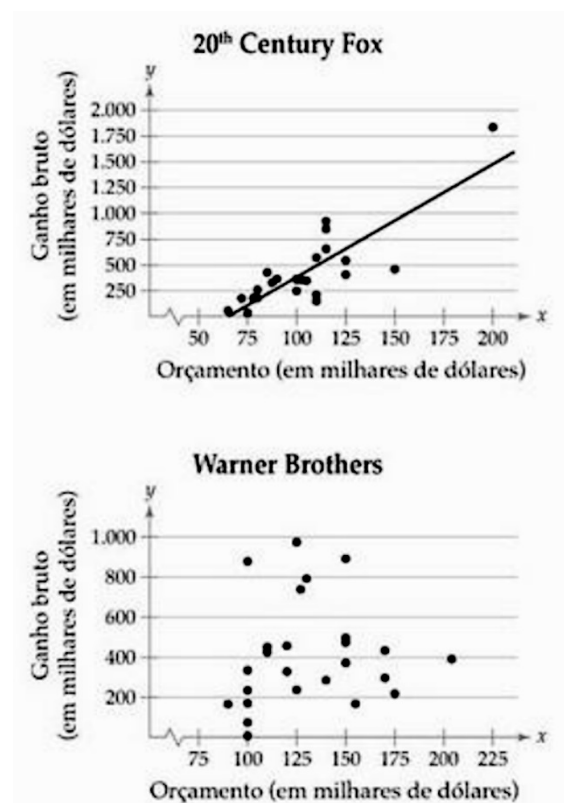
2. Coeficiente de Correlação de Pearson (r):

- Indica o grau de correlação linear entre as variáveis e varia de -1 a +1.
 - ✓ O índice +1 representa correlação linear positiva máxima;
 - ✓ O índice -1 indica correlação linear negativa máxima;
 - ✓ O índice 0 (zero) indica ausência de correlação linear.
- No Microsoft Excel: $r = \text{CORREL}(\text{Matriz A}; \text{Matriz B})$.

3. Exemplo:

O orçamento dos 25 filmes mais caros da 20th Century Fox e seus respectivos ganhos brutos têm alguma correlação linear?

25 filmes mais caros da XX Century Fox (US\$ milhões)		
N	Orçamento	Ganhos brutos
1	200	1.835,4
2	150	459,4
3	125	406,4
4	125	542,7
5	115	924,3
6	115	656,7
7	115	848,5
8	110	571,1
9	110	211,4
10	110	150,5
11	105	348,8
12	102	358,8
13	100	365,3
14	100	359,1
15	100	249,0
16	90	365,0
17	87,5	329,5
18	85	427,2
19	80	260,7
20	80	197,1
21	78	179,3
22	75	34,0
23	75	36,8
24	72	176,1
25	65	56,4



No Microsoft Excel, elaborar o gráfico de dispersão (x, y) e adicionar linha de tendência habilitando a exibição do R^2 . Calcular o coeficiente de correlação de Pearson.

Como fica o valor do R^2 e o coeficiente de correlação de Pearson se excluirmos do cálculo o filme mais caro e com maior ganhos brutos?

4. Coeficiente de Correlação r e Coeficiente de Determinação R^2

São diversas as fórmulas utilizadas para o cálculo do coeficiente de correlação r . Todas apresentam resultados equivalentes em termos práticos. O coeficiente de determinação R^2 é calculado elevando-se o coeficiente de correlação r ao quadrado.

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y}$$

$$r = \frac{\sum (x_1 - x_m)(y_1 - y_m)}{\sqrt{\sum (x_1 - x_m)^2} \sqrt{\sum (y_1 - y_m)^2}}$$

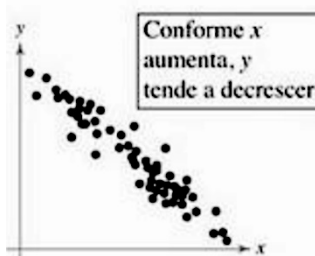
$$r = \frac{\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y} \right]}{n - 1}$$

$$r = \frac{s_{xy}}{\sqrt{s_{xx}} \sqrt{s_{yy}}}$$

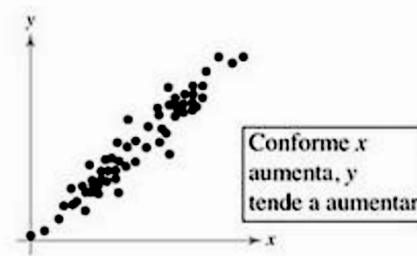
O diagrama ao lado é chamado “diagrama de dispersão”. A variável independente (explanatória) é representada no eixo horizontal e a variável dependente (resposta) é representada no eixo vertical.

O quadrado do coeficiente de correlação r , chamado coeficiente de determinação R^2 , indica o percentual da variância de uma variável que pode ser explicado a partir do valor da outra variável.

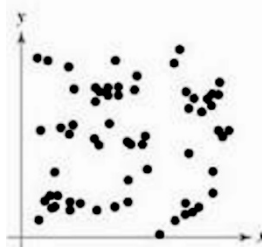
Correlação linear negativa



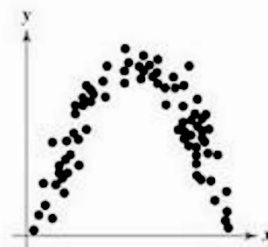
Correlação linear positiva



Não há correlação



Correlação não linear



5. Classificação do Grau de Correlação

São várias as propostas de interpretação da força da correlação a partir do coeficiente r . Uma classificação comumente encontrada é:

Valor de ρ (+ ou -)	Interpretação
0.00 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

6. Exercícios:

Construa o diagrama de dispersão, calcule o índice de correlação r de Pearson e o coeficiente de determinação R^2 das situações abaixo:

a) Gastos com propaganda e vendas na empresa em determinado período:

em U\$ mil

Gastos com propaganda	Vendas da Empresa
2,4	225
1,6	184
2,0	220
2,6	240
0,4	180
1,6	184
2,0	186
2,2	215

c) Peso do indivíduo e consumo diário de água em determinado mês:

Peso: libras – água (onças)

Peso	Água
142	54
201	86
119	32
102	50
141	64
124	82
220	39
154	21

b) Nível de renda e % da renda doada em determinado período:

em U\$ mil

Nível de Renda	% de Doações
50	8
65	6
48	10
42	9
59	5
72	3

7. Teste de Significância do Coeficiente de Correlação r

Indica se há evidência para considerar significativo o coeficiente de correlação r , isto é, ao se analisar uma amostra, em que nível de significância pode-se fazer uma inferência sobre a população. Os níveis limites são, em geral, 0,05 (5%) ou 0,01 (1%).

Utiliza-se a tabela de valores críticos para o coeficiente de correlação r de Pearson.

Se o valor absoluto de r for maior que o valor encontrado na tabela para o alfa adotado e o número de pares de amostras testado, rejeita-se a hipótese nula (a correlação é significativa).

8. Exercício

Verifique se a correlação é significativa para as situações apresentadas no exercício do item 6 e no exemplo do item 3 (página anterior).

n	$\alpha = 0,05$	$\alpha = 0,01$
4	0,950	0,990
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
19	0,456	0,575
20	0,444	0,561
21	0,433	0,549
22	0,423	0,537
23	0,413	0,526
24	0,404	0,515
25	0,396	0,505
26	0,388	0,496
27	0,381	0,487
28	0,374	0,479
29	0,367	0,471
30	0,361	0,463
35	0,334	0,430
40	0,312	0,403
45	0,294	0,380
50	0,279	0,361
55	0,266	0,345
60	0,254	0,330
65	0,244	0,317
70	0,235	0,306
75	0,227	0,296
80	0,220	0,286
85	0,213	0,278
90	0,207	0,270
95	0,202	0,263
100	0,197	0,256

9. Teste t para o coeficiente de correlação r

Pode-se também utilizar a tabela t para avaliar a significância do coeficiente de correlação r :

- Calcula-se o valor t utilizando-se a equação:
 - ✓ r é o coeficiente de correlação de Pearson
 - ✓ n é o número de pares de amostras
- Compara-se o valor t calculado com o valor t limite correspondente ao valor alfa adotado:
 - ✓ Se o t calculado é maior que o t limite adotado, rejeita-se a hipótese nula;
 - ✓ Se o t calculado é menor que o t limite adotado, aceita-se a hipótese nula.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

10. Exemplo

Ao se analisar a correlação de 8 pares de amostra, encontrou-se um valor $r = 0,9129$. Teste a significância desse coeficiente de correlação utilizando um alfa de 0,05.

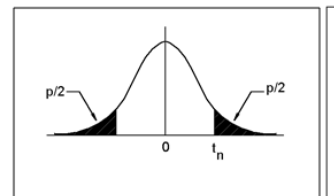
- Hipótese nula: A correlação não tem significância estatística para alfa = 0,05.
- Hipótese alternativa: a correlação tem significância estatística para alfa = 0,05.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$t = \frac{0,9129}{\sqrt{\frac{1-0,9129^2}{8-2}}}$$

$$t \approx 5,478$$

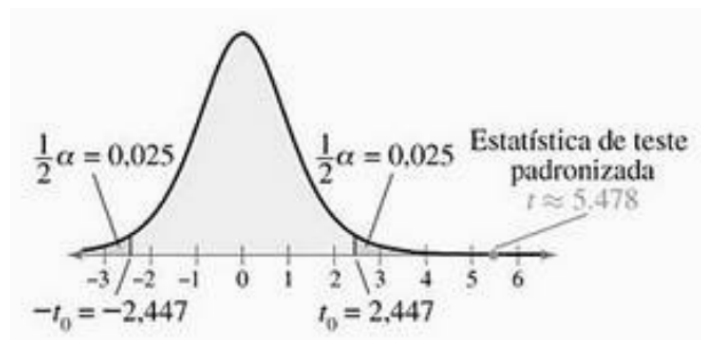
Distribuição t
Valores críticos



Probabilidade P

df	.25	.20	.15	.10	.05	.025	.02
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89
2	.816	1.061	1.386	1.886	2.920	4.303	4.849
3	.765	.978	1.250	1.638	2.353	3.182	3.482
4	.741	.941	1.190	1.533	2.132	2.776	2.999
5	.727	.920	1.156	1.476	2.015	2.571	2.757
6	.718	.906	1.134	1.440	1.943	2.447	2.612
7	.711	.896	1.119	1.415	1.895	2.365	2.517

Como o valor t calculado (5,478) é maior que o valor t limite (2,447) rejeita-se a hipótese nula: a correlação tem significância estatística para alfa = 0,05.



11. Observações importantes:

- O fato de duas variáveis serem fortemente correlacionadas não implica uma relação direta de causa e efeito entre elas.
 - ✓ Não se pode afirmar que X causa Y e nem que Y causa X.
- É possível que a correlação entre as duas variáveis possa ser causada por uma terceira variável, ou combinação de diversas outras variáveis.
 - ✓ A arrecadação de um filme não depende só do orçamento, mas também da temática, dos atores, do marketing, do momento, etc.
- É possível que uma correlação forte entre duas variáveis seja apenas coincidência.

12. Mais Exercícios:

Construa o diagrama de dispersão, calcule o índice de correlação de Pearson e o coeficiente de determinação R^2 , verifique se a correlação é significativa considerando um alfa de 0,05 e comente a credibilidade da associação nas situações abaixo:

a) Idade (anos) e pressão sistólica (mm.Hg)

Idade (x)	16	25	39	45	49	64	70	29	57	22
Pressão Sistólica (y)	109	122	143	132	199	185	199	130	175	118

b) Idade (anos) e vocabulário (número de palavras)

Idade (x)	1	2	6	4	3	6	3	5	2	4	5
Vocabulário (y)	3	440	2500	1500	1200	2600	1100	200	500	1525	2100

c) Horas de estudo e notas nos testes

Horas de estudo (x)	0	1	2	4	4	5	5	5	6	6	7	7	8
Notas nos testes (y)	40	41	51	48	64	69	73	75	68	93	84	90	95

d) Horas on-line e notas nos testes

Horas on line (x)	0	1	2	3	3	5	5	5	6	7	7	10
Notas nos testes (y)	96	85	82	74	95	68	76	84	58	65	75	50

e) Altitude (mil pés) e velocidade do som (pés por segundo)

Altitude (x)	0	5	10	15	20	25	30	35	40	45	50
V. do som (y)	1116,3	1096,9	1077,3	1057,2	1036,8	1015,8	994,5	969	967,7	967,7	967,7

f) Peso do veículo (libras) e variabilidade da distância de frenagem (pés) em superfície molhada a uma determinada velocidade.

Peso (x)	5890	5340	6500	4800	5940	5600	5100	5580
Variabilidade frenagem (y)	2,92	2,4	4,09	1,72	2,88	2,53	2,32	2,78

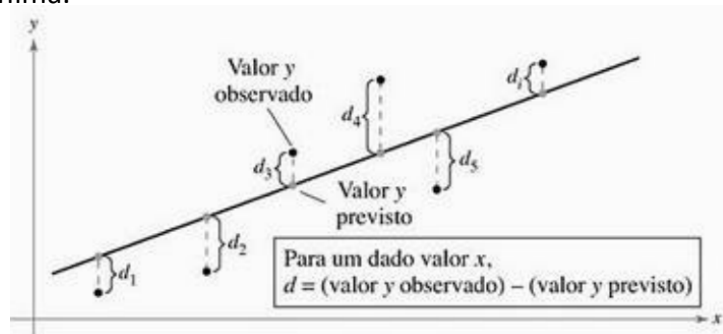
g) Número de Crimes (milhões) e número de prisões (milhões) em 14 anos.

Crimes (x)	1,66	1,65	1,6	1,55	1,44	1,4	1,32	1,23	1,22	1,23	1,22	1,18	1,16	1,19
Prisões (y)	0,72	0,72	0,78	0,8	0,73	0,72	0,68	0,64	0,63	0,63	0,62	0,6	0,59	0,6

13. Regressão Linear - Linhas de Regressão

É a linha que melhor modela os dados, ou “linha de melhor ajuste”, para a qual a soma dos quadrados dos resíduos é mínima.

- Resíduo é a diferença entre o valor observado e o valor previsto pela equação da linha de regressão.
- Equação da reta:
 $y = mx + b$



- A inclinação (m) e a interseção no eixo y (b) são dadas pelas equações:

$$m = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2}$$

$$b = \bar{y} - m\bar{x} = \frac{\sum y}{n} - m \frac{\sum x}{n}$$

n é o número de pares de valores xy

- No Microsoft Excel a linha de regressão é fornecida pela Linha de Tendência linear.
 - ✓ Selecione os dados e Insira o gráfico de dispersão xy
 - ✓ Clique com o botão direito sobre um ponto no gráfico e selecione “Adicionar Linha de tendência”
 - ✓ Escolha o tipo Linear de linha de tendência
 - ✓ Habilite Exibir Equação no gráfico
 - A equação da linha de regressão é exibida no gráfico;
 - ✓ Habilite Exibir valor de R-quadrado no gráfico
 - O valor de R^2 é exibido no gráfico

15. O TESTES QUI QUADRADO

1. Experimento Polinomial

É um experimento de probabilidade que consiste em um número fixo de testes nos quais há mais de dois resultados possíveis para cada teste independente. (obs: um experimento binomial tem apenas dois resultados possíveis).

- A probabilidade para cada resultado é fixa
- Cada resultado é classificado em categorias

2. Teste Qui-Quadrado

Um teste de ajuste qui-quadrado é usado para testar se uma distribuição de frequência se encaixa em uma distribuição esperada. Utiliza-se a tabela da distribuição χ^2 .

Para que o teste possa ser utilizado, duas condições têm que ser satisfeitas:

- As observações devem ser obtidas utilizando uma amostra aleatória;
- Cada frequência esperada deve ser igual ou maior a 5;
 - ✓ Se uma frequência esperada para uma categoria for menor que 5, pode-se agrupar essa categoria com outra para atender a essa exigência.

3. Exemplo

Uma matriz de uma estação de rádio afirma que a distribuição de preferências musicais dos ouvintes na região de uma determinada filial é:

Estilo	Clássica	Country	Gospel	Flash back	Pop	rock
% ouvintes	4%	36%	11%	2%	18%	29%

A estação local faz uma pesquisa para verificar se a distribuição alegada pela matriz corresponde à distribuição local considerando um alfa igual a 0,01. Pesquisa aleatoriamente 500 ouvintes e encontra o seguinte resultado (frequência observada **O**):

Estilo	Clássica	Country	Gospel	Flash back	Pop	rock
n. ouvintes	8	210	72	10	75	125

- A hipótese nula (H_0) é: a distribuição de frequência de preferências musicais na região da filial é a alegada pela matriz;
- A hipótese alternativa (H_a) é: a distribuição de frequência de preferências musicais na região da filial difere daquela alegada pela matriz;

Se a hipótese nula fosse verdadeira, o número de ouvintes em cada estilo musical encontrados na pesquisa efetuada na região da filial deveria ser igual à porcentagem alegada pela matriz multiplicada pelo número de ouvintes pesquisados (frequência esperada **E**):

Estilo	Clássica	Country	Gospel	Flash back	Pop	rock
n. ouvintes	20	180	55	10	90	145

A estatística de teste qui-quadrado é dada pela equação:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Onde:

O é a frequência observada de cada categoria e
E é a frequência esperada de cada categoria

- Quando as frequências observadas são próximas às frequências esperadas, as diferenças entre O e E serão pequenas e assim a estatística de teste qui-quadrado também será pequena (próxima de zero). Dessa forma, a hipótese nula não pode ser rejeitada.
 - Quando as frequências observadas são discrepantes das frequências esperadas, as diferenças entre O e E serão maiores e assim a estatística de teste qui-quadrado será maior. Se o valor for maior que um determinado valor limite, rejeita-se a hipótese nula, assumindo a hipótese alternativa.
- ✓ O teste qui-quadrado é um teste unicaudal à direita.

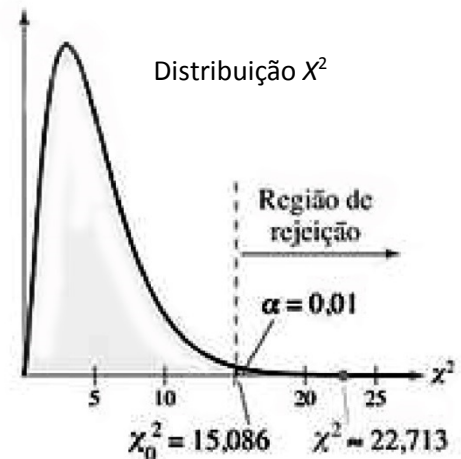
ESTILO	O	E
Clássica	8	20
Country	210	180
Gospel	72	55
Flash Back	10	10
Pop	75	90
Rock	125	145
TOTAL	500	500

O valor da estatística qui-quadrado para o exemplo acima é:

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(8-20)^2}{20} + \frac{(210-180)^2}{180} + \frac{(72-55)^2}{55} + \frac{(10-10)^2}{10} + \frac{(75-90)^2}{90} + \frac{(125-145)^2}{145} \approx 22,713$$

Considerando um alfa de 0,01 encontramos na tabela da distribuição χ^2 , com $K-1$ graus de liberdade ($6-1 = 5$), o valor crítico de 15,086.

Como o valor calculado é maior que o valor limite (o valor calculado está na região de rejeição da hipótese nula) rejeitamos a hipótese nula e assumimos a hipótese alternativa (a diferença de proporções é estatisticamente significativa).



4. Exemplo

Um instituto de pesquisa afirma que, em estudo efetuado no Rio de Janeiro, a opinião dos homens sobre os motivos para poupar eram: aposentadoria (44%), educação superior dos filhos (40%) e não sabem (16%). Para testar se essa afirmação é válida na cidade de São Paulo, o instituto refaz a pesquisa selecionando aleatoriamente 400 homens nesta cidade e encontra a seguinte distribuição: aposentadoria (186), educação superior dos filhos (143) e não sabem (71). Com alfa igual a 0,05 pode-se afirmar que as distribuições são diferentes?

- H_0 : a distribuição em São Paulo é igual à distribuição no Rio de Janeiro;
- H_a : a distribuição em São Paulo é diferente da distribuição no Rio de Janeiro.
- Como há três categorias, o número de graus de liberdade é $3-1=2$
- Consultando a tabela χ^2 encontramos, para $\alpha = 0,05$ e g.l. = 2 o valor 5,991;

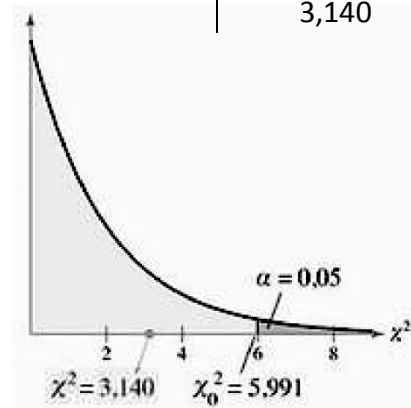
Motivo	Freq.Observada O	Freq.Esperada E
Aposentadoria	186	$400 \times 0,44 = 176$
Educação filhos	143	$400 \times 0,40 = 160$
Não sabe	71	$400 \times 0,16 = 64$

O cálculo da estatística χ^2 é mostrado na tabela abaixo:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
186	176	10	100	0,568182
143	160	-17	289	1,80625
71	64	7	49	0,765625
				3,140

Observa-se que o valor calculado não está na região de rejeição da hipótese nula, portanto com nível de significância de 5% não há evidência estatística de que as distribuições são diferentes e a hipótese nula é aceita: a distribuição de opiniões no Rio de Janeiro é igual à distribuição de opiniões em São Paulo.



5. Exercícios

- a) No caso do exemplo 4 acima, a distribuição de opiniões das mulheres no Rio de Janeiro foi: aposentadoria (41%), educação superior dos filhos (46%) e não sabem (13%). A pesquisa em São Paulo com 300 mulheres resultou em: aposentadoria (129), educação superior dos filhos (149) e não sabem (22). Com alfa igual a 0,05 pode-se afirmar que as distribuições são diferentes?
- b) Um sociólogo afirma que a distribuição de idades dos moradores de certa cidade é diferente do que era 10 anos antes. Você seleciona aleatoriamente 400 moradores e registra a idade de cada um deles. Os resultados são registrados na tabela abaixo. Pode-se afirmar, com alfa igual a 5%, que a distribuição de idades foi alterada nesses 10 anos? E com alfa igual a 1%?

Idade	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70+
Anterior	16%	20%	8%	14%	15%	12%	10%	5%
Pesquisa	76	84	30	60	54	40	42	14

- c) O fabricante de chocolate “confeti” afirma que o número de unidades de cores diferentes é uniformemente distribuído em cada embalagem. Você seleciona aleatoriamente uma embalagem com 500 unidades e verifica a distribuição abaixo. Utilizando um alfa de 10% pode-se afirmar que as cores são uniformemente distribuídas nas embalagens?

Cor	Marrom	Amarelo	Vermelho	Azul	Laranja	Verde
Frequência	80	95	88	83	76	78

6. Tabela de Contingência e Teste de Independência

O teste qui-quadrado para independência é utilizado para testar a independência de duas variáveis. Pode determinar se a ocorrência de uma das variáveis afeta a probabilidade de ocorrência da outra.

- Eventos independentes são aqueles que a ocorrência de um evento não afeta a probabilidade de ocorrência do outro.

Uma tabela de contingência $r \times c$ (row x column) mostra as frequências observadas para duas variáveis. As frequências observadas são arranjadas nas fileiras r e colunas c . A interseção de uma fileira e uma coluna é chamada célula.

A frequência esperada para uma determinada célula E em uma tabela de contingência é:

$$\text{Frequência esperada } E_{rc} = \frac{(\text{soma das fileiras } r) * (\text{soma das colunas } c)}{\text{tamanho da amostra}}$$

Para que o teste possa ser aplicado, duas condições têm que ser observadas:

- As frequências observadas devem ser obtidas utilizando-se uma amostra aleatória;
- Cada frequência esperada deve ser maior ou igual a 5.

7. Exemplo:

O gerente da academia universitária quer determinar se o número de dias por semana que os alunos frequentam a academia está relacionado ao gênero. Uma amostra aleatória de 275 alunos é selecionada e os resultados são classificados na tabela abaixo. Com alfa de 0,05, há evidência suficiente para concluir que o número de dias na semana que se frequenta a academia está relacionado ao gênero do aluno?

Gênero	0 a 1	2 a 3	4 a 5	6 a 7	Total
Homens	40	53	26	6	125
Mulheres	34	68	37	11	150
Total	74	121	63	17	275

- H_0 : O número de dias frequentando academia é independente do gênero
- Há: o número de dias frequentando academia depende do gênero
- Cálculo das frequências esperadas:

E_{rc}	0 a 1	2 a 3	4 a 5	6 a 7	Total
Homens	$\frac{125 * 74}{275}$	$\frac{125 * 121}{275}$	$\frac{125 * 63}{275}$	$\frac{125 * 17}{275}$	125
Mulheres	$\frac{150 * 74}{275}$	$\frac{150 * 121}{275}$	$\frac{150 * 63}{275}$	$\frac{150 * 17}{275}$	150
Total	74	121	63	17	275

E_{rc}	0 a 1	2 a 3	4 a 5	6 a 7
Homens	33,64	55,00	28,64	7,73
Mulheres	40,36	66,00	34,36	9,27

Teste do qui-quadrado (χ^2) para a independência:

- O : frequência observada;
- E : frequência esperada.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O número de graus de liberdade é dado por

- r : quantidade de linhas da tabela;
- c : quantidade de colunas da tabela.

$$(r - 1) * (c - 1)$$

A hipótese nula e hipótese alternativa

- H_0 : as variáveis são independentes;
- H_a : as variáveis são dependentes.

Se a estatística de teste for menor que o valor crítico, aceita-se a hipótese nula.

O cálculo da estatística de teste:

χ^2	0 a 1	2 a 3	4 a 5	6 a 7	Total
Homens	$\frac{(40 - 33,64)^2}{33,64}$	$\frac{(53 - 55,00)^2}{55,00}$	$\frac{(26 - 28,64)^2}{28,64}$	$\frac{(6 - 7,73)^2}{7,73}$	
Mulheres	$\frac{(34 - 40,36)^2}{40,36}$	$\frac{(68 - 66,00)^2}{66,00}$	$\frac{(37 - 34,36)^2}{34,36}$	$\frac{(11 - 9,27)^2}{9,27}$	
				χ^2:	

χ^2	0 a 1	2 a 3	4 a 5	6 a 7	Total
Homens	1,203931	0,072727	0,242713	0,386096	1,905468
Mulheres	1,003276	0,060606	0,202261	0,321747	1,587890
				χ^2:	3,493357

O número de graus de liberdade: $(2-1)*(4-1) = 3$

Consultando-se a tabela χ^2 encontra-se, para $\alpha = 0,05$ e g.l. = 3 o valor $\chi^2 = 7,815$.

Como a estatística de teste (3,493) é menor que o valor limite (7,815), não se pode rejeitar a hipótese nula com um α de 0,05.

8. Exercício:

Um pesquisador quer determinar se o número de minutos que adultos ficam on-line está relacionado ao seu gênero. Uma amostra aleatória de 450 adultos é selecionada e os resultados são classificados segundo a tabela abaixo. Com $\alpha = 0,05$, há evidência suficiente para concluir que o número de minutos que adultos ficam on-line está relacionado ao seu gênero?

Gênero	0 a 15	15 a 30	30 a 45	45 a 60	60 ou +	Total
Homens	19	36	75	90	55	275
Mulheres	21	72	45	19	18	175
Total	40	108	120	109	73	450

16. NOÇÕES DE ESTATÍSTICA NÃO PARAMÉTRICA**1. Testes de Hipótese Não Paramétricos**

Os testes paramétricos vistos anteriormente exigem, para sua validade, alguns requisitos referentes à população, como:

- A população tenha uma distribuição normal,
- As variâncias sejam equivalentes

Os testes não paramétricos não requerem condições específicas sobre o formato das populações ou o valor de nenhum parâmetro da população.

Os testes paramétricos são mais fáceis de se realizar, mas em geral, são menos eficientes. Nos testes não paramétricos, são necessárias evidências mais fortes para se rejeitar a hipótese nula.

2. Teste dos Sinais (Sign Test)

É um teste não paramétrico que pode ser utilizado para testar uma mediana de população contra um valor hipotético K .

- Muitas vezes, em testes não paramétricos, se testa a mediana em lugar da média.

3. Teste de Postos Sinalizados de Wilcoxon (Signed Rank Test)

O teste de postos sinalizados de Wilcoxon é um teste não paramétrico que pode ser utilizado para determinar se duas amostras dependentes foram selecionadas de populações que possuem a mesma distribuição.

4. Teste de Soma de Postos de Wilcoxon (Rank Sum Test)

O teste de soma de postos de Wilcoxon é um teste não paramétrico que pode ser utilizado para determinar se duas amostras independentes foram selecionadas de populações que possuem a mesma distribuição.

5. Teste de Kruskal-Wallis

O teste de Kruskal-Wallis é um teste não paramétrico que pode ser utilizado para determinar se três ou mais amostras independentes foram selecionadas de populações que possuem a mesma distribuição.

6. Coeficiente de Correlação de Postos de Spearman (Rank Correlation Coefficient)

O coeficiente de correlação de postos de Spearman (r_s) é uma medida da força da relação entre duas variáveis.

- Pode ser utilizado para descrever relações lineares e não lineares entre dados;
- Não requer que as populações de cada variável sejam normalmente distribuídas.

7. Teste de Corridas para Aleatoriedade (Runs Test for Randomness)

O teste de corridas para aleatoriedade é um teste não paramétrico que pode ser utilizado para determinar se uma sequência de dados é aleatória.

8. Uso dos Testes Não Paramétricos

Antes de executar um teste de hipóteses paramétrico é preciso garantir que certas condições sobre a população sejam preenchidas. Por exemplo, antes de utilizar um teste t, é necessário verificar se a população é distribuída normalmente. Para executar um teste não paramétrico não é necessário verificar nenhuma informação em particular sobre a população ou populações que estão sendo testadas. Outra vantagem dos testes não paramétricos é que eles são mais fáceis de serem executados do que seus equivalentes paramétricos. Isso significa que eles são mais fáceis de entender e mais rápidos para serem utilizados. Testes não paramétricos frequentemente podem ser utilizados quando os dados estão no nível nominal ou ordinal.

9. Abuso dos Testes Não Paramétricos

Evidência insuficiente: evidência mais forte é necessária para se rejeitar a hipótese nula em um teste não paramétrico do que em um teste paramétrico correspondente (é mais fácil rejeitar a hipótese nula em um teste paramétrico). Isto é, quando se está tentando apoiar uma afirmação representada pela hipótese nula, pode-se precisar de uma amostra maior quando se executa um teste não paramétrico.

Utilizar teste não apropriado: Em geral, quando informações sobre as populações (como a condição de normalidade) são conhecidas, é mais eficiente utilizar um teste paramétrico. No entanto, se as informações sobre a população não são conhecidas, testes não paramétricos são úteis.

CHAPTER 2

$$\text{Class Width} = \frac{\text{Range of data}}{\text{Number of classes}}$$

(round up to next convenient number)

$$\text{Midpoint} = \frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}$$

$$\text{Relative Frequency} = \frac{\text{Class frequency}}{\text{Sample size}} = \frac{f}{n}$$

$$\text{Population Mean: } \mu = \frac{\sum x}{N}$$

$$\text{Sample Mean: } \bar{x} = \frac{\sum x}{n}$$

$$\text{Weighted Mean: } \bar{x} = \frac{\sum (x \cdot w)}{\sum w}$$

$$\text{Mean of a Frequency Distribution: } \bar{x} = \frac{\sum (x \cdot f)}{n}$$

$$\text{Range} = (\text{Maximum entry}) - (\text{Minimum entry})$$

$$\text{Population Variance: } \sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Population Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

$$\text{Sample Variance: } s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$\text{Sample Standard Deviation: } s = \sqrt{s^2} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

Empirical Rule (or 68-95-99.7 Rule) For data with a (symmetric) bell-shaped distribution:

1. About 68% of the data lies between $\mu - \sigma$ and $\mu + \sigma$.
2. About 95% of the data lies between $\mu - 2\sigma$ and $\mu + 2\sigma$.
3. About 99.7% of the data lies between $\mu - 3\sigma$ and $\mu + 3\sigma$.

Chebyshev's Theorem The portion of any data set lying within k standard deviations ($k > 1$) of the mean is at

$$\text{least } 1 - \frac{1}{k^2}.$$

Sample Standard Deviation of a Frequency Distribution:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2 f}{n - 1}}$$

$$\text{Standard Score: } z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

CHAPTER 3

Classical (or Theoretical) Probability:

$$P(E) = \frac{\text{Number of outcomes in event } E}{\text{Total number of outcomes in sample space}}$$

Empirical (or Statistical) Probability:

$$P(E) = \frac{\text{Frequency of event } E}{\text{Total frequency}} = \frac{f}{n}$$

Probability of a Complement: $P(E') = 1 - P(E)$

Probability of occurrence of both events A and B :

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

$P(A \text{ and } B) = P(A) \cdot P(B)$ if A and B are independent

Probability of occurrence of either A or B or both:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$P(A \text{ or } B) = P(A) + P(B)$ if A and B are mutually exclusive

Permutations of n objects taken r at a time:

$${}_nP_r = \frac{n!}{(n-r)!}, \text{ where } r \leq n$$

Distinguishable Permutations: n_1 alike, n_2 alike, ..., n_k alike:

$$\frac{n!}{n_1! \cdot n_2! \cdot n_3! \cdots n_k!},$$

where $n_1 + n_2 + n_3 + \cdots + n_k = n$

Combination of n objects taken r at a time:

$${}_nC_r = \frac{n!}{(n-r)!r!}$$

CHAPTER 4

Mean of a Discrete Random Variable: $\mu = \sum xP(x)$

Variance of a Discrete Random Variable:

$$\sigma^2 = \sum (x - \mu)^2 P(x)$$

Standard Deviation of a Discrete Random Variable:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x - \mu)^2 P(x)}$$

Expected Value: $E(x) = \mu = \sum xP(x)$

Binomial Probability of x successes in n trials:

$$P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x}$$

Population Parameters of a Binomial Distribution:

$$\text{Mean: } \mu = np \quad \text{Variance: } \sigma^2 = npq$$

$$\text{Standard Deviation: } \sigma = \sqrt{npq}$$

Geometric Distribution: The probability that the first success will occur on trial number x is $P(x) = p(q)^{x-1}$, where $q = 1 - p$.

Poisson Distribution: The probability of exactly x

occurrences in an interval is $P(x) = \frac{\mu^x e^{-\mu}}{x!}$, where

$e \approx 2.71828$ and μ is the mean number of occurrences per interval unit.

CHAPTER 5

Standard Score, or z -Score:

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}$$

Transforming a z -Score to an x -Value: $x = \mu + z\sigma$

Central Limit Theorem ($n \geq 30$ or population is normally distributed):

$$\text{Mean of the Sampling Distribution: } \mu_{\bar{x}} = \mu$$

$$\text{Variance of the Sampling Distribution: } \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Standard Deviation of the Sampling Distribution (Standard Error):

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$z\text{-Score} = \frac{\text{Value} - \text{Mean}}{\text{Standard Error}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

CHAPTER 6

c -Confidence Interval for μ : $\bar{x} - E < \mu < \bar{x} + E$,

where $E = z_c \frac{\sigma}{\sqrt{n}}$ if σ is known and the population is

normally distributed or $n \geq 30$, or $E = t_c \frac{s}{\sqrt{n}}$ if the population is normally or approximately normally distributed, σ is unknown, and $n < 30$

Minimum Sample Size to Estimate μ : $n = \left(\frac{z_c \sigma}{E} \right)^2$

Point Estimate for p , the population proportion of

$$\text{successes: } \hat{p} = \frac{x}{n}$$

c -Confidence Interval for Population Proportion p (when $np \geq 5$ and $nq \geq 5$): $\hat{p} - E < p < \hat{p} + E$, where

$$E = z_c \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Minimum Sample Size to Estimate p : $n = \hat{p}\hat{q} \left(\frac{z_c}{E} \right)^2$

c -Confidence Interval for Population Variance σ^2 :

$$\frac{(n-1)s^2}{\chi_R^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_L^2}$$

c -Confidence Interval for Population Standard Deviation σ :

$$\sqrt{\frac{(n-1)s^2}{\chi_R^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_L^2}}$$

CHAPTER 7

z -Test for a Mean μ : $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$, for σ known with a normal population, or for $n \geq 30$

t -Test for a Mean μ : $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$, for σ unknown,

population is normal or nearly normal, and $n < 30$.
(d.f. = $n - 1$)

z -Test for a Proportion p (when $np \geq 5$ and $nq \geq 5$):

$$z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

Chi-Square Test for a Variance σ^2 or Standard Deviation σ :

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad (\text{d.f.} = n - 1)$$

CHAPTER 8

Two-Sample z -Test for the Difference Between Means (Independent samples; n_1 and $n_2 \geq 30$ or normally distributed populations):

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}},$$

$$\text{where } \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Two-Sample t -Test for the Difference Between Means (Independent samples from normally distributed populations, n_1 or $n_2 < 30$):

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

If population variances are equal, d.f. = $n_1 + n_2 - 2$ and

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

If population variances are not equal, d.f. is the

smaller of $n_1 - 1$ or $n_2 - 1$ and $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$

t -Test for the Difference Between Means (Dependent samples):

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}, \text{ where } \bar{d} = \frac{\sum d}{n}, s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n - 1}}$$

and d.f. = $n - 1$

Two-Sample z -Test for the Difference Between Proportions ($n_1\bar{p}$, $n_1\bar{q}$, $n_2\bar{p}$, and $n_2\bar{q}$ must be at least 5):

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ where } \bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

and $\bar{q} = 1 - \bar{p}$.

CHAPTER 9

Correlation Coefficient:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}}$$

t -Test for the Correlation Coefficient:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (\text{d.f.} = n - 2)$$

Equation of a Regression Line: $\hat{y} = mx + b$,

$$\text{where } m = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \text{ and}$$

$$b = \bar{y} - m\bar{x} = \frac{\sum y}{n} - m\frac{\sum x}{n}$$

Coefficient of Determination:

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Standard Error of Estimate: $s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$

c -Prediction Interval for y : $\hat{y} - E < y < \hat{y} + E$, where

$$E = t_{c s_e} \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n\sum x^2 - (\sum x)^2}} \quad (\text{d.f.} = n - 2)$$

CHAPTER 10

Chi-Square: $\chi^2 = \sum \frac{(O - E)^2}{E}$

Goodness-of-Fit Test: d.f. = $k - 1$

Test of Independence:

d.f. = (no. of rows - 1)(no. of columns - 1)

Two-Sample F -Test for Variances: $F = \frac{s_1^2}{s_2^2}$, where

$s_1^2 \geq s_2^2$, d.f._N = $n_1 - 1$, and d.f._D = $n_2 - 1$

One-Way Analysis of Variance Test:

$$F = \frac{MS_B}{MS_W}, \text{ where } MS_B = \frac{SS_B}{k - 1} = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k - 1}$$

$$\text{and } MS_W = \frac{SS_W}{N - k} = \frac{\sum (n_i - 1)s_i^2}{N - k}$$

(d.f._N = $k - 1$, d.f._D = $N - k$)

CHAPTER 11

Test Statistic for Sign Test:

When $n \leq 25$, the test statistic is the smaller number of + or - signs.

When $n > 25$, $z = \frac{(x + 0.5) - 0.5n}{\frac{\sqrt{n}}{2}}$, where x is the

smaller number of + or - signs and n is the total number of + and - signs.

Test Statistic for Wilcoxon Rank Sum Test:

$$z = \frac{R - \mu_R}{\sigma_R}, \text{ where } R = \text{sum of the ranks for the}$$

smaller sample, $\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$,

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}, \text{ and } n_1 \leq n_2$$

Test Statistic for the Kruskal-Wallis Test:

Given three or more independent samples, the test statistic for the Kruskal-Wallis test is

$$H = \frac{12}{N(N+1)} \left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \dots + \frac{R_k^2}{n_k} \right) - 3(N+1). \quad (\text{d.f.} = k - 1)$$

Spearman Rank Correlation Coefficient:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Test Statistic for the Runs Test:

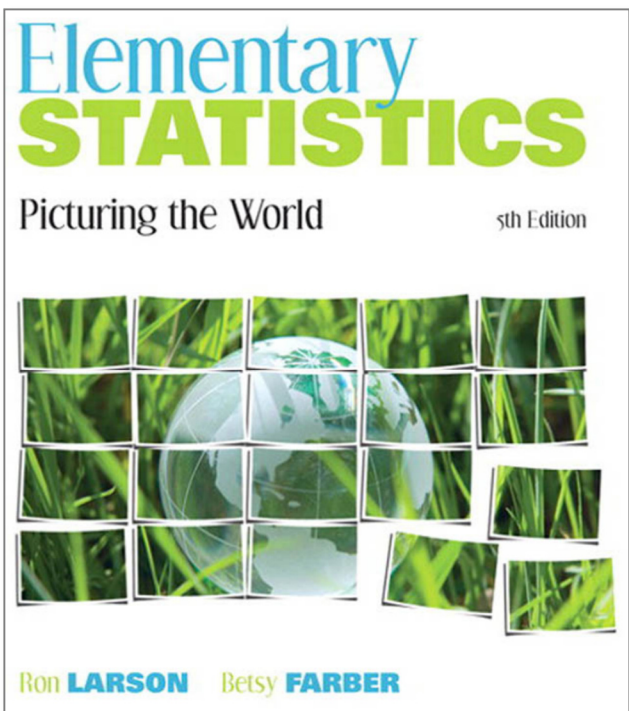
When $n_1 \leq 20$ and $n_2 \leq 20$, the test statistic is G , the number of runs.

When $n_1 > 20$ or $n_2 > 20$, the test statistic is

$$z = \frac{G - \mu_G}{\sigma_G}, \text{ where } G = \text{number of runs,}$$

$$\mu_G = \frac{2n_1 n_2}{n_1 + n_2} + 1, \text{ and}$$

$$\sigma_G = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}.$$

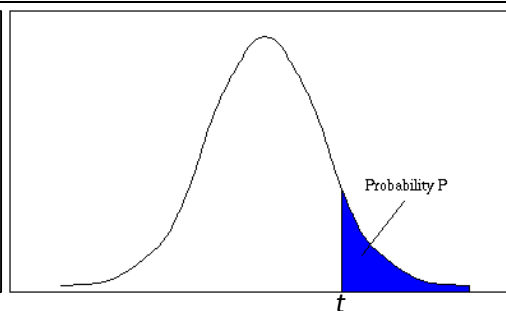
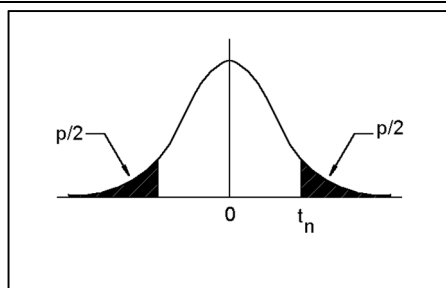


DISTRIBUIÇÃO Z (-)

Table A-1 **The Z-Table**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0003	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

DISTRIBUIÇÃO T Valores críticos



Probabilidade P

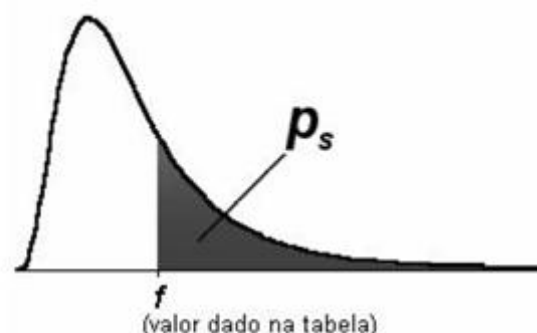
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.663.	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.15	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.295	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
inf.	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291


TABLE of CRITICAL VALUES for STUDENT'S t DISTRIBUTIONS

 Column headings denote probabilities (α) **above** tabulated values.

d.f.	0.40	0.25	0.10	0.05	0.04	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	7.916	12.706	15.894	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	3.320	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	2.605	3.182	3.482	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.333	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.191	2.571	2.757	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.104	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.046	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.004	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	1.973	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	1.948	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	1.928	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	1.912	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	1.899	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	1.887	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	1.878	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	1.869	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	1.862	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	1.855	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	1.850	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	1.844	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	1.840	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	1.835	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	1.832	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	1.828	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	1.825	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	1.822	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	1.819	2.052	2.158	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	1.817	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	1.814	2.045	2.150	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	1.812	2.042	2.147	2.457	2.750	3.030	3.385	3.646
31	0.256	0.682	1.309	1.696	1.810	2.040	2.144	2.453	2.744	3.022	3.375	3.633
32	0.255	0.682	1.309	1.694	1.808	2.037	2.141	2.449	2.738	3.015	3.365	3.622
33	0.255	0.682	1.308	1.692	1.806	2.035	2.138	2.445	2.733	3.008	3.356	3.611
34	0.255	0.682	1.307	1.691	1.805	2.032	2.136	2.441	2.728	3.002	3.348	3.601
35	0.255	0.682	1.306	1.690	1.803	2.030	2.133	2.438	2.724	2.996	3.340	3.591
36	0.255	0.681	1.306	1.688	1.802	2.028	2.131	2.434	2.719	2.990	3.333	3.582
37	0.255	0.681	1.305	1.687	1.800	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.255	0.681	1.304	1.686	1.799	2.024	2.127	2.429	2.712	2.980	3.319	3.566
39	0.255	0.681	1.304	1.685	1.798	2.023	2.125	2.426	2.708	2.976	3.313	3.558
40	0.255	0.681	1.303	1.684	1.796	2.021	2.123	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	1.781	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.254	0.678	1.292	1.664	1.773	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.254	0.677	1.290	1.660	1.769	1.984	2.081	2.364	2.626	2.871	3.174	3.390
120	0.254	0.677	1.289	1.658	1.766	1.980	2.076	2.358	2.617	2.860	3.160	3.373
140	0.254	0.676	1.288	1.656	1.763	1.977	2.073	2.353	2.611	2.852	3.149	3.361
160	0.254	0.676	1.287	1.654	1.762	1.975	2.071	2.350	2.607	2.847	3.142	3.352
180	0.254	0.676	1.286	1.653	1.761	1.973	2.069	2.347	2.603	2.842	3.136	3.345
200	0.254	0.676	1.286	1.653	1.760	1.972	2.067	2.345	2.601	2.838	3.131	3.340
250	0.254	0.675	1.285	1.651	1.758	1.969	2.065	2.341	2.596	2.832	3.123	3.330
inf	0.253	0.674	1.282	1.645	1.751	1.960	2.054	2.326	2.576	2.807	3.090	3.290

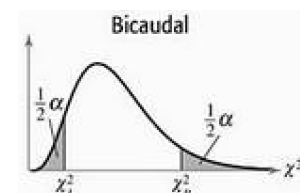
DISTRIBUIÇÃO F



		Degrees of freedom in numerator (df1)											
		1	2	3	4	5	6	7	8	12	24	1000	
Degrees of freedom in denominator (df2)	1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.30
		0.050	181.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.2
		0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	976.7	997.3	1017.8
		0.010	4052	4999	5404	5624	5764	5859	5928	5981	6107	6234	6383
		0.001	405312	499725	540257	562668	576496	586033	593185	597954	610352	623703	638101
	2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
		0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.49
		0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
		0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.42	99.46	99.50
		0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
	3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
		0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
		0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.91
		0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.14
		0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	128.32	125.93	123.52
	4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
		0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
		0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
		0.010	21.20	18.00	16.89	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.47
		0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	47.41	45.77	44.09
	5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.11
		0.050	6.81	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
		0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
		0.010	18.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.03
		0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	26.42	25.13	23.82
	6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
		0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
		0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.86
		0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.89
		0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.77
	7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
		0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
		0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.15
		0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.66
		0.001	29.25	21.89	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.72
	8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.30
		0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
		0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.68
		0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.87
		0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.30	9.36
	9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
		0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
		0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.34
		0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.32
		0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.84



Distribuição Qui-Quadrado



$\nu \backslash \alpha$	0,995	0,99	0,975	0,95	0,90	0,75	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
1	0,0004	0,002	0,001	0,004	0,016	0,102	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991	7,378	9,210	10,597	13,816
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815	9,348	11,345	12,838	16,266
4	0,207	0,297	0,484	0,711	1,064	1,923	3,357	5,385	7,779	9,488	11,143	13,277	14,860	18,467
5	0,412	0,554	0,831	1,145	1,610	2,675	4,351	6,626	9,236	11,071	12,833	15,086	16,750	20,515
6	0,676	0,872	1,237	1,635	2,204	3,455	5,348	7,841	10,645	12,592	14,449	16,812	18,548	22,458
7	0,989	1,239	1,690	2,167	2,833	4,255	6,346	9,037	12,017	14,067	16,013	18,475	20,278	24,322
8	1,344	1,646	2,180	2,733	3,490	5,071	7,344	10,219	13,362	15,507	17,535	20,090	21,955	26,125
9	1,735	2,088	2,700	3,325	4,168	5,899	8,343	11,389	14,684	16,919	19,023	21,666	23,589	27,877
10	2,156	2,558	3,247	3,940	4,865	6,737	9,342	12,549	15,987	18,307	20,483	23,209	25,188	29,588
11	2,603	3,053	3,816	4,575	5,578	7,584	10,341	13,701	17,275	19,675	21,920	24,725	26,757	31,264
12	3,074	3,571	4,404	5,226	6,304	8,438	11,340	14,845	18,549	21,026	23,337	26,217	28,299	32,909
13	3,565	4,107	5,009	5,892	7,042	9,299	12,340	15,984	19,812	22,362	24,736	27,688	29,819	34,528
14	4,075	4,660	5,629	6,571	7,790	10,165	13,339	17,117	21,064	23,685	26,119	29,141	31,319	36,123
15	4,601	5,229	6,262	7,261	8,547	11,036	14,339	18,245	22,307	24,996	27,488	30,578	32,801	37,697
16	5,142	5,812	6,908	7,962	9,312	11,912	15,338	19,369	23,542	26,296	28,845	32,000	34,267	39,252
17	5,697	6,408	7,564	8,672	10,085	12,792	16,338	20,489	24,769	27,587	30,191	33,409	35,718	40,790
18	6,265	7,015	8,231	9,390	10,865	13,675	17,338	21,605	25,989	28,869	31,526	34,805	37,156	43,312
19	6,844	7,633	8,907	10,117	11,651	14,562	18,338	22,718	27,204	30,144	32,852	36,191	38,582	43,820
20	7,434	8,260	9,591	10,851	12,443	15,452	19,337	23,828	28,412	31,410	34,170	37,566	39,997	45,315
21	8,034	8,897	10,283	11,591	13,240	16,344	20,337	24,935	29,615	32,671	35,479	38,932	41,401	46,797
22	8,643	9,542	10,982	12,338	14,042	17,240	21,337	26,039	30,813	33,924	36,781	40,289	42,796	48,268
23	9,260	10,196	11,689	13,091	14,848	18,137	22,337	27,141	32,007	35,172	38,076	41,638	44,181	49,728
24	9,886	10,856	12,401	13,848	15,659	19,037	22,337	28,241	33,196	36,415	39,364	42,980	45,559	51,179
25	10,520	11,524	13,120	14,611	16,473	19,939	24,337	29,339	34,382	37,652	40,646	44,314	46,928	52,620
26	11,160	12,198	13,844	15,379	17,292	20,843	25,336	30,434	35,563	38,885	41,923	45,642	48,290	54,052
27	11,808	12,879	14,573	16,151	18,114	21,749	26,336	31,528	36,741	40,113	43,194	46,963	49,645	55,476
28	12,461	13,565	15,308	16,928	18,939	22,657	27,336	32,620	37,916	41,337	44,461	48,278	50,993	56,892
29	13,121	14,257	16,047	17,708	19,768	23,567	28,336	33,711	39,087	42,557	45,722	49,588	52,336	58,302
30	13,787	14,954	16,791	18,493	20,599	24,478	29,336	34,800	40,256	43,773	46,979	50,892	53,672	59,703
31	14,458	15,655	17,539	19,281	21,434	25,390	30,336	35,887	41,422	44,985	48,232	52,191	55,003	61,098
32	15,134	16,362	18,291	20,072	22,271	26,304	31,336	36,973	42,585	46,194	49,480	53,486	56,328	62,487
33	15,815	17,074	19,047	20,867	23,110	27,219	32,336	38,058	43,745	47,400	50,725	54,776	57,648	63,870
34	16,501	17,789	19,806	21,664	23,952	28,136	33,336	39,141	44,903	48,602	51,966	56,061	58,964	65,247
35	17,192	18,509	20,569	22,465	24,797	29,054	34,336	40,223	46,059	49,802	53,203	57,342	60,275	66,619
36	17,887	19,233	21,336	23,269	25,643	29,973	35,336	41,304	47,212	50,998	54,437	58,619	61,581	67,985
37	18,586	19,960	22,106	24,075	26,492	30,893	36,336	42,383	48,363	52,192	55,668	59,892	62,883	69,346
38	19,289	20,691	22,878	24,884	27,343	31,815	37,335	43,462	49,513	53,384	56,896	61,162	64,181	70,701
39	19,996	21,426	23,654	25,695	28,196	32,737	38,335	44,539	50,660	54,572	58,120	62,428	65,476	72,055
40	20,707	22,164	24,433	26,509	29,051	33,660	39,335	45,616	51,805	55,758	59,342	63,691	66,766	73,402
41	21,421	22,906	25,215	27,326	29,907	34,585	40,335	46,692	52,949	56,942	60,561	64,950	68,053	74,745
42	22,138	23,650	25,999	28,144	30,765	35,510	41,335	47,766	54,090	58,124	61,777	66,206	69,336	76,084
43	22,859	24,398	26,785	28,965	31,625	36,436	42,335	48,840	55,230	59,304	62,990	67,459	70,616	77,419
44	23,584	25,148	27,575	29,787	32,487	37,363	43,335	49,913	56,369	60,481	64,201	68,710	71,893	78,750
45	24,311	25,901	28,366	30,612	33,350	38,291	44,335	50,985	57,505	61,656	65,410	69,957	73,166	80,077
50	27,991	29,707	32,357	34,764	37,689	42,942	49,335	56,334	63,167	67,505	71,420	76,154	79,490	86,661
60	35,534	37,485	40,482	43,188	46,459	52,294	59,335	66,981	74,397	79,082	83,298	88,379	91,952	99,607
70	43,275	45,442	48,758	51,739	55,329	61,698	69,335	77,577	85,527	90,531	95,023	100,425	104,215	112,317
80	51,172	53,540	57,153	60,391	64,278	71,145	79,335	88,130	96,578	101,879	106,629	112,329	116,321	124,839
90	59,196	61,754	65,647	69,126	73,291	80,625	89,335	98,650	107,565	113,145	118,136	124,116	128,299	137,208
100	67,328	70,065	74,222	77,929	82,358	90,133	99,335	109,141	118,498	124,342	129,561	135,807	140,169	149,449



UM ESTATÍSTICO É AQUELE QUE,
TENDO A CABEÇA A ARDER E OS
PÉS ENTERRADOS NO GELO,
AINDA DIZ QUE NA MÉDIA
ESTÁ TUDO BEM!...