

APOSTILA DE ESTATÍSTICA I

Este trabalho contém uma compilação de textos de diversos autores, tendo sido elaborado com o objetivo exclusivo de ser um apoio didático para o aluno em sala de aula.

Professora Walquiria Torezani

Vila Velha 2004

I - A Natureza da Estatística

1- Panorama Histórico

A origem da palavra Estatística está associada à palavra latina STATUS (Estado). Há indícios de que 3000 anos A.C. já se faziam censos na Babilônia, China e Egito e até mesmo o 4o. livro do Velho Testamento faz referência a uma instrução dada a Moisés, para que fizesse um levantamento dos homens de Israel que estivessem aptos para guerrear. Usualmente, estas informações eram utilizadas para a taxação de impostos ou para o alistamento militar. O Imperador César Augusto, por exemplo, ordenou que se fizesse o Censo de todo o Império Romano.

Contudo, mesmo que a prática de coletar dados sobre colheitas, composição da população humana ou de animais, impostos, etc., fosse conhecida pelos egípcios, hebreus, caldeus e gregos, e se atribuem a Aristóteles cento e oitenta descrições de Estados, apenas no século XVII a Estatística passou a ser considerada disciplina autônoma, tendo como objetivo básico a descrição dos BENS do Estado.

A palavra Estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall (1719-1772), que foi um notável continuador dos estudos de Hermann Conrig (1606-1681). Gottfried determinou os objetivos da Estatística e suas relações com as demais ciências.

Com a Escola Alemã as tabelas tornaram-se mais completas, surgiram as representações gráficas e o cálculo das probabilidades, e a Estatística deixou de ser simples catalogação de dados numéricos coletivos para se tornar o estudo de como chegar a conclusões sobre o todo ("população"), partindo da observação de partes desse todo ("amostras").

Atualmente, os estudos estatísticos têm avançado rapidamente e, com seus processos e técnicas, Têm contribuído para a organização dos negócios e recursos do mundo moderno.

2- O que é Estatística?

A Estatística é uma parte da Matemática Aplicada que fornece métodos para coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões.

A utilização de técnicas, destinadas à análise de situações complexas ou não, tem aumentado e faz parte do nosso cotidiano. Tome-se, por exemplo, as transmissões esportivas. Em um jogo de futebol, o número de escanteios, o número de faltas cometidas e o tempo de posse de bola são dados fornecidos ao telespectador e fazem com que a conclusão sobre qual time foi melhor em campo se torne objetiva (não que isso implique que tenha sido o vencedor...). O que tem levado a essa *qualificação* de nossas vidas no dia a dia?

Um fator importante é a popularização dos computadores. No passado, tratar uma grande massa de números era uma tarefa custosa e cansativa, que exigia horas de trabalho tedioso. Recentemente, no entanto, grandes quantidades de informações podem ser analisadas rapidamente com um computador pessoal e programas adequados. Desta forma o computador contribui, positivamente, na difusão e uso de métodos estatísticos. Por outro lado, o computador possibilita uma automação que pode levar um indivíduo sem preparo específico a utilizar técnicas inadequadas para resolver um dado problema. Assim, é necessária a compreensão dos conceitos básicos da Estatística, bem como as suposições necessárias para o seu uso de forma criteriosa.

A grosso modo podemos dividir a Estatística em três áreas:

- Estatística Descritiva
- Probabilidade
- Inferência Estatística

Vamos caracterizar estas três áreas

Estatística Descritiva

A Estatística Descritiva pode ser definida como um conjunto de técnicas destinadas a descrever e resumir dados, a fim de que possamos tirar conclusões a respeito de características de interesse. Em geral utilizamos a Estatística Descritiva na etapa inicial da análise quando tomamos contato com os dados pela primeira vez. Objetivando tirar conclusões de modo informal e direto, a maneira mais simples seria a observação dos valores colhidos. Entretanto ao depararmos com uma grande massa de dados percebemos, imediatamente, que a tarefa pode não ser simples. Para tentar retirar dos dados informações a respeito do fenômeno sob estudo, é preciso aplicar algumas técnicas que nos permitam simplificar a informação daquele particular conjunto de valores. A finalidade da Estatística Descritiva é tornar as coisas mais fáceis de entender, de relatar e discutir.

A média industrial Dow-Jones, a taxa de desemprego, o custo de vida, o índice pluviométrico, a quilometragem média por litro de combustível, as médias de estudantes são exemplos de dados tratados pela Estatística Descritiva.

Probabilidade

A Probabilidade pode ser pensada como o teoria matemática utilizada para estudar a *incerteza* oriunda de fenômenos que envolvem o *acaso*. Jogos de dados e de cartas, ou o lançamento de uma moeda para o ar enquadram-se na categoria do acaso. A maioria dos jogos esportivos também é influenciada pelo acaso até certo ponto. A decisão de um fabricante de cola de empreender uma grande campanha de propaganda visando a aumentar sua participação no mercado, a decisão de parar de imunizar pessoas com menos de vinte anos contra determinada doença, a decisão de arriscar-se a atravessar uma rua no meio do quarteirão, todas utilizam a probabilidade consciente ou inconscientemente.

Inferência Estatística

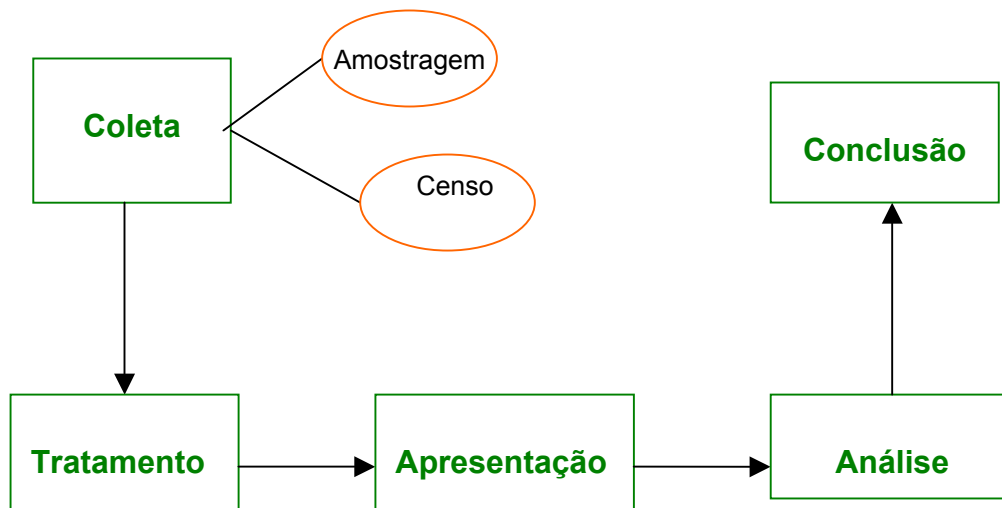
Inferência Estatística é o estudo de técnicas que possibilitam a extrapolação, a um grande conjunto de dados, das informações e conclusões obtidas a partir de subconjuntos de valores, usualmente de dimensões muito menores. Deve-se notar que se tivermos acesso a todos os elementos que desejamos estudar, não é necessário o uso das técnicas de inferência estatística; entretanto, elas são indispensáveis quando existe a impossibilidade de acesso a todo o conjunto de dados, por razões de natureza econômica, ética ou física.

Estudos complexos que envolvem o tratamento estatístico dos dados, usualmente incluem as três áreas citadas acima.

3 - Fases do Trabalho Estatístico

O trabalho estatístico é um método científico, que consiste das cinco etapas básicas seguintes:

- 1- Coleta e crítica de dados
- 2- Tratamento dos dados
- 3- Apresentação dos dados
- 4- Análise e interpretação dos resultados
- 5- Conclusão



Vamos tratar cada uma dessas etapas:

Coleta e crítica dos dados

Após definirmos cuidadosamente o problema que se quer pesquisar, damos início à coleta dos dados numéricos necessários à sua descrição.

A coleta pode ser direta ou indireta.

A coleta é direta quando feita sobre elementos informativos de registro obrigatório (nascimentos, casamentos e óbitos, importação e exportação de mercadorias), elementos pertinentes aos prontuários dos alunos de uma escola ou, ainda, quando os dados são coletados pelo próprio pesquisador através de inquéritos e questionários.

A coleta direta de dados pode ser classificada relativamente ao fator tempo em:

- Contínua – quando feita continuamente, tal como a de nascimentos e óbitos e a de frequência dos alunos às aulas.
- Periódica – quando feita em intervalos constantes de tempo, como os censos e as avaliações mensais dos alunos.
- Ocasional – Quando feita extemporaneamente, a fim de atender a uma conjuntura ou a uma emergência, como no caso de epidemias que assolam ou dizimam rebanhos inteiros.

A coleta se diz indireta quando é inferida de elementos conhecidos (coleta direta) e/ou do conhecimento de outros fenômenos relacionados com o fenômeno estudado. Como por exemplo, podemos citar a pesquisa sobre a mortalidade infantil, que é feita através de dados colhidos por uma coleta direta.

Obtidos os dados, eles devem ser cuidadosamente criticados, à procura de possíveis falhas e imperfeições, a fim de não incorrerem em erros grosseiros ou certo vulto, que possam influir sensivelmente nos resultados.

A crítica é externa quando visa às causas dos erros por parte do informante, por distração ou má interpretação das perguntas que lhe foram feitas; é interna quando visa observar os elementos originais dos dados da coleta.

Tratamento dos dados

Nada mais é do que a soma e o processamento dos dados obtidos e a disposição mediante critérios de classificação. Pode ser manual ou eletrônica.

Apresentação dos dados

Por mais diversa que seja a finalidade que se tenha em vista, os dados devem ser apresentados sob forma adequada – tabelas e gráficos – tornando mais fácil o exame daquilo que está sendo objeto de tratamento estatístico.

Análise dos resultados

Após a apresentação dos dados devemos calcular as medidas típicas convenientes para fazermos uma análise dos resultados obtidos, através dos métodos da Estatística Indutiva ou Inferencial, e tirarmos desses resultados conclusões e previsões.

Conclusão

É de responsabilidade de um especialista no assunto que está sendo pesquisado, que não é necessariamente um estatístico, relatar as conclusões de maneira que sejam facilmente entendidas por quem as for usar na tomada de decisões.

4 - A Estatística nas Empresas

No mundo atual, a empresa é uma das vigas-mestra da economia dos povos. A direção de uma empresa é de qualquer tipo, incluindo as estatais e governamentais, exige de seu administrador a importante tarefa de tomar decisões, e o conhecimento e o uso da Estatística facilitaram seu tríplice trabalho de organizar, dirigir e controlar a empresa.

Por meio de sondagem, de coleta de dados e de recenseamento de opiniões, podem conhecer a realidade social, os recursos naturais, humanos e financeiros disponíveis, as expectativas da comunidade sobre a empresa, e estabelecer suas metas, seus objetivos com maior possibilidade de serem alcançados a curto, médio ou longo prazo.

A Estatística ajudará em tal trabalho, como também na seleção e organização da estratégia a ser adotada no empreendimento e, ainda, na escolha das técnicas de verificação e avaliação da qualidade e da quantidade do produto e mesmo das possíveis lucros e/ou perdas.

Tudo isso que se pensou, que se planejou, precisa ficar registrado, documentado para evitar esquecimento, a fim de garantir o bom uso do tempo, da energia e do material e, ainda, para um controle eficiente do trabalho.

O esquema do planejamento é o plano, que pode ser resumido, com o auxílio da Estatística, em tabelas e gráficos, que facilitarão a compreensão visual dos cálculos matemáticos-estatísticos que lhes deram origem.

O homem de hoje, em suas múltiplas atividades, lança mão de técnicas e processos estatísticos, e só estudando-os evitaremos o erro das generalizações apresentadas a respeito de tabelas e gráficos apresentados em jornais, revistas e televisão, freqüentemente cometido quando se conhece apenas “por cima” um pouco de Estatística.

II – Amostragem

1- População e Amostras

Ao conjunto de entes portadores de, pelo menos, uma característica comum denominamos **População estatística** ou **universo estatístico**.

Esse termo refere-se não somente a uma coleção de indivíduos, mas também ao alvo sobre o qual reside nosso interesse. Assim, nossa população pode ser tanto todos os habitantes de Vila Velha, como todas as lâmpadas produzidas por uma fábrica em um certo período de tempo, ou todo o sangue no corpo de uma pessoa.

Como em qualquer estudo estatístico temos em mente pesquisar uma ou mais características dos elementos de alguma população, esta característica deve estar perfeitamente definida. E isso se dá quando, considerando um elemento qualquer, podemos afirmar, sem ambigüidade, se esse elemento pertence ou não à população.

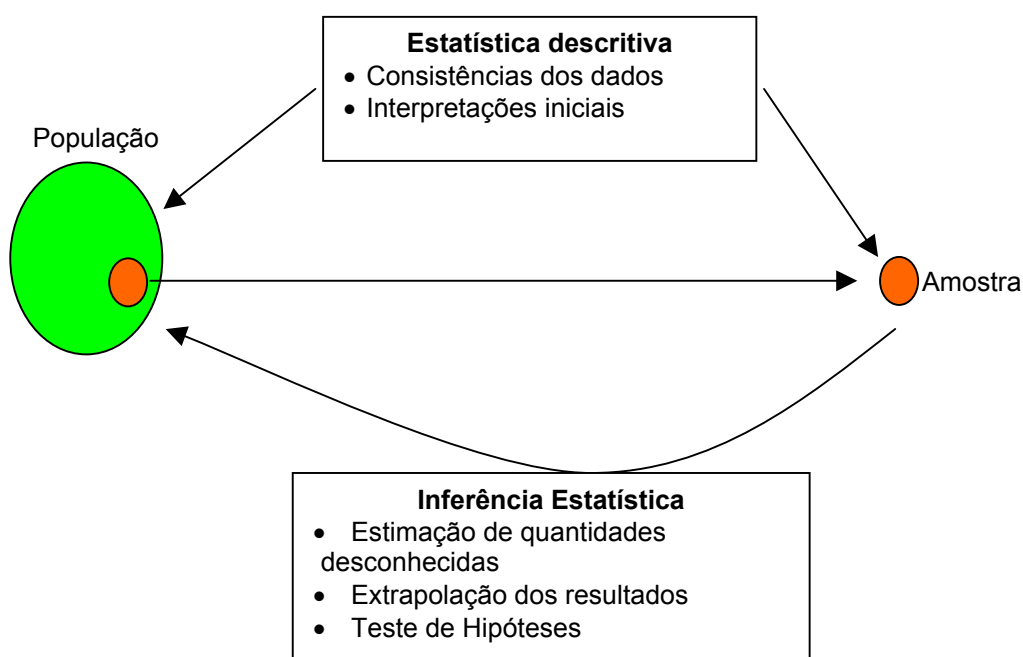
Vamos entender que, em Estatística, a palavra *população* tem significado muito mais amplo do que no vocabulário leigo. Para o estatístico, todos os valores que uma variável pode assumir, nos elementos de um conjunto, constitui uma população.

Algumas vezes podemos acessar toda a população para estudarmos características de interesse, mas em muitas situações, tal procedimento não pode ser realizado, por impossibilidade ou inviabilidade econômica ou temporal. Por exemplo, uma empresa não dispõe de verba suficiente para saber o que pensa todos os consumidores de seus produtos. Há ainda razões éticas, quando, por exemplo, os experimentos de laboratório envolvem o uso de seres vivos. Além disso, existem casos em que a impossibilidade de acessar toda a população de interesse é incontornável como no caso da análise do sangue de uma pessoa ou em um experimento para determinar o tempo de funcionamento das lâmpadas produzidas por uma indústria.

Tendo em vista as dificuldades de várias naturezas para observar todos os elementos da população, tomaremos alguns deles para formar um grupo a ser estudado. A essa parte proveniente da população em estudo denominamos **amostra**.

Uma **amostra** é um subconjunto finito de uma população.

Etapas da análise estatística.



Os pesquisadores trabalham com amostras. Primeiro, porque as *populações infinitas* só podem ser estudadas através de amostras. As *populações finitas muito grandes* também só podem ser estudadas através de amostras. Finalmente, o estudo cuidadoso de uma amostra tem mais valor científico do que o estudo rápido de toda a população.

Exemplos:

- De população infinita:

A produção futura de uma máquina.
As extrações, com repetição das bolas de uma urna.
Os nascimentos de bebês.

- O número de peixes do mar constitui uma população finita muito grande, pois esse número é, em dado momento, matematicamente finito, mas tão grande que pode ser considerado infinito para finalidade prática.
- Os alunos de uma sala de aula, os produtos de um supermercado, os livros de uma biblioteca, os automóveis de vila velha, são exemplos de populações finitas.

A distinção entre população e amostra é fundamental porque é com base nos dados de uma amostra que os estatísticos inferem sobre a população.

Exemplo: Uma pesquisa de opinião para saber o resultado das eleições para o governo do estado de São Paulo em 1988, a população considerada foram todos os eleitores do estado e para constituir a amostra o IBOPE coletou a opinião de cerca de 1600 eleitores.

As medidas estatísticas obtidas com base na população são denominadas *parâmetros*. As medidas obtidas com base em amostras são denominadas *estimativas*. Tanto parâmetros quanto estimativas são numéricos a única diferença é o fato de os parâmetros serem obtidos com base na população e as estimativas com base nas amostras.

Os parâmetros são em geral desconhecidos porque, na prática, não é possível observar toda a população. Mas, como já disse alguém, não é preciso beber todo o vinho para saber que gosto ele tem. Então o pesquisador obtém uma amostra para “ter uma idéia” do valor do parâmetro.

Embora nenhum plano de amostragem possa garantir que a amostra seja exatamente semelhante à população da qual foi extraída, se a amostra for suficientemente grande e obtida com a técnica correta, na maioria das vezes, poderemos estimar o valor do erro possível, isto é dizer “quão próxima” esta a amostra da população, em termos de representatividade. Mas ainda, amostras sucessivas da mesma população tendem a fornecer estimativas similares entre si e com valores em torno do verdadeiro, ou seja, o valor do parâmetro.

Exemplo: Imagine que a prefeitura de uma metrópole quer tomar uma medida administrativa que afeta os lojistas: Metade (0,5) deles é a favor e metade (0,5) é contra, mas ninguém conhece essas proporções. Você toma então uma amostra de dois lojistas para afirmar a proporção de lojistas favoráveis. Na sua amostra, podem ser favoráveis à medida:

- a) Nenhum dos dois
- b) Somente um deles
- c) Os dois

As estimativas da proporção de lojistas favoráveis à medida seriam:

- a) 0
- b) 0,5
- c) 1

Assim, vimos que, amostras diferentes dão estimativas diferentes do parâmetro. É o que os estatísticos chamam de *flutuação amostral*.

Das estimativas possíveis desse parâmetro, com base em uma amostra de tamanho 2, que é muitíssimo pequena, apenas um caso coincide com o valor verdadeiro do parâmetro e as outras são muito ruins.

Se a amostra fosse de tamanho 1000, uma proporção muito maior de estimativas estaria em torno do valor verdadeiro do parâmetro.

Existe uma técnica especial, a **amostragem**, para recolher amostras, que garantam, tanto quanto possível, o caráter de representatividade do todo, que possam ser usadas para permitir fazer inferências acerca da população de que originou. Quanto mais complexa for a amostragem, maiores cuidados deverão ser tomados nas análises estatísticas utilizadas; em contrapartida, o uso de um esquema de amostragem mais elaborado pode levar a uma diminuição no tamanho da amostra necessário para uma dada precisão.

Antes de escolher a amostra, é preciso definir a técnica de amostragem, isto é, os critérios que serão usados para escolher os elementos da população que constituirão a amostra. De acordo com a técnica usada, tem-se um tipo de amostra.

2- Amostragem X Censo

Uma amostra usualmente envolve o estudo de uma parcela dos itens de uma população, enquanto que um censo requer o exame de todos os itens. Embora concentremos nossa atenção nas amostras, é conveniente considerar também a alternativa do censo.

À primeira vista pode parecer que a inspeção completa ou total de todos os itens de uma população seja mais conveniente do que a inspeção de apenas uma amostra deles. Na prática, o contrário é que é quase válido

Firmas comerciais e entidades governamentais recorrem à amostragem por várias razões. O custo é usualmente um fator relevante. Colher dados e analisar resultados custam dinheiro e, em geral quanto maior o número de dados colhidos, maior o custo. Outra razão para o emprego de amostragem é que o valor da informação dura pouco. Para ser útil, a informação deve ser obtida e usada rapidamente. A amostragem é a única maneira de se fazer isso. Por vezes, o exame de determinado artigo o destrói. Testar cadeiras quanto a sua resistência ao peso obviamente as destrói; se fôssemos testar todas as cadeiras, não sobrariam cadeiras para a venda.

A amostragem é preferível ao censo quando:

- a) A população pode ser infinita, e obviamente não seria possível examinar todos os itens da população o que tornaria então o censo impossível.
- b) Uma amostra pode ser mais atualizada do que o censo. Se se necessita de uma informação rapidamente, um estudo de toda a população pode consumir demasiado tempo e perder utilidade. Além disso, se a população tende a modificar-se com o tempo, um censo poderá, na realidade, combinar várias populações.
- c) Os testes podem apresentar caráter destrutivo, ou seja, os itens examinados são destruídos no próprio ato do experimento. Então o censo nos daria o panorama preciso de uma população que não existe mais.
- d) O custo de um censo pode ser proibitivo, normalmente se o custo individual é elevado e se existem muitos itens na população.
- e) A precisão pode sofrer no caso de um censo de uma grande população. A amostragem envolve menor número de observações e, conseqüentemente, menor número de coletores de dados. Com grande número de agentes, há menor coordenação e controle, aumentando a chance de erros. A amostragem pode

revelar maior uniformidade nos métodos de coleta de dados, e maior comparabilidade entre os dados, do que um censo.

- f) O tipo de informação pode depender da utilização de uma amostra ou de um censo. Frequentemente as despesas com coleta de dados sofrem restrições orçamentárias. Existe também a premência do tempo. Se nos decidirmos por um censo, os problemas de custo e de tempo podem conduzir a uma limitação do censo a apenas uma ou a poucas características por item. Uma amostra com o mesmo custo e mesmo tempo, poderia proporcionar resultados mais aprofundados sobre um menor número de itens.

Entretanto, há certas situações em que é mais vantajoso fazer um censo. Entre essas situações destacamos:

- a) A população pode ser tão pequena que o custo e o tempo de um censo sejam pouco maiores que para uma amostra.
- b) Se o tamanho da amostra é grande em relação ao da população, o esforço adicional requerido por um censo pode ser pequeno, além disso, o censo eliminará a variabilidade amostral.
- c) Se se exige precisão completa, então o censo é o único método aceitável. Em face da variabilidade amostral, nunca podemos ter certeza de quais são os parâmetros verdadeiros da população. Um censo nos dará essa informação, embora erros na coleta dos dados e outros tipos de tendenciosidade possam afetar a precisão dos resultados.
- d) Ocasionalmente, já se dispõe da informação completa, de modo que não há necessidade de amostra.

3- Amostragem Probabilística

Uma amostragem será probabilística se todos os elementos da população tiverem probabilidade conhecida, e diferente de zero, de pertencer à amostra. Desta forma, a amostragem probabilística implica um sorteio com regras bem determinadas, cuja realização só será possível se a população for finita e totalmente acessível.

Consideraremos aqui os seguintes planos de amostragem probabilística:

- 1 – Amostragem Aleatória Simples
- 2 – Amostragem Proporcional Estratificada
- 3 – Amostragem Sistemática

Amostragem Aleatória Simples

Este tipo de amostragem é equivalente a um sorteio lotérico. A Amostragem Aleatória Simples é constituída de elementos retirados ao acaso da população. Então todo elemento da população tem probabilidade fixa de ser amostrado. Por isso é que a esse tipo de amostragem tende a produzir amostras representativas.

Exemplo: Geralmente são considerados aleatórios os seguintes processos:

- A chegada de carros a um posto de pedágio
- As chamadas telefônicas numa grande mesa de operação
- A chegada de clientes aos caixas de um supermercado
- A produção de qualquer processo mecânico
- Sucessivos lances de moeda ou de dado
- Tempo de serviço em estações de pedágio

É de máxima importância dar cuidadosa atenção à maneira como se escolhem os itens, bem como se eles são igualmente prováveis.

Exemplo: Imagine que 500 clientes estão cadastrados em sua empresa e você precisa obter uma amostra aleatória de 2% dos cadastros. O que você faria?

Como queremos uma amostra de 2% dos cadastros, precisamos sortear 10 deles. Faremos isso seguindo os seguintes passos:

1 – Numeramos os cadastros de 001 a 500.

2 - Para o sorteio exibiremos duas opções:

- a) Escreva os números de 001 a 500, em pedaços iguais de um mesmo papel, colocando-os dentro de uma caixa. Agite sempre a caixa para misturar bem os pedaços de papel e retire, um a um, dez números que formarão a amostra.
- b) Coloque em uma urna, bolas numeradas de zero a nove, inclusive, misture bem e retire uma. Anote o número dessa bola que será o primeiro dígito do número do cadastro que será amostrado. Volte a bola retirada à urna, misture bem e retire outra. O número dessa segunda bola será o segundo dígito do número do cadastro que será amostrado. O procedimento deverá ser repetido até completar os três dígitos da numeração utilizada. Como a população é constituída por 500 cadastros, devem ser desprezados os números maiores do que 500, bem como os números que já foram sorteados e o número 000. O sorteio deverá ser repetido até se conseguir a amostra de 10 cadastros.

O processo de seleção exige que se atribuam números consecutivos aos itens listados escolhendo-se depois, aleatoriamente, os números dos itens que comporão a amostra. Conceitualmente, podemos usar cartas, dados, fichas numeradas ou bolas numeradas para gerar números aleatórios para gerar números aleatórios correspondentes aos números de nossa listagem.

Na prática, tais dispositivos são empregados raramente, por várias razões. Uma delas é que cada dispositivo deixa algo a desejar; os métodos não são perfeitamente aleatórios. As cartas, por exemplo, podem aderir umas às outras, impedindo um embaralhamento perfeito. As arestas de um dado podem estar desgastadas. E sempre há o perigo de as bolas de uma urna não terem sido convenientemente misturadas. Em vista disso, e porque a amostragem aleatória é vital para a inferência estatística, existem tabelas especialmente elaboradas, chamadas **Tabelas de Números Aleatórios**, construída de modo que os dez algarismos (0 a 9) são distribuídos ao acaso nas linhas e colunas.

Na tabela de números aleatórios os dez algarismos 0,1,2,...,7,8,9, podem ser lidos isoladamente ou em grupos; podem ser lidos em qualquer ordem, como por colunas, num sentido ou noutro, por linhas, diagonalmente etc., e podem ser considerados aleatórios. A opção de leitura, porém, deve ser feita, antes de iniciado o processo. Para usar uma tabela de números aleatórios devemos:

- 1 – Fazer uma lista dos números da população
- 2 – Numerar consecutivamente os itens na lista, a começar do zero,
- 3 – Ler os números na tabela de números aleatórios de modo que o número de algarismos em cada um seja igual ao número de algarismos do último número da sua listagem.
- 4 – Desprezar quaisquer números que não correspondam a números da lista ou que sejam repetições de números lidos anteriormente. Continue o processo até ter o número desejado de observações.
- 5 – Usar os números assim escolhidos para identificar os itens da lista a serem incluídos na amostra.

EXEMPLO DE UMA TABELA DE NÚMEROS ALEATÓRIOS

(retirada de: STEVENSON, William J. Estatística aplicada
à administração, São Paulo: Harbra, 1981)

3690	2492	7171	7720	6509	7549	2330	5733	4730
0813	6790	6858	1489	2669	3743	1901	4971	8280
6477	5289	4092	4223	6454	7632	7577	2816	9202
0772	2160	8236	0812	4195	5589	0830	8261	9232
5692	9870	3583	8997	1533	6566	8830	7271	3809
2080	3828	7880	0586	8482	7811	6807	3309	2729
1039	3382	7600	1077	4455	8806	1822	1669	7501
7227	0104	4141	1521	9104	5563	1392	8238	4882
8506	6348	4612	8252	1062	1757	0964	2983	2244
5086	0303	7423	3298	3979	2831	2257	1508	7642
0092	1629	0377	3590	2209	4839	6332	1490	3092
0935	5565	2315	8030	7651	5189	0075	9353	1921
2605	3973	8204	4143	2677	0034	8601	3340	8383
7277	9889	0390	5579	4620	5650	0210	2082	4664
5484	3900	3485	0741	9069	5920	4326	7704	6525
6905	7127	5933	1137	7583	6450	5658	7678	3444
8387	5323	3753	1859	6043	0294	5110	6340	9137
4094	1957	0163	9717	4118	4276	9465	8820	4127
4951	3781	5101	1815	7068	6379	7252	1086	8919
9047	0199	5068	7447	1664	9278	1708	3625	2864
7274	9512	0074	6677	8676	0222	3335	1976	1645
9192	4011	0255	5458	6942	8043	6201	1587	0972
0554	1690	6333	1931	9433	2661	8690	2313	6999
8231	5627	1815	7171	8036	1832	2031	6298	6073
3995	9677	7765	3194	3222	4191	2734	4469	8617
2402	6250	9362	7373	4757	1716	1942	0417	5921
5295	7385	5474	2123	7035	9983	5192	1840	6176
5177	1191	2106	3351	5057	0967	4538	1246	3374
7315	3365	7203	1231	0546	6612	1038	1425	2709
5775	7517	8974	3961	2183	5295	3096	8536	9442
5500	2276	6307	2346	1285	7000	5306	0414	3383
3251	8902	8843	2112	8567	8131	8116	5270	5994
4675	1435	2192	0874	2897	0262	5092	5541	4014
3543	6130	4247	4859	2660	7852	9096	0578	0097
3521	8772	6612	0721	3899	2999	1263	7017	8057
5573	9396	3464	1706	9204	3389	5678	2589	0288
7478	7569	7551	3380	2152	5411	2647	7242	2800
3339	2854	9691	9562	3252	9848	6030	8472	2266
5505	8474	3167	8552	5409	1556	4247	4652	2953
6381	2086	5457	7703	2758	2963	8167	6712	9820

Exemplo: Imagine que 500 clientes estão cadastrados em sua empresa e você precisa obter uma amostra aleatória de 2% dos cadastros. Como você usaria a tabela de números aleatórios para extrair essa amostra?

Depois de numerar os cadastros podemos escolher, por exemplo, percorrer a última coluna da tabela de cima para baixo lendo os três primeiros algarismos de cada linha. Os números obtidos dessa forma são:

473, 828, 920, 923, 380, 272, 750, 488, 224,
764, 309, 192, 838, 466, 652, 344, 913, 412.

Desprezando os números que são maiores do que 500 (e eventuais repetições) devemos tomar para a amostra os cadastros de números:

473, 380, 272, 488, 224, 309, 192, 466, 344, 412.

Dispondo-se de uma lista precisa dos itens da população, é relativamente simples escolher uma amostra aleatória com o auxílio de uma tabela de números aleatórios. Na realidade, a lista não precisa conter todos os itens. As locações dos itens podem constituir uma alternativa, como por exemplo, os quarteirões de uma cidade, ou os arquivos de uma firma etc.

Amostragem Sistemática

Quando os elementos da população já se acham ordenados, não há necessidade de construir um sistema de referência. São exemplos os prontuários médicos de um hospital, os prédios de uma rua, as linhas de produção etc. Nestes casos, a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador.

A Amostragem Sistemática é constituída de elementos retirados da população segundo um sistema preestabelecido.

Exemplo 1: Imagine que 500 clientes estão cadastrados em sua empresa e você precisa obter uma amostra aleatória de 2% dos cadastros. Como você obterá uma amostra sistemática?

Precisamos obter uma amostra de tamanho 10. Para obter a amostra podemos dividir 500 por 10, e obter 50. Sorteamos um número entre 1 e 50, inclusive, para ser o primeiro cadastro da mostra e a partir desse número, contamos 50 cadastros e retiramos o último para fazer parte da amostra. Procedemos dessa forma até completarmos os 10 cadastros da amostra.

Exemplo 2: No caso de uma linha de produção, podemos, a cada dez itens produzidos, retirar um para pertencer a uma amostra da produção diária. Neste caso, estaríamos fixando o tamanho da amostra em 10% da população.

É preciso especial cuidado com o sistema de seleção. Não forme uma amostra com as primeiras pessoas de uma fila ou, se são atendidos 10 clientes por dia, não escolha para a amostra, o décimo de cada dez clientes. Estes procedimentos podem determinar amostras tendenciosas. Recomenda-se sempre sortear o primeiro elemento que será selecionado para a mostra e, a partir daí, usar o sistema de seleção.

Amostragem Proporcional Estratificada

Muitas vezes a população se divide em subpopulações, denominadas de Estratos. Como é provável que a característica em estudo dessa população apresente, de estrato em estrato, um comportamento heterogêneo e, dentro de cada estrato, um comportamento homogêneo, convém que o sorteio dos elementos da amostra leve em consideração tais estratos.

A amostra proporcional estratificada é composta por elementos proveniente de todos os estratos.

Exemplo: Vamos obter uma amostra proporcional estratificada de 10% para a pesquisa da estatura de 90 alunos de uma escola onde 54 são meninos e 36 são meninas.

Temos aqui dois estratos, sexo masculino e sexo feminino.

a) O primeiro passo é determinar o tamanho da amostra em cada estrato:

Sexo	População	10%	Amostra
M	54	$\frac{10 \times 54}{100} = 5,4$	5
F	36	$\frac{10 \times 36}{100} = 3,6$	4
Total	90	$\frac{10 \times 90}{100} = 9$	9

b) Numeramos os alunos de 01 a 90, sendo que de 01 a 54 correspondem meninos e de 55 a 90 meninas.

c) obtemos uma amostra aleatória ou sistemática de cada sexo e reunimos as informações numa só amostra, denominada amostra estratificada.

4- Amostragem Não Probabilística

Quando nem todos os elementos da população tiverem uma probabilidade diferente de zero de pertencerem à amostra, dizemos que a amostragem é não-probabilística.

Este processo de amostragem é subjetivo e seu regimento depende do conhecimento que possui o pesquisador a respeito da estrutura da população. É empregada, muitas vezes, por simplicidade ou pela impossibilidade de se obter amostragens probabilísticas.

Se os efeitos das amostragens não-probabilísticas podem ser considerados equivalentes aos das amostragens probabilísticas, então os processos de amostragens não-probabilísticas também podem ser considerados válidos.

Consideraremos aqui os seguintes planos de amostragem não probabilística:

1 – Amostragem a Esmo

2 – Amostragem por Julgamento

Amostragem a Esmo

É o caso em que o pesquisador procura ser aleatório, sem, no entanto, utilizar um sorteio aleatório rigoroso

Exemplo: Se tirarmos 100 parafusos de uma caixa que contém 10.000 parafusos do mesmo modelo e tamanho, de certo não faríamos uma amostragem aleatória simples, pois seria extremamente trabalhosa, mas faríamos retiradas a esmo.

Os resultados de uma amostragem a esmo são os mesmos de uma amostragem probabilística se a população é homogênea e se não existe a possibilidade de o amostrador ser influenciado (mesmo que inconscientemente) por alguma característica dos elementos da população. Não seria o caso da amostra dos parafusos, acima, se estes tivessem modelos e tamanhos diferentes, e isto afetasse a característica observada nos parafusos.

Amostragem por Julgamento

Neste tipo de amostragem, a amostra é colhida na parte da população que é acessível. Então se faz uma distinção entre população-objeto (aquela que se tem em mente ao realizar o estudo) e a população-amostrada (a parte da população que é acessível). Se essas duas populações tiverem as mesmas características, este tipo de amostragem vai ser equivalente a uma amostragem probabilística.

Se o tamanho da amostra é bem pequeno, digamos, de um a cinco itens, a amostragem aleatória pode dar totalmente não-representativos, ao passo que uma

pessoa familiarizada com a população pode especificar quais os itens mais representativos da população.

Exemplo: Uma cadeia de restaurantes pode querer experimentar uma nova técnica de serviço, empregando bandejas com aquecimento. Problemas de custo podem fazer com que a experiência se limite a dois restaurantes, os quais podem diferir consideravelmente em termos de tamanho, localização, clientela e lucratividade. Ao invés de uma seleção aleatória dos dois locais a ser usado como teste, será melhor confiar no conhecimento da administração para fazer tal escolha.

Ocasionalmente, os itens amostrais, se apresentam convenientemente agrupados. Uma pesquisa médica deve trabalhar com os pacientes disponíveis. Este grupo não pode ser considerado como uma amostra aleatória do público em geral e seria perigoso tentar tirar conclusões gerais com base em tal estudo. Entretanto, os resultados poderiam proporcionar uma base para a elaboração de um plano de amostragem aleatória para validar os resultados básicos. Os perigos inerentes à pesquisa médica, bem como a outros tipos de pesquisa, freqüentemente obrigam a limitar a pesquisa inicial a um pequeno grupo de voluntários. Outros exemplos similares seriam portadores de doenças fatais, cadáveres, animais, etc.

Finalmente, a amostragem por julgamento pode ser mais rápida e menos custosa porque não é preciso construir uma listagem dos itens da população.

Tenha-se em mente que a amostragem por julgamento não permite a avaliação objetiva do erro amostral, de modo que é conveniente usar a amostragem probabilística sempre que possível.

Comparação de Planos de Amostragem	
Tipo	Caracterizado por
Aleatória	Lista de itens
Sistemática	Lista aleatória de itens
Estratificada	Subgrupos Homogêneos
A Esmo e Por julgamento	Não necessita de uma listagem de itens

5- Amostras Tendenciosas

Talvez você nunca faça um trabalho que exija amostragem. Mas muito provavelmente você lerá ou usará resultados de trabalhos cujos dados foram obtidos por amostragem. Então saiba que é importante entender o que é uma amostra tendenciosa. Primeiro, as inferências devem ser feitas apenas para a população de onde a amostra foi retirada. Não tem sentido, por exemplo, estudar os hábitos de higiene dos índios bolivianos e fazer inferência para a população da periferia da cidade de São Paulo. Também é preciso verificar se a amostra foi retirada da população usando um processo delineado segundo critérios estatísticos.

A amostra deve ter o tamanho usual da área em que a pesquisa se enquadra. Na prática, o tamanho da amostra é determinado mais por considerações reais ou imaginárias a respeito do custo de cada unidade amostrada do que por técnicas estatísticas. Veja o que se faz na sua área de trabalho, consultando a literatura e verifique o que seu orçamento permite fazer.

As amostras muito pequenas podem ser excelentes estudos de casos, mas não permitem fazer inferência estatística. Mas também desconfie de amostras muito grandes. Será que o pesquisador tinha tempo e dinheiro para fazer um bom levantamento de tantos dados? E veja como foi feito o questionário. São mais confiáveis as respostas obtidas através de entrevistas, desde que o entrevistador tenha sido treinado. Estude as perguntas. Elas eram claras? As respostas podem ser, por alguma razão, mentirosas? Leia o artigo e se pergunte:

- 1 – Qual é a população?
- 2 – Como a mostra foi selecionada?
- 3 – Qual é o tamanho da amostra?
- 4 – Como o questionário foi feito?
- 5 – As perguntas eram claras?

Se estas perguntas não tiverem resposta satisfatória, a amostra pode ser tendenciosa.

Exemplo: Para estimar o tamanho dos morangos de uma caixa, não seria correto examinar o tamanho dos 10 morangos que estão na parte de cima, pois provavelmente, a amostra seria tendenciosa uma vez que os vendedores de morango arrumam as caixas de maneira a colocar as frutas maiores nas camadas mais superficiais.

É preciso cuidado na forma de tomar a amostra porque os erros de amostragem fazem com que os resultados da amostragem sejam diferentes dos resultados do censo.

Quando a pesquisa exige que os participantes respondam a um questionário, é preciso especial atenção à forma de obter as respostas. Se o questionário é enviado às pessoas pelo correio, a taxa de resposta é baixa. Tendem a responder mais as pessoas que têm opinião formada, seja a favor ou contra o assunto. Então a amostra pode ser tendenciosa pelo fato de não conter pessoas indiferentes ao tema tratado. Por telefone obtêm-se respostas mais rapidamente, mas no Brasil a maioria das residências não tem telefone. Então a amostra pode ser tendenciosa pelo fato de só terem sido entrevistadas pessoas suficientemente ricas para ter um telefone.

As entrevistas pessoais têm a taxa mais alta de respostas, mas exigem do entrevistador tempo e treinamento. Mesmo assim, muitas unidades podem não responder. A falta de resposta se explica pela inabilidade do pesquisador para entrar em contato com as pessoas ou para conseguir respostas daqueles que, a princípio, se recusam a responder. Se essas pessoas diferem da população a amostra é tendenciosa.

Mas existem, ainda outras fontes de erros nos resultados de um levantamento por amostragem.

Exemplo: A origem do erro pode estar na própria pergunta que é feita ao entrevistado.

- A pessoa pode mentir sobre sua idade ou sobre sua renda
- A pessoa pode não lembrar e dar resposta errada quando perguntada sobre questões do tipo: "Quantos cigarros o senhor fumou na semana passada?"
- Quem não entende a pergunta pode dar qualquer resposta, apenas para não mostrar ignorância.
- Perguntas mal feitas que induzem a certa resposta. Por exemplo, "Você acha justo pessoas de idade ficarem passeando de ônibus de graça, enquanto estudantes e trabalhadores têm que pagar?"

Finalmente, cabe tratar aqui algumas questões de ética que surgem quando se faz um levantamento de dados. É claro que podem existir abusos. Esses abusos não são, contudo, da ordem dos que, vez por outra, ocorrem nos experimentos científicos. De qualquer forma, a primeira questão de ética que pode surgir é o fato de a pessoa que responde estar sendo usada para uma pesquisa, sem saber disso. As pessoas devem ser avisadas de que estão sendo entrevistadas para uma pesquisa e têm o direito de se recusar a participar. Ainda, na maioria das vezes, quem responde não sabe o uso que se fará do dado. Mas todo pesquisador deve informar às pessoas que participam da pesquisa quais são os seus objetivos e tem o dever de informar, depois de terminado o trabalho, quais foram as suas conclusões.

Outra questão ética, muito mais séria do que as anteriores é a dos pesquisadores sociais que se infiltram em certos grupos, fingindo pertencer a eles, só para obter informações. Essa atitude é defensável apenas nos raríssimos casos em que a informação obtida pela pesquisa é absolutamente essencial e não pode ser obtida por outros meios. E, por último, vem a questão da identificação, sem a necessária autorização, de quem responde, usando expedientes como questionários com códigos secretos. Definitivamente, a identificação por código não é ética.

Um estatístico conhecido disse, certa vez, que é possível mentir usando estatística, mas que se mente mais, e melhor, sem estatísticas. É preciso entender que as amostras podem levar a conclusões erradas. Contudo, as opiniões pessoais, sem base em dados, levam em geral, a conclusões muito mais erradas.

III – Organização de Dados

1- Variáveis

Os dados estatísticos se obtêm mediante um processo que envolve a observação ou outra mensuração de características de uma população ou amostra tais como renda anual numa comunidade, sexo dos indivíduos de uma tribo indígena, percentagem de açúcar em cereais, etc. Cada uma dessas características é chamada de **variável**, porque originam valores que tendem a exibir certo grau de variabilidade quando se fazem mensurações sucessivas.

Exemplo: Suponha que um questionário foi aplicado aos alunos do 3º período do curso de administração da UNIVILA fornecendo as seguintes informações:

- 1) Id: Identificação do aluno
- 2) Turma: Turma em que o aluno foi alocado (A ou B)
- 3) Sexo: F se feminino, M se masculino.
- 4) Idade: Idade em anos
- 5) Altura: Altura em metros
- 6) Peso: Peso em quilogramas
- 7) Filhos: Número de filhos na família

- Para a variável “sexo” são dois os valores possíveis: F ou M
- Para a variável “Filhos” os valores possíveis são expressos através de números naturais: 0, 1, 2, 3,...
- Para a variável “Altura” temos uma situação diferente, pois os resultados podem assumir um número infinito de valores dentro de um determinado intervalo.

Claramente tais variáveis têm naturezas diferentes no que tange aos possíveis valores que podem assumir. Tal fato deve ser levado em conta nas análises dos dados, pois para cada tipo de variável existe um tratamento diferente.

2- Classificação das variáveis

Vamos considerar dois grandes tipos de variáveis:

A) Quantitativas (numéricas): São as variáveis cujos valores são expressos em números. Elas podem ser subdivididas em *quantitativas discretas* e *quantitativas contínuas*. As variáveis discretas podem ser vistas como resultantes de contagens, assumindo assim, valores inteiros. Já as variáveis contínuas geralmente provêm de uma mensuração e podem assumir qualquer valor em intervalos dos números reais.

Exemplos:

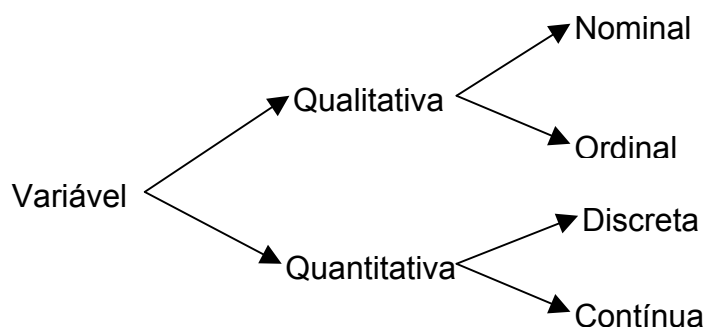
- São variáveis quantitativas discretas: Número de irmãos, de alunos numa sala de aula, de defeitos num carro novo, etc.
- São variáveis quantitativas contínuas: Altura, peso, comprimento, espessura, velocidade, etc.

B) Qualitativas (não numéricas): São as variáveis cujos possíveis valores que assumem representam atributos e/ou quantidades. Se tais variáveis têm uma ordenação natural, indicando intensidades crescentes de realização, então elas serão classificadas como *qualitativas ordinais ou por postos*. Caso contrário, quando não é possível estabelecer uma ordem natural entre seus valores definindo apenas uma categoria, elas são classificadas como *qualitativas nominais*.

Exemplos:

- São variáveis qualitativas nominais: Turma (A ou B), sexo (F ou M), cor dos olhos, campo de estudo, etc.

- São variáveis qualitativas ordinais: Tamanho (pequeno, médio ou grande), Classe social (baixa, média ou alta), etc.
Podemos resumir a classificação das variáveis no seguinte esquema:



É interessante notar que muitas populações podem originar os quatro tipos de dados como ilustramos na tabela abaixo.

Populações	Tipo de variáveis			
	Quantitativas		Qualitativas	
	Contínua	Discreta	Nominal	Ordinal
Alunos de Estatística	Idade, peso	Nº na classe	sexo	Por período
Automóveis	Velocidade em km/h	Nº de defeitos por carro	cores	limpeza
Venda de Imóveis	Valor em reais	Nº de ofertas	Acima do preço	Muito dispendioso

3- Distribuição de Frequência

Uma distribuição de frequência é um método de grupamento de dados em classes, ou intervalos, de tal forma que se possa determinar o número ou a percentagem de observações em cada classe. O número ou percentagem numa classe chama-se *frequência de classe*. Uma distribuição de frequência pode ser apresentada sob forma gráfica ou tabular.

Tipos de frequências

- **Frequência simples ou absoluta (f):** São os valores que realmente representam o número de dados de cada classe.
- **Frequência relativa (fr):** São os valores das razões entre as frequências simples e a frequência total. Normalmente calcula-se a frequência relativa para efeito de comparação com outros grupos ou conjunto de dados. Convém notar que, quando estivermos comparando dois grupos com relação às frequências de ocorrência dos valores de uma dada variável, grupos com um número total de dados maior tendem a ter maiores frequências de ocorrência dos valores da variável. Dessa forma, o uso de frequência relativa vem resolver este problema.
- **Frequência acumulada (F):** É o total das frequências de todos os valores inferiores ao limite superior do intervalo de uma dada classe. Normalmente utilizamos esse tipo de frequência quando tratamos de variáveis qualitativas ordinais ou quantitativas em geral.
- **Frequência acumulada relativa (Fr):** É o total das frequências relativas de todos os valores inferiores ao limite superior do intervalo de uma dada classe. Como no caso anterior utilizamos esse tipo de frequência quando tratamos de variáveis qualitativas ordinais ou quantitativas em geral.

O processo de construção de uma distribuição de frequência para determinado conjunto de dados depende do tipo de dados em estudo, isto é, contínuos, discretos, nominais ou ordinais. Vamos estudar cada caso.

Os dados coletados são registrados em fichas que contêm, além dos dados de interesse, diversas outras informações. Portanto, terminada a fase de coleta dos dados, é preciso retirar os dados das fichas e organizá-los. Esta fase do trabalho é denominada, tecnicamente, de *tratamento dos dados*.

Tabelas e Gráficos

Os dados depois de tratados podem ser apresentados em tabelas. Existem normas nacionais para a organização de tabelas, ditadas pela ABNT. Essas normas não serão tratadas aqui, mas convém saber que as tabelas devem ter os seguintes componentes:

- **Título:** Precede a tabela e explica, em poucas palavras, o dado em estudo. Se for o caso, indica o tempo e o lugar a que os dados se referem.
- **Cabeçalho:** Especifica o conteúdo de cada coluna
- **Coluna Indicadora:** Especifica em cada linha os valores que os dados podem assumir.
- **Corpo da tabela:** Apresenta a frequência dos dados.
- **Fonte:** Especifica a entidade, o pesquisador ou pesquisadores que forneceram os dados, quando esses não foram coletados por você.

Exemplo

Título	Valor em dólares dos principais produtos que o Brasil vende à Argentina	
Cabeçalho	Produto	Valor em dólares (em bilhões)
Coluna Indicadora	Automóveis	606
	Veículos de carga	541
	Autopeças	531
	Motores	264
	Minério	248
	Tratores	130
	Fonte: Época 25 de janeiro de 1999	

A organização dos dados em tabelas de frequência proporciona um meio eficaz de estudo do comportamento de características de interesse. Muitas vezes, a informação contida nas tabelas pode ser mais facilmente visualizada através de gráficos. Meios de comunicação apresentam, diariamente, gráficos das mais variadas formas para auxiliar na apresentação das informações. Órgãos públicos e empresas se municiam de gráficos e tabelas em documentos internos e relatórios de atividades e desempenho. Graças à proliferação de recursos gráficos, cuja construção tem sido cada vez mais simplificada em programas computacionais, existe hoje uma infinidade de tipos de gráficos que podem ser utilizados.

Deve ser notado, entretanto, que a utilização de recursos visuais na criação de gráficos deve ser feita cuidadosamente; um gráfico desproporcional em suas medidas pode dar falsa impressão de desempenho e conduzir a conclusões equivocadas. Obviamente, questões de manipulação incorreta da informação podem ocorrer em qualquer área e não cabe culpar a Estatística. O uso e a divulgação ética e criteriosa de dados devem ser pré-requisitos indispensáveis e inegociáveis.

Exemplos: Vamos definir quatro tipos básicos de gráficos:

- **Gráfico de disco:** É usado para mostrar a importância relativa das proporções. Esse tipo de gráfico se adapta melhor às variáveis qualitativas nominais.

- **Gráfico de barras:** Utiliza o plano cartesiano com os valores da variável no eixo das abscissas e as frequências ou porcentagens no eixo das ordenadas. Esse tipo de gráfico se adapta melhor às variáveis discretas ou qualitativas ordinais.
- **Histograma:** É a representação gráfica de uma distribuição de frequência por meio de retângulos justapostos. Esse tipo de gráfico se adapta melhor às variáveis quantitativas contínuas.
- **Polígono de frequência:** É uma alternativa ao histograma construído mediante a conexão dos pontos médios dos intervalos do histograma com linhas retas.

Distribuição de Frequência para Variáveis Quantitativas Contínuas

Os principais estágios na construção de uma distribuição de frequência para dados contínuos são:

- 1 – Organizar os dados brutos em um rol de ordem crescente ou decrescente.
- 2 – Determinar a amplitude total dos dados que é a diferença entre o maior e menor dos dados.
- 3 – Determinar quanto ao número de classes a usar (k). É aconselhável usar entre 5 e 15 classes. Menos que cinco classes pode ocultar detalhes importantes dos dados, e mais que quinze torna a apresentação demasiado detalhada. Uma regra prática consiste em tomar a raiz quadrada do número total de dados (\sqrt{n}) e ajustá-la, se necessário, aos limites de 5 a 15.
- 4 – Determinar a amplitude de cada classe dividindo a amplitude total por k. Se necessário o valor encontrado deve ser aproximado para cima com o mesmo número ou mais casas decimais que os valores das variáveis.
- 5 – Estabelecer os intervalos das classes começando com um inteiro logo abaixo do menor valor observado e somando a amplitude das classes. Os intervalos de classe devem ser escritos, de acordo com a Resolução 866/66 do IBGE em termos de “desta quantidade até menos aquela”, empregando, para isso, o símbolo --- (inclusão por limite inferior e exclusão do limite superior).
- 6 – Relacionar os intervalos e fazer a contagem dos pontos por classe. A contagem total deve ser igual a n.
- 7 – construir uma tabela de frequência ou um gráfico de frequência.

Exemplo: Considere os dados brutos que representam a safra, em alqueires, por árvore, para um conjunto de 40 pessegueiros.

Safra atual em (alqueire/ árvore) para 40 Pessegueiros							
11,1	12,5	32,4	7,8	21,0	16,4	11,2	22,3
4,4	6,1	27,5	32,8	18,5	16,4	15,1	6,0
10,7	15,8	25,0	18,2	12,2	12,6	4,7	23,5
14,8	22,6	16,0	19,1	7,4	9,2	10,0	26,2
3,5	16,2	14,5	3,2	8,1	12,9	19,1	13,7

Vamos construir uma tabela para representar esses dados:

1- Construção do Rol em ordem crescente

Safra atual em (alqueire/ árvore) para 40 Pessegueiros							
3,2	3,5	4,4	4,7	6	6,1	7,4	7,8
8,1	9,2	10,0	10,7	11,1	11,2	12,2	12,5
12,6	12,9	13,7	14,5	14,8	15,1	15,8	16,0
16,2	16,4	16,4	18,2	18,5	19,1	19,1	21,0
22,3	22,6	23,5	25,0	26,2	27,5	32,4	32,8

2 – Amplitude Total (AT) = $32,8 - 3,2 = 29,6$

3 – Número de classes (K) = $\sqrt{40} = 6,32 \approx 6$

Podemos construir uma distribuição de frequência, sem perda dos valores originais, utilizando como classes os inteiros de 0 a 9.

Nº de acidentes diários num estacionamento, durante 50 dias.		
Nº de acidentes	f	F
0	3	3
1	2	5
2	5	10
3	6	16
4	9	25
5	7	32
6	7	39
7	6	45
8	4	49
9	1	50
Total	50	

Podemos ver pela frequência acumulada que em 25 dias (dos 50) ocorreram até 4 acidentes por dia.

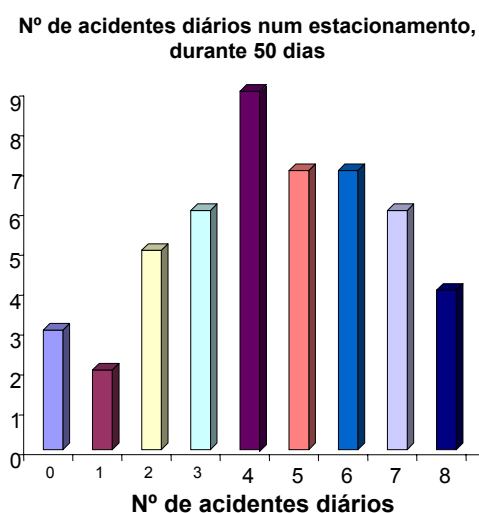
Dizemos que não há perda de informações porque poderíamos reconstruir os dados originais a partir desta tabela. Por outro lado, poderíamos tratar essa variável como se fosse contínua, gerando a seguinte tabela:

Nº de acidentes diários num estacionamento, durante 50 dias.		
Nº de acidentes	f	fr
0 - 2	5	0,10
2 - 4	11	0,22
4 - 6	16	0,32
6 - 8	13	0,26
8 - 10	5	0,10
Total	50	1,00

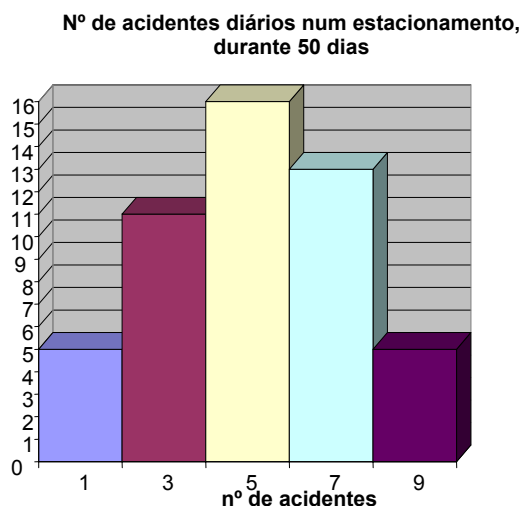
Para a distribuição de frequência sem perda de informações construímos o gráfico

de barras, enquanto que para a distribuição com perda de informações construímos um histograma.

a) Gráfico de barras (sem perdas)



b) Histograma (com perdas)



De modo geral, prefere-se uma distribuição de frequência sem perda de informações quando:

- Os dados são constituídos de valores inteiros
- Há menos de, digamos, 16 dados
- Há suficientes observações para originar uma distribuição significativa.

Por outro lado, uma distribuição de frequência com perda de informações é útil quando:

- Estão em jogo inteiros e não-inteiros (ou não inteiros somente)
- Só existem inteiros, porém em número demasiadamente elevado para permitir uma distribuição útil.
- A perda de informações é de importância secundária.

Distribuição de Frequência para Variáveis Qualitativas Nominais ou Ordinais

Talvez as distribuições de frequência mais simples sejam as relativas as variáveis nominais ou ordinais. Tal simplicidade decorre do fato de que as classes são facilmente reconhecíveis, tornando mínimos os cálculos.

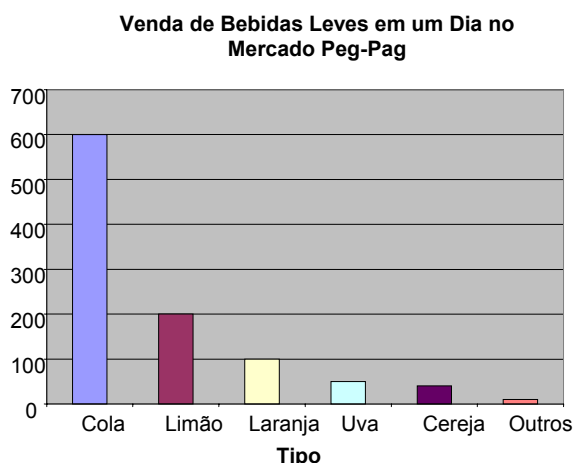
Exemplo 1: Considere os dados nominais referentes à venda de bebidas leves em um dia no Mercado Peg-Pag, dispostos na tabela de frequência abaixo:

Venda de Bebidas Leves em um Dia no Mercado Peg-Pag		
Tipo de Bebida	f	fr
Cola	600	0,60
Limão	200	0,20
Laranja	100	0,10
Uva	50	0,05
Cereja	40	0,04
Outros	10	0,01
Total	1000	1,00

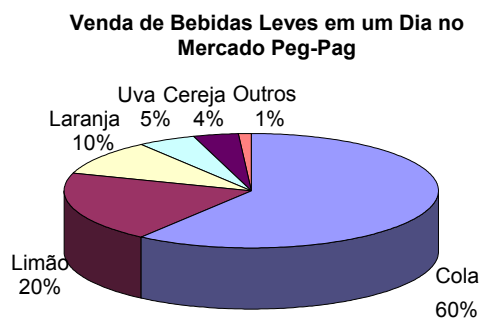
As categorias são os diversos tipos de bebidas. Pode haver diversos tipos de bebidas com vendas bastante baixas, tais como soda, cerveja e chocolate, que foram englobadas numa única categoria, que chamamos de “Outros”, para tornar os dados mais abrangentes.

Podemos optar pela construção de um gráfico de barras horizontais ou verticais usando as frequências simples ou destacar os percentuais de vendas de cada bebida construindo um gráfico de setores com as frequências relativas.

a) Gráfico de barras verticais



b) Gráfico de setores



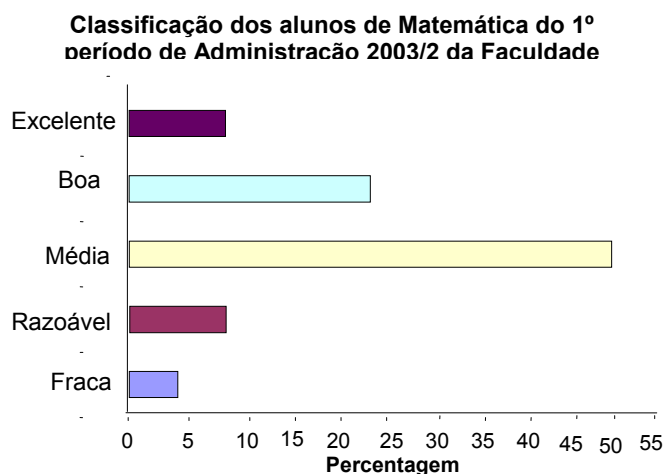
Exemplo 2: Consideremos os dados relativos ao aproveitamento num curso de Matemática para o 1º período de Administração 2003/2 da Faculdade UNIVILA, apresentados abaixo de forma ligeiramente diferente das tabelas de freqüências anteriores, apenas para ilustrar outra maneira de preparar uma tabela de freqüência.

Classificação dos alunos de Matemática do 1º período de Administração 2003/2 da Faculdade UNIVILA

Classificação	Fraca	Razoável	Média	Boa	Excelente	Total
Número de alunos	2	4	20	10	4	40
Percentagem	0,05	0,10	0,50	0,25	0,10	1,00

Podemos representar esses dados em um gráfico de barras horizontais ou de setores usando os valores das freqüências relativas

a) Gráfico de barras horizontais



b) Gráfico de setores

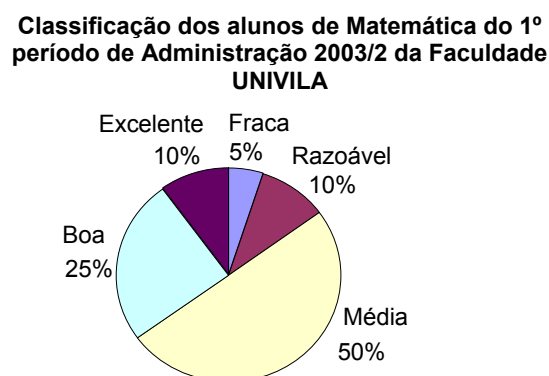


Diagrama de Pareto

É usado na gestão de qualidade para estabelecer a ordem em que as causas das perdas ou de outros tipos de fracasso devem ser sanadas.

O Diagrama de Pareto apresenta fracassos e insucessos em ordem de freqüência. Tem-se, então, a ordem em que devem ser sanados os erros, resolvidos os problemas, atendidas as reclamações, diminuindo o desperdício. Diz-se, por isso, que o Diagrama de Pareto estabelece prioridades. Mas o Diagrama de Pareto também

pode ser usado para identificar causas de sucesso como, por exemplo, as causas do aumento de venda de um produto.

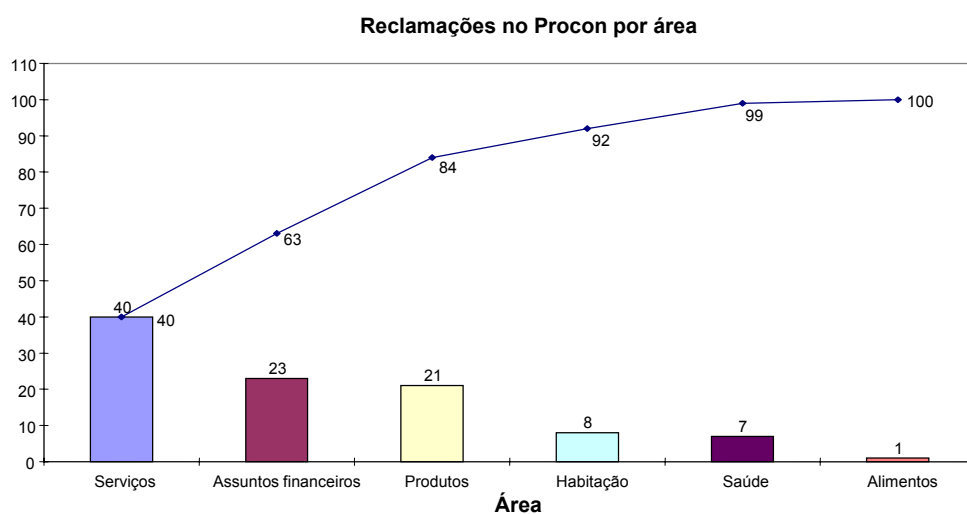
Exemplo: Consideremos a distribuição de frequência das reclamações feitas no Procon por área em 1999.

Reclamação do Procon por área		
Área	fr	Fr
Serviços	40	40
Assuntos financeiros	23	63
Produtos	21	84
Habitação	8	92
Saúde	7	99
Alimentos	1	100

Fonte: Procon, 1999.

Observe que os dados estão apresentados em ordem decrescente de frequências relativas.

O Diagrama de Pareto, neste caso, é:



Note que a linha que está desenhada acima das colunas é obtida com base na soma dos valores das colunas.

O Diagrama de Pareto para essas reclamações no Procon deve ser interpretado da seguinte forma: *"Como a maior parte de reclamações se concentram nas áreas de serviços e assuntos financeiros, o Procon deverá centralizar suas investigações nessas áreas"*.

IV – Séries Estatísticas

Denominamos **Série Estatística** toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do tempo ou da espécie.

Dai podemos inferir que numa série estatística observamos a existência de três elementos ou fatores:

- Tempo
- Espaço
- Espécie

Basicamente existem três tipos de séries estatísticas: Temporais ou Cronológicas, Geográficas e Categóricas.

1- Séries Temporais

São constituídas por dados produzidos e monitorados ao longo do tempo. Também são chamadas de séries históricas ou cronológicas.

Exemplos: São séries temporais:

Preço do Acém no varejo em São Paulo segundo o ano

Anos	Preço médio (US\$)
1989	2,24
1990	2,73
1991	2,12
1992	1,89
1993	2,04
1994	2,62

Fonte: APA

Número de apartamentos vendidos no mês de janeiro na cidade de São Paulo por ano.

Ano	Nº de Apartamentos
1995	1299
1996	533
1997	659
1998	1040
1999	402

Fonte: Secovi – SP (1999)

2- Séries Geográficas

São constituídas por dados provenientes de diferentes regiões geográficas. Também são chamadas de séries espaciais, territoriais ou de localização.

Exemplos: São séries geográficas:

Duração média dos estudos superiores 1994

Países	Número de anos
Itália	7,5
Alemanha	7,0
França	7,0
Holanda	5,9
Inglaterra	Menos de 4

Fonte: Revista Veja

Número de desempregados, em milhões, nos dez países com mais desemprego em 1998.

Países	Nº de desempregados (em Milhões)
Índia	38.960,1
Indonésia	10.625,7
Rússia	8.028,8
Brasil	6.649,9
EUA	6.173,4
China	6.125,1
Alemanha	4.075,7
Espanha	3.347,1
Japão	2.930,3
Itália	2.896,2

Fonte: Pochmann (1999)

3 – Séries Categóricas

São constituídas por dados obtidos nas diferentes categorias de uma mesma variável. Também são chamadas de séries específicas.

Exemplos: São séries categóricas:

Rebanhos brasileiros em 1992		Patrimônio líquido, segundo o banco, em 1998.	
Espécies	Quantidade (1000 cabeças)	Banco	Patrimônio líquido (milhões de reais)
Bovinos	154.440,8	Banco do Brasil	6.629,9
Bufalinos	1.423,3	Banespa	4.143,2
Eqüinos	549,5	Bank Boston	693,0
Asininos	47,1	Boa Vista	419,6
Muare	208,5	Bradesco	6.320,9
Suínos	34.532,2	HSBC	1.176,5
Ovinos	19.955,6	Itaú	4.650,7
Caprinos	12.159,6	Safr	931,2
Coelhos	6,1	Santander	960,7
Fonte: IBGE		Unibanco	2.906,3
		Fonte: EFC/Bancos (1999)	

4 – Séries Conjugadas

Muitas vezes temos necessidade de apresentar, em uma única tabela, a variação de valores de mais de uma variável, isto é, fazer uma conjugação de duas ou mais séries.

Conjugando duas séries em uma única tabela. Obtemos uma tabela de dupla entrada. Em uma tabela desse tipo ficam criadas duas ordens de classificação: uma horizontal e uma vertical.

Exemplo: A série conjugada abaixo é uma série geográfico-temporal.

Terminais telefônicos em serviço de 1991 a 1993			
Regiões	1991	1992	1993
Norte	342.938	375.658	403.494
Nordeste	1.287.813	1.379.101	1.486.649
Sudeste	6.234.501	6.729.467	7.231.634
Sul	1.497.315	1.608.989	1.746.232
Centro-oeste	713.357	778.925	884.822
Fonte: Ministério das comunicações			

V – Construção de tabelas e gráficos usando o Excel

1- Construção de tabelas de frequência para Variáveis Discretas

Suponhamos que um grupo tenha feito quarenta entrevistas e que uma das variáveis observadas tenha sido “Número de filhos”.

A função FREQUÊNCIA no Excel calcula a frequência de cada um dos valores observados. Para isso é preciso que os dados estejam digitados (em colunas ou em linhas) e que forneça os intervalos com os quais queremos que se faça a contagem.

Exemplo: Considere a seguinte tabela primitiva de dados:

Número do quest.	Nº de filhos	Número do quest.	Nº de filhos	Número do quest.	Nº de filhos	Número do quest.	Nº de filhos
1	1	11	1	21	0	31	1
2	0	12	2	22	1	32	0
3	0	13	0	23	1	33	1
4	1	14	2	24	1	34	0
5	0	15	0	25	0	35	1
6	1	16	1	26	1	36	0
7	0	17	2	27	0	37	0
8	1	18	2	28	3	38	0
9	1	19	1	29	0	39	0
10	2	20	4	30	2	40	0

Usando o Excel faça um pequeno banco de dados conforme mostrado a seguir.

Nas células A2 a A41 digite os números dos questionários.

Nas células B2 a B41 digite os dados (número de filhos).

Nas células C2 a C6 digite os valores 0, 1, 2, 3, 4 (são os limites das classes).

Marcar as células D2 a D7. Na barra de ferramentas escolha INSERIR → FUNÇÃO

Escolha primeiramente a opção ESTATÍSTICA.

Depois escolha a opção FREQUÊNCIA → OK.

O primeiro dado que será pedido é Matriz_dados e você deve, neste campo marcar os valores onde estão os dados; no nosso caso B2 a B41.

A seguir, pede-se Matriz_bin e você deve, neste campo, marcar os valores que correspondem aos limites das classes, no nosso caso C2 a C6.

Logo abaixo da Matriz_bin aparecerá o resultado: {17; 15; 6;1;1;0}.

Para que estes valores sejam mostrados na planilha dê CTRL + SHIFT + ENTER.

O resultado deve ser interpretado desta forma:

Nº de filhos	f
0	17
1	15
2	6
3	1
4	1
5 ou mais	0
Total	40

2- Construção de tabelas de frequência para Variáveis Qualitativas

Para construir tabelas de frequência para Variáveis Qualitativas procede-se da mesma maneira que para as Variáveis Discretas tomando-se apenas o cuidado de codificar os possíveis valores da variável e digitar a tabela primitiva no Excel usando esse código.

Exemplo: Uma variável qualitativa com três valores (A, B, C) foi observada em cada um dos vinte elementos de uma amostra. Obteve-se:

B	A	A	C	A	B	B	B	A	A
C	A	B	B	A	C	B	B	B	A

Podemos codificar os valores dessa variável como: A = 0, B = 1 e C = 2.

Devemos prosseguir como na construção de tabelas para variáveis quantitativas discretas digitando o seguinte banco de dados no Excel:

1	0	0	2	0	1	1	1	0	0
2	0	1	1	0	2	1	1	1	0

A tabela fornecida pelo Excel será:

0	8
1	9
2	3
	0

Que para ser interpretada deve ser decodificada. O resultado deve ser interpretado desta forma:

Variável	f
A	8
B	9
C	3
outros	0
Total	20

3 - Construção de tabelas de frequência para Variáveis Contínuas

Suponhamos que um grupo tenha feito vinte e duas entrevistas e que uma das variáveis observadas tenha sido “Tempo de realização de um dado exercício”.

Como esta variável é contínua, não podemos usar a função FREQUÊNCIA para construir uma tabela de frequências.

ATENÇÃO! Confira se no seu computador tem a opção Análise de dados... na opção FERRAMENTAS.

Se não tiver, escolha, na opção FERRAMENTAS a opção SUPLEMENTOS.

Marque no quadrinho Ferramentas de análise e → OK.

Exemplo: Considere a seguinte tabela primitiva de dados:

Número do quest.	Tempo	Número do quest.	Tempo
1	0,0	12	2,7
2	0,5	13	2,3
3	0,6	14	4,5
4	1,5	15	3,7
5	3,3	16	2,9
6	4,2	17	0,7
7	3,9	18	4,7
8	1,0	19	3,1
9	5,8	20	4,1
10	4,8	21	4,9
11	4,0	22	1,2

Usando o Excel faça um pequeno banco de dados conforme mostrado a seguir.

Nas células A2 a A23 digite os números dos questionários.

Nas células B2 a B23 digite os dados (tempos).

Agora vá para:

FERRAMENTAS → ANÁLISE DE DADOS → HISTOGRAMA → OK

Os dados pedidos são:

- Intervalo de entrada: Marcar no banco de dados, a coluna referente aos dados, no nosso caso B2 a B23.

- Intervalo de saída: Indique uma célula vazia, a partir da qual será mostrada a tabela (p.ex. G2).

A planilha apresentada será a seguinte:

0	1
1,45	5
2,9	4
4,35	7
mais	5

Que deve ser interpretada desta forma:

Tempos		f
0	1,45	6
1,45	2,90	4
2,90	4,35	7
4,35	5,80	5
Total		22

4- Construção de gráficos de barras.

O gráfico estatístico é uma forma de apresentar os dados estatísticos. A principal vantagem de um gráfico sobre uma tabela é que ele permite uma visualização imediata da distribuição dos valores observados.

O primeiro passo para se descrever graficamente um conjunto de dados observados é verificar as freqüências dos diversos valores das variáveis.

Para fazer um gráfico de barras no Microsoft Excel acompanhe o exemplo:

Exemplo: Considere a seguinte tabela de freqüência:

Nº de filhos	f
0	17
1	15
2	6
3	1
4	1
Total	40

1º) Digitar os valores da variável (antecedidos de uma 'aspa simples') e as freqüências, como abaixo:

	A	B
1	Nº de filhos	f
2	'0	17
3	'1	15
4	'2	6
5	'3	1
6	'4	1

2º) Marcar as células A1 a B6

3º) na barra de ferramentas: INSERIR → GRÁFICO.

Tipo de gráfico: Colunas → Avançar → Avançar

Colocar o título do gráfico:

Distribuição do número de filhos de 40 entrevistados

Colocar nome no eixo X: *Nº de filhos*

Colocar nome no eixo Y: *f*

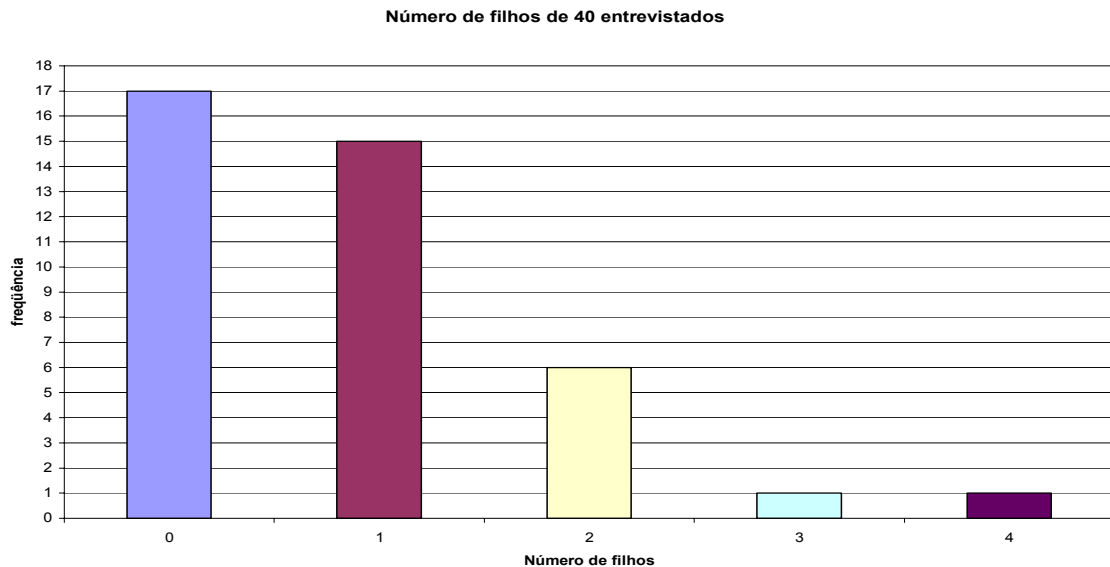
Ainda nesta janela escolha, no alto, “legenda” e desmarque a opção “mostrar legenda”
→ Avançar.

4º) Escolher a opção “Como nova planilha” e “concluir”.

5º) (opcional) Clique 2 vezes na parte colorida do gráfico.

No alto, escolha a opção “padrões” → escolha variar cor por ponto.

O gráfico será apresentado da seguinte forma:



Obs: Quando a variável é quantitativa é preciso digitar os valores da variável com uma aspa simples antes de cada valor ('5 ENTER, por exemplo) para diferenciar para o Excel quem é a variável e quem é a frequência.

5- Construção de Histogramas.

A Construção de um Histograma no Excel requer alguns detalhes. Acompanhe a partir do exemplo.

Exemplo: Para a variável “Tempo de realização de um dado exercício”, construímos a seguinte tabela de frequência:

Tempos		f
0	1,45	6
1,45	2,90	4
2,90	4,35	7
4,35	5,80	5
Total		22

Vamos construir o histograma que representa essa variável:

Procedimento:

1º) É preciso calcular os pontos médios de cada classe:

0 | 1,45 → ponto médio = 0,725

1,45 | 2,90 → ponto médio = 2,125

e assim sucessivamente.

2º) Digitar os pontos médios (antecedidos de uma ‘aspa simples’) e as frequências, como abaixo:

	A	B
1	Pontos médios	f
2	'0,725	6
3	'2,125	4
4	'3,625	7
5	'5,175	5

3º) Marcar as células A1 a B5

4º) na barra de ferramentas: INSERIR → GRÁFICO.

Tipo de gráfico: Colunas → Avançar → Avançar

Colocar o título do gráfico:

Distribuição dos tempos de 22 pessoas para a realização de um exercício.

Colocar nome no eixo X: *Tempo*

Colocar nome no eixo Y: *f*

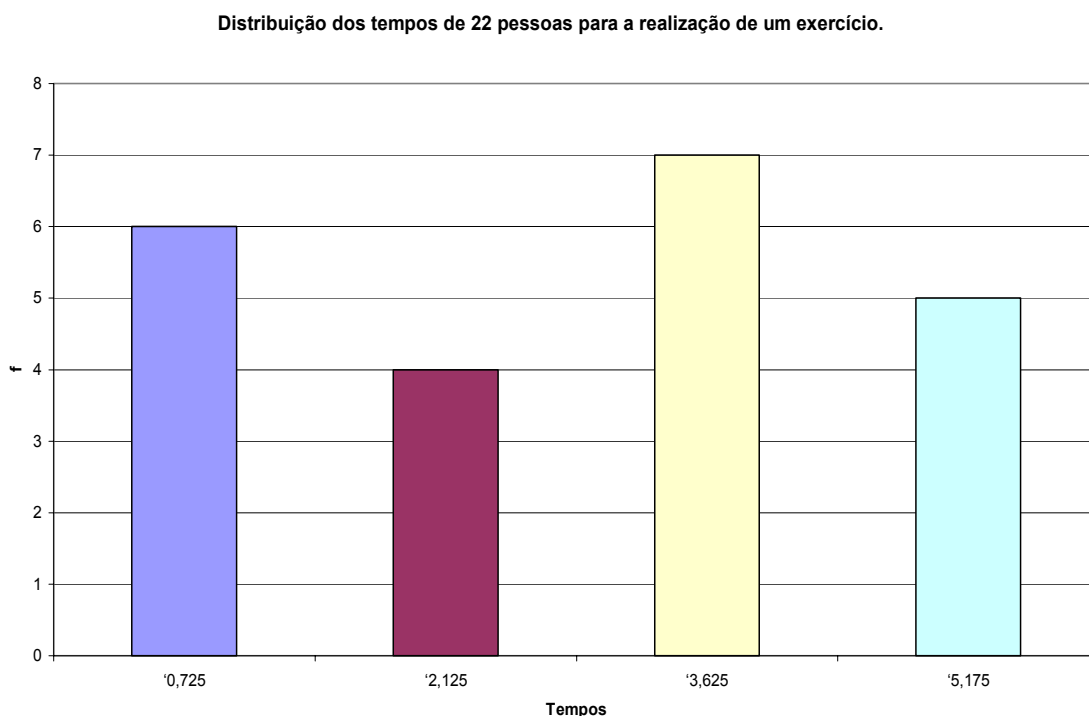
Ainda nesta janela escolha, no alto, “legenda” e desmarque a opção “mostrar legenda” → Avançar.

5º) Escolher a opção “Como nova planilha” e “concluir”.

6º) (opcional) Clique 2 vezes na parte colorida do gráfico.

No alto, escolha o opção “padrões” → escolha variar cor por ponto.

O gráfico será apresentado da seguinte forma:



Para que este gráfico seja um histograma é preciso fazer as seguintes alterações:

1º) Clique 2 vezes em alguma das barras. A janela “Formatar seqüência de dados” irá aparecer.

2º) Escolha, no alto, Opções.

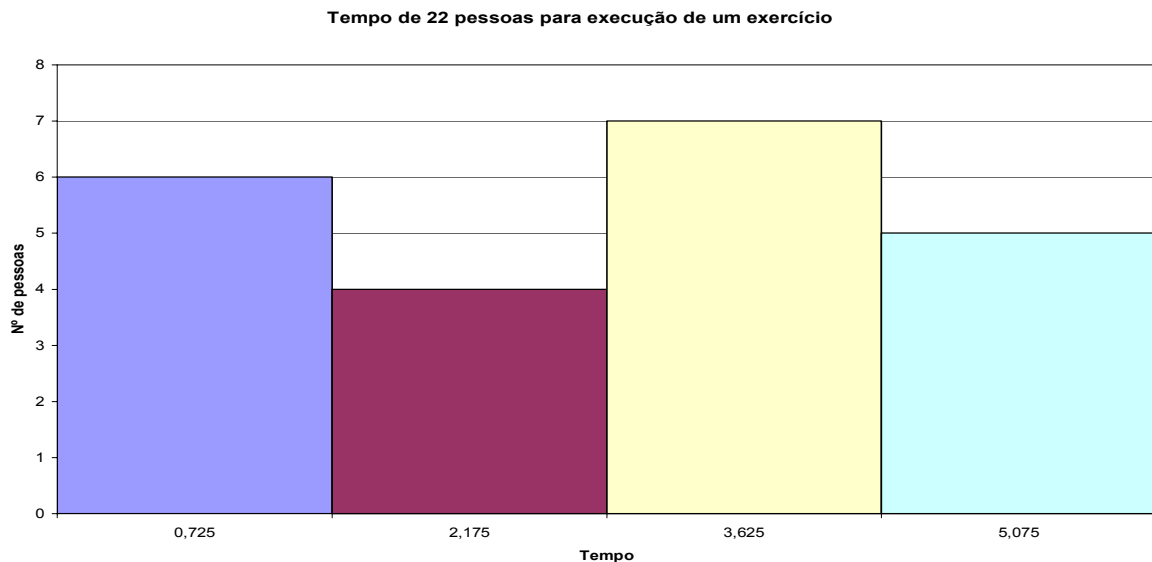
Na opção: sobreposição, passe o número para 100

Na opção: largura do espaçamento, passe para 0 → OK

3º) (Opcional) Clique 2 vezes na parte cinza do gráfico.

Na parte *Área* escolha a opção “nenhuma” →OK.

O gráfico agora estará como o da esquerda:



Observação: Para alterar o tipo e o tamanho das letras dos eixos, clique duas vezes sobre o eixo → vai aparecer uma janela “Formatar eixos” → escolha “fonte”, altere o tipo e → OK.

6 - Construção de Polígonos de Frequências.

Para se construir um polígono de frequências, é preciso entrar com os dados da seguinte forma:

	A	B
1	Pontos médios	f
2	'0,725	6
3	'2,125	4
4	'3,625	7
5	'5,175	5

1º) Marcar as células A1 a B12

2º) na barra de ferramentas: INSERIR → GRÁFICO.

Tipo de gráfico: Linha → Avançar → Avançar

3º) Colocar o título do gráfico:

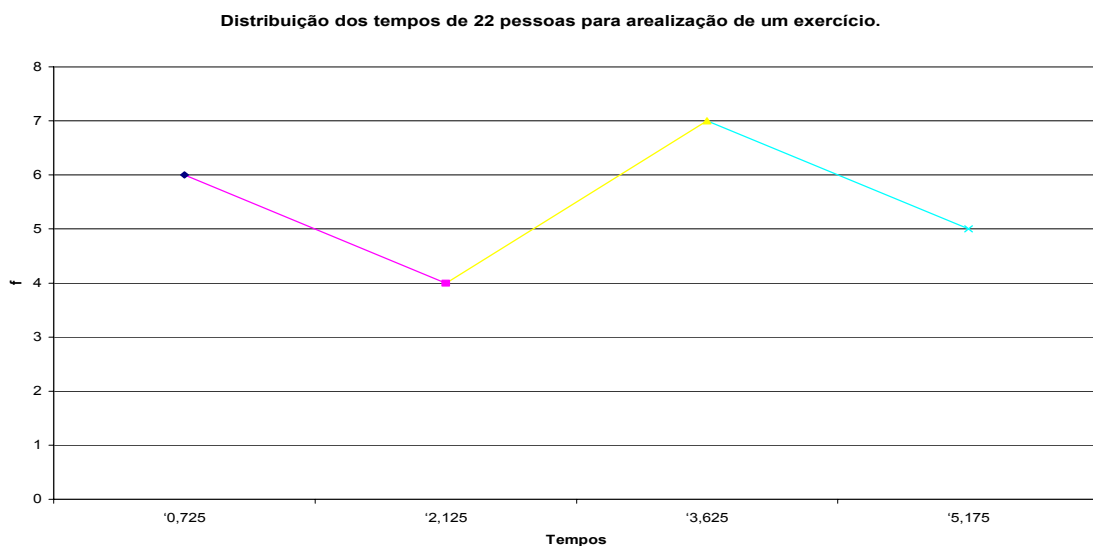
Distribuição dos tempos de 22 pessoas para a realização de um exercício.

Colocar nome no eixo X: *Tempo*

Ainda nesta janela escolha, no alto, “legenda” e desmarque a opção “mostrar legenda” → Avançar

4º) Escolher a opção “Como nova planilha” e “concluir”.

O gráfico agora estará como o da esquerda:



VI – Medidas de Tendência Central

1- Introdução

Se estivermos numa parada de ônibus urbano e nos pedirem alguma informação sobre a demora em passar um determinado ônibus, que diremos? Ninguém imagina que poderíamos dar como resposta uma tabela de frequências que pacientemente coletamos no último mês, ou ano! Quem pergunta deseja uma resposta breve e rápida que sintetiza a informação que dispomos e não uma completa descrição dos dados coletados.

Para resumir a quantidade de informação contida em um conjunto de dados, os estatísticos definem medidas que descreve, através de um só número, características dos dados. Algumas dessas medidas descrevem a *tendência central*, isto é, a tendência que os dados têm de se agrupar em torno de certos valores. Dentre as medidas de tendência central, destacamos:

- A média Aritmética
- A Mediana
- A Moda

2 – A Média Aritmética

A média aritmética é a idéia que ocorre à maioria das pessoas quando se fala em “média”. E como ela possui certas propriedades matemáticas convenientes, é a mais importante das três medidas que estudaremos.

A média aritmética de um conjunto de dados é a soma de todos eles dividido pelo número deles.

Exemplo: Calcule a média dos dados: 0, 2, 4, 6, 8.

Basta somar todos os valores e dividir o resultado pelo número de parcelas que é 5. Assim temos:

$$\frac{0 + 2 + 4 + 6 + 8}{5} = 4$$

Ou seja, a média aritmética, nesse caso, é 4.

Quando alguém fala sobre um conjunto de dados, tanto pode estar se referindo a uma amostra como a uma população. Utilizamos o símbolo μ para a média de uma população e o símbolo \bar{x} para representar a média de uma amostra. A média da população também é obtida dividindo a soma dos dados pelo número de elementos da população. Não calculamos μ porque, em geral, temos apenas uma amostra da população. Mas a média da amostra é uma *estimativa* da média da população.

Às vezes, a média pode ser um número diferente de todos os da série de dados que ela representa, por isso costumamos dizer que a média aritmética não tem existência concreta.

Exemplo: Calculando a média dos dados: 2, 4, 6, 8 temos:

$$\frac{2 + 4 + 6 + 8}{4} = 5$$

Isto é, a média aritmética, nesse caso, é 5. Esse será o número representativo dessa série de valores, embora não esteja representado nos dados originais.

Desvio em relação à média

Denominamos desvio em relação à média a diferença entre cada elemento de um conjunto de valores e a média aritmética.

Designando o desvio pelo símbolo d_i , podemos escrever:

$$d_i = x_i - \bar{x}$$

Onde x_i é um elemento do conjunto de valores.

Exemplo: Sabendo-se que a produção leiteira diária da vaca Mimosa, durante uma semana, foi de 10, 14, 13, 15, 16, 18, 12 litros, temos, para a produção média semanal:

$$\frac{10 + 14 + 13 + 15 + 16 + 18 + 12}{7} = \frac{98}{7} = 14$$

Logo, $\bar{x} = 14$ litros.

Os desvios em relação à média são dados por:

$$d_1 = x_1 - \bar{x} = 10 - 14 = -4$$

$$d_2 = x_2 - \bar{x} = 14 - 14 = 0$$

$$d_3 = 13 - 14 = -1$$

$$d_4 = 15 - 14 = 1$$

$$d_5 = 16 - 14 = 2$$

$$d_6 = 18 - 14 = 4$$

$$d_7 = 12 - 14 = -2$$

A soma dos desvios dos números de um conjunto a contar da média é zero.

Exemplo: No exemplo anterior temos:

$$-4 - 1 + 2 - 2 + 0 + 1 + 4 = 0$$

Propriedades da média aritmética

A média aritmética tem certas propriedades interessantes e úteis, que explicam por que é ela a medida de tendência central mais usada:

1 - A média aritmética de um conjunto de números pode sempre ser calculada.

2 - Para um dado conjunto de números a média aritmética é única.

3 - A média é sensível a (ou afetada por) todos os valores do conjunto. Assim, se um valor se modifica, a média também se modifica.

Exemplo: A média dos números 1, 2, 2, 3, 4, é dada por

$$\frac{1 + 2 + 2 + 3 + 4}{5} = 2,4$$

Se alterarmos o conjunto para 1, 2, 3, 3, 4 a média passa a ser dada por

$$\frac{1 + 2 + 3 + 3 + 4}{5} = 2,6$$

4 - Somando-se (ou subtraindo-se) uma constante (**c**) de todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante. Simbolicamente se,

$$y_i = x_i \pm c$$

então

$$\bar{y} = \bar{x} \pm c.$$

Exemplo: Somando-se 2 litros de leite a cada produção diária da Mimosa temos que:

$$y_1 = 12, y_2 = 16, y_3 = 15, y_4 = 17, y_5 = 18, y_6 = 20 \text{ e } y_7 = 14$$

Dai:

$$\bar{y} = \frac{12 + 16 + 15 + 17 + 18 + 20 + 14}{7} = \frac{112}{7} = 16$$

Lembrando que a média anterior era $\bar{x} = 14$, temos que:

$$\bar{y} = 16 = 14 + 2 = \bar{x} + 2.$$

5 - Multiplicando-se (ou dividindo-se) uma constante (**c**) de todos os valores de uma variável, a média do conjunto fica multiplicada (ou dividida) dessa constante. Simbolicamente se,

$$y_i = x_i \times c$$

então

$$\bar{y} = \bar{x} \times c.$$

Exemplo: Multiplicando-se por 3 cada produção diária da Mimosa temos que:

$$y_1 = 30, y_2 = 42, y_3 = 39, y_4 = 45, y_5 = 48, y_6 = 54 \text{ e } y_7 = 36$$

Daí:

$$\bar{y} = \frac{30 + 42 + 39 + 45 + 48 + 54 + 36}{7} = \frac{294}{7} = 42$$

Lembrando que a média anterior era $\bar{x} = 14$, temos que:

$$\bar{y} = 42 = 14 \times 3 = \bar{x} \times 3.$$

A Média Aritmética Ponderada

A fórmula anterior para calcular a média aritmética supõe que cada observação tenha a mesma importância. Apesar desse caso ser o mais geral, há exceções. Podemos considerar casos em que as observações tenham importâncias diferentes. Nesses casos, devemos ponderar a importância de cada variável para calcular a média aritmética.

Exemplo 1: Em uma faculdade a média semestral de cada disciplina é calculada considerando as duas médias bimestrais com peso 3 cada uma e um exame final com peso 4. Se um aluno obtém 8,0 no 1º bimestre, 9,0 no 2º bimestre e 9,6 no exame final de Estatística, Qual será a sua média semestral em Estatística?

O cálculo da média aritmética deve levar em conta os pesos desiguais das notas. Assim, para esse aluno temos:

Avaliação	Notas	Peso
1º Bimestre	8,0	3
2º Bimestre	9,0	3
Exame final	9,6	4
Total		10

Logo:

$$\text{média ponderada} = \frac{3(8,0) + 3(9,0) + 4(9,6)}{10} = \frac{89,4}{10} = 8,94$$

Isto é, A média semestral desse aluno em Estatística é 8,94.

Exemplo 2: Suponha que em outra faculdade a média semestral de cada disciplina é calculada considerando um exame no meio do período e um exame final, este último com o dobro do peso daquele. Qual será a média semestral em Estatística de um estudante com 8,5 no primeiro exame e 9,6 no exame final?

Neste caso temos:

Avaliação	Notas	Peso
Meio do Período	8,5	1
Exame final	9,6	2
Total		3

Logo:

$$\text{média ponderada} = \frac{1(8,5) + 2(9,6)}{3} = \frac{27,7}{3} = 9,23$$

Portanto sua média semestral em Estatística é 9,23.

Cálculo da Média Aritmética para Dados Agrupados

Se os dados se repetem para calcular a média é preciso multiplicar cada valor pela sua respectiva frequência. Na verdade, ao fazer isso, estamos usando uma variante da fórmula de cálculo da média ponderada, onde os pesos são substituídos pelas frequências das classes.

Exemplo 1: Durante uma manhã, um feirante vendeu determinado produto a preços variados: 12 unidades foram vendidas a 2 reais; 10 unidades foram vendidas a 3 reais e 8 unidades foram vendidas a 6 reais. Qual foi o preço médio de venda desse produto naquela manhã?

Podemos construir um rol para a variável “Preço de venda do produto”:

2	2	2	2	2	2	2	2	2	2
2	2	3	3	3	3	3	3	3	3
3	3	6	6	6	6	6	6	6	6

Para calcular o preço médio basta somar todos esse valores e dividir o resultado por 30 que é o número total de produtos vendidos. Observe essa conta:

$$\bar{x} = \frac{\overbrace{2+2+2+2+2+2+2+2+2+2}^{12 \times 2} + \overbrace{3+3+3+3+3+3+3+3+3+3}^{10 \times 3} + \overbrace{6+6+6+6+6+6+6+6+6+6}^{8 \times 6}}{30}$$
$$\frac{12 \times 2 + 10 \times 3 + 8 \times 6}{30} = \frac{24 + 30 + 48}{30} = \frac{102}{30} = 3,40$$

Assim o preço médio do produto pela manhã foi R\$ 3,40.

Observe agora, a tabela de frequência dessa variável:

Valor	f
2	12
3	10
6	8
Total	30

A segunda igualdade da expressão acima, nos mostra que basta ponderar cada preço praticado pela frequência com que foi praticado, para calcular o preço médio de venda.

Para facilitar as contas, podemos criar uma coluna auxiliar na tabela de frequência, chamada de “multiplicação”, correspondente aos produtos $x_i f_i$, e calcular o preço médio dividindo o total dessa coluna pelo total da coluna de frequência simples.

Valor	f	Multiplicação
2	12	2 x 12 = 24
3	10	3 x 10 = 30
6	8	6 x 8 = 48
Total	30	102

Nesse caso o cálculo da média aritmética se resume à conta:

$$\bar{x} = \frac{102}{30} = 3,40$$

Exemplo 2: Foram coletadas 150 observações da variável X, representando o número de vestibulares (um por ano) que um mesmo estudante prestou. Os dados estão na tabela abaixo:

X	f
1	75
2	47
3	21
4	7
Total	150

Quantas vezes, em média, cada aluno prestou vestibular?
Criando a coluna multiplicação temos:

X	f	Multiplicação
1	75	75
2	47	94
3	21	63
4	7	28
Total	150	260

$$\bar{X} = \frac{260}{150} = 1,73$$

Pode ser de interesse estudar o gasto dos alunos associado com as despesas do vestibular. Para simplificar um pouco a situação, vamos supor que se atribui, para cada aluno, uma despesa fixa de R\$ 1300,00, relativo à preparação e mais R\$ 50,00 para cada vestibular prestado. De posse dessas informações, vamos calcular a média da variável D: Despesa com vestibular. Pela definição desta nova quantidade temos:

$$D = 50 \times X + 1300$$

Logo do cálculo que fizemos é imediato que:

$$\bar{D} = 50 \times 1,73 + 1300 = 1386,50 .$$

Se os dados estão distribuídos em classes, para calcular a média aritmética multiplique o ponto médio de cada classe por sua respectiva frequência, some e divida o total pela soma das frequências e a média resultante é uma aproximação.

Nesse caso, convencionamos que todos os valores incluídos em um determinado intervalo de classe coincidem com o seu ponto médio, o que nem sempre é o caso. Entretanto, se não dispomos dos dados originais, não há outra alternativa razoável.

Exemplo: Imagine que a margem de lucro na venda de um produto é variável, mas que, ao longo de seis meses, foram registrados os valores apresentados na tabela abaixo. Calcule a média aritmética da margem de lucro nesse período.

Margem de lucro, em termos de percentual do valor de compra, segundo a classe.

Classe	f	Ponto médio	Multiplicação
15 - 25	30	20	600
25 - 35	45	30	1350
35 - 45	150	40	6000
45 - 55	45	50	2250
55 - 65	30	60	1800
Total	300		12000

Logo a margem média de lucro será dada por:

$$\bar{x} = \frac{12000}{300} = 40 .$$

3 – A Mediana.

Uma segunda medida de tendência central de um conjunto de números é a mediana. **Mediana** é o valor que ocupa a posição central do conjunto dos dados ordenados.

Da definição de mediana, segue-se que sua característica principal é dividir um conjunto ordenado de dados em dois grupos iguais; a metade terá valores inferiores à mediana, a outra metade terá valores superiores à mediana. A mediana de uma amostra será indicada por **md**.

Para calcular a mediana, é necessário primeiro ordenar os valores (comumente) do mais baixo ao mais alto. Em seguida, conta-se até a metade dos valores para achar a mediana.

Em geral, a mediana ocupa a posição $\frac{n+1}{2}$, onde n representa a quantidade de valores do conjunto.

O processo para determinara a mediana é o seguinte:

- 1 – Ordenar os valores
- 2 – Se o número de dados é ímpar, a mediana é o valor que está no centro da série.
- 3 – Se o número de dados é par, a mediana é a média dos dois valores que estão no centro da série.

Exemplo 1: Calcule a mediana dos dados: 3,9,5,7,9,1,9.

Para obter a mediana, ordene os dados:

1,3,5,7,9,9,9.

A posição da mediana é dada por:

$$\frac{7+1}{2} = \frac{8}{2} = 4$$

Assim a mediana é o elemento que ocupa a 4ª posição na série dos dados ordenados. Isto é, a mediana é 7.

Note que antes do 7, existem três números menores: 1,3,5 e, depois do 7, três números maiores: 9,9,9.

Exemplo 2: Calcule a mediana dos dados: 42,3,9,5,7,9,1,9.

Para obter a mediana, ordene os dados:

1,3,5,7,9,9,9,42.

A posição da mediana é dada por:

$$\frac{8+1}{2} = \frac{9}{2} = 4,5$$

Como o número de dados é par, a mediana é a média aritmética dos valores que ocupam a 4ª e a 5ª posição na série dos dados ordenados.

Assim, a mediana é dada por:

$$\frac{7+9}{2} = \frac{16}{2} = 8.$$

Logo temos quatro valores maiores do que 8 (1,3,5,7), e quatro valores menores do que 8 (9,9,9,42).

Cálculo da Mediana para Dados Agrupados

Se os dados se agrupam em uma distribuição de freqüência, o cálculo da mediana se processa de modo muito semelhante àquele dos dados não agrupados, implicando, porém, a determinação prévia das freqüências acumuladas. Ainda aqui, temos que determinar um valor tal que divida a distribuição em dois grupos que contenham o mesmo número de elementos. Também o processo e os resultados diferem, dependendo de dispormos ou não dos dados originais.

Se dispusermos dos dados originais, o processo será o seguinte:

- 1 – Determine as freqüências acumuladas.

- 2 – Determinar a posição da mediana.
- 3 – Identificar a frequência acumulada imediatamente superior ao valor determinado em (1).
- 4 – A mediana será aquele valor da variável que corresponde a tal frequência acumulada se o número de dados for ímpar.
- 5 – A mediana será a média aritmética entre o valor da variável correspondente a essa frequência acumulada e o da posição anterior se o número de dados for par.

Exemplo 1: Considere a distribuição relativa a 33 famílias de quatro filhos, tomando para variável o número de filhos do sexo masculino:

Nº de meninos	f	F
0	2	2
1	6	8
2	10	18
3	11	29
4	4	33
Total	33	

Determine a mediana desse conjunto.

A posição da mediana é dada por:

$$posição = \frac{33 + 1}{2} = \frac{34}{2} = 17$$

A menor frequência acumulada que supera esse valor é 18, que corresponde ao valor 2 da variável, sendo esse o valor mediano. Logo:

$Md = 2$ meninos.

Exemplo 2: Considere a distribuição relativa a 8 pessoas, tomando para variável o número de vezes que vão ao cinema por mês:

Nº de idas ao Cinema	f	F
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8
Total	8	

Determine a mediana desse conjunto.

A posição da mediana é dada por:

$$posição = \frac{8 + 1}{2} = \frac{9}{2} = 4,5$$

Devemos determinar os valores da variável que estão na posição 4 e 5.

Para a posição 5 temos que a menor frequência acumulada que supera esse valor é 6, que corresponde ao valor 16 da variável, e para a posição 4 temos que o valor da variável correspondente é 15. Logo:

$$Md = \frac{15 + 16}{2} = \frac{31}{2} = 15,5$$

Sem os dados originais ficamos restritos à suposição de que os valores na classe que contém a mediana são distribuídos de forma uniforme. Assim, o processo será o seguinte:

- 1 – Determine as frequências acumuladas.
- 2 – Identificar o intervalo que contém a mediana.
- 3 – Determinar a posição da mediana nesse intervalo.
- 4 – Ordenar os valores daquela classe.

5 – Identificar a mediana.

Exemplo 1: Considere a distribuição relativa a 40 pessoas, tomando para variável estatura de cada uma.

Estaturas	f	F
150-154	4	4
154-158	9	13
158-162	11	24
162-166	8	32
166-170	5	37
170-172	3	40
Total	40	

Vamos calcular a mediana desses dados:

A posição da mediana é dada por:

$$posição = \frac{40 + 1}{2} = \frac{41}{2} = 20,5$$

Devemos determinar os valores da variável que estão na posição 20 e 21.

Esses dois valores estão na 3ª classe que correspondem às estaturas de 158cm a 162cm.

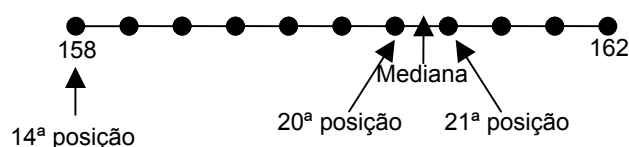
Existem 11 valores nessa classe ($24 - 13 = 11$) e a mediana é a média aritmética dos valores que estão nas posições 7 ($20 - 13$) e 8 ($21 - 13$) dessa classe.

A amplitude dessa classe é:

$$162 - 158 = 4$$

Supondo que os valores se distribuem uniformemente pela classe devemos dividir a amplitude da classe por 11 para determinar o “suposto” valor de cada elemento.

Graficamente temos:



Ou ainda:

Elemento	158	158,37	158,73	159,10	159,46	159,82	160,19	160,55	160,91	161,27	161,64
Posição na classe	0	1	2	3	4	5	6	7	8	9	10

Assim, devemos tomar a média aritmética dos valores:

$$158 + 7 \times \frac{4}{11} = 158 + 2,55 = 160,55. \text{ Correspondente à } 20^{\text{a}} \text{ posição.}$$

e

$$158 + 8 \times \frac{4}{11} = 158 + 2,91 = 160,91. \text{ Correspondente à } 21^{\text{a}} \text{ posição.}$$

Logo:

$$Md = \frac{160,55 + 160,91}{2} = \frac{321,46}{2} = 160,73$$

Dizemos então que metade dos alunos mede menos de 160,73cm e a outra metade mede mais de 160,73cm.

Comparação entre Média Aritmética e Mediana

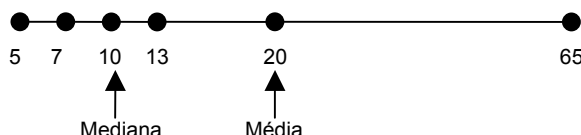
A escolha da média, ou da mediana, como medida de tendência central de um conjunto, depende de diversos fatores. A média é sensível a (ou influenciada por) cada valor do conjunto, inclusive os extremos. Por outro lado, a mediana é relativamente insensível aos valores extremos.

A mediana descreve bem os grandes conjuntos de dados. De qualquer forma, em algumas circunstâncias, a mediana descreve, melhor do que a média, a tendência central dos dados.

Exemplo: A mediana dos dados 5,7,10,13,65 é o número 10. Calculando a média desses dados temos que:

$$\bar{X} = \frac{5 + 7 + 10 + 13 + 65}{5} = \frac{100}{5} = 20$$

Graficamente temos:



Neste conjunto há um dado discrepante que é o 65. Esse valor “puxa” a média para cima, mas não afeta a mediana.

Por outro lado, para o conjunto 5,7,10,13,15, a mediana ainda é o número 10, e sua média é dada por:

$$\bar{X} = \frac{5 + 7 + 10 + 13 + 15}{5} = \frac{50}{5} = 10$$

Neste conjunto não há discrepância de valores.

Um exemplo clássico em que a mediana pode descrever melhor a tendência central dos dados do que a média é dado pelos salários de uma categoria profissional. É o caso dos salários dos jogadores de futebol no Brasil. A existência de alguns salários muito altos afeta mais a média do que a mediana. Então, a mediana dá, melhor do que a média, idéia do salário típico dessa categoria de profissionais.

De modo geral, a média possui certas propriedades matemáticas que a tornam atraente. Além disso, a ordenação dos dados para determinar a mediana pode ser enfadonha, e o cálculo da mediana não pode ser feito com máquina de calcular, ao contrário do que ocorre com a média.

4 – A Moda.

A moda é o valor que ocorre com maior frequência num conjunto. A moda de uma amostra será indicada por *mo*.

Exemplo: Determine a moda dos dados: 0,0,2,5,3,7,4,7,8,7,9,6.

A moda é 7, porque é o valor que ocorre o maior número de vezes.

Um conjunto de dados pode não ter moda porque nenhum valor se repete o maior número de vezes é o caso do conjunto:

3,5,8,10,12,13

Dizemos que esse conjunto é amodal.

Em outros casos, ao contrário, pode haver dois ou mais valores de concentração. Dizemos, então, que o conjunto tem dois ou mais valores modais. No conjunto:

2,3,4,4,4,5,6,7,7,7,8,9.

Temos duas modas: 4 e 7. Esse conjunto se diz bimodal.

A moda funciona como medida descritiva quando se trata de contar dados. Essa medida não se presta a manipulações matemáticas. De um ponto de vista puramente descritivo, a moda indica o valor “típico” em termos de maior ocorrência. Além disso, se as frequências são razoavelmente uniformes, a moda perde muito de sua importância como medida descritiva. Por outro lado, a utilidade da moda se acentua quando um ou dois valores, ou um grupo de valores, ocorre com muito mais frequência que os outros. Na maior parte das vezes, a média aritmética e a mediana fornecem melhor descrição da tendência central dos dados.

Quando existe uma grande quantidade de dados e, principalmente, os dados estão dispostos em uma tabela de frequências, a idéia da moda é útil. A moda de uma distribuição de frequência indica qual porção da distribuição tem a maior frequência de ocorrências. Em geral é bastante simples identificar a moda de dados que estão agrupados em uma tabela de frequências.

Exemplo: Considere a tabela de distribuição de frequências relativas ao percentual de vezes que 50 alunos formulam questões durante uma palestra.

Nº de vezes	f (%)
1	5
2	3
3	17
4	35
5	18
6	15
7	7
Total	100

A moda neste caso é 4 pois a maior parte dos alunos (35%) formularam 4 perguntas para o palestrante.

Quando há perda de informação, a moda se refere a uma “classe modal”, e não a um valor único. Ela mostra a tendência central dos dados identificando a área em que os dados estão mais concentrados.

Exemplo: Imagine que a margem de lucro na venda de um produto é variável, mas que, ao longo de seis meses, foram registrados os valores apresentados na tabela abaixo. Calcule a classe modal da margem de lucro nesse período.

Classe	f
15 - 25	30
25 - 35	45
35 - 45	150
45 - 55	45
55 - 65	30
Total	300

A classe modal é 35 - 45, porque é a classe com maior frequência, uma vez que 150 produtos foram vendidos com margem de lucro entre 35% e 45%.

Finalmente, a moda também pode ser usada para descrever dados qualitativos. Como, nesse caso, a moda é a categoria que ocorre com maior frequência, ela mostra a categoria que mais concentra dados.

Exemplo: O site de Veja na Internet perguntou que gastos as pessoas cortam devido à crise econômica. Qual é a moda?

Distribuição percentual dos respondentes, segundo o corte nos gastos.

Cortes	Percentual
Jantar fora	24
Viagens	23
Curso de idiomas	10
Cinema	7
Nenhum	36

A moda foi não cortar gastos, o que não significa que seja esse o comportamento da população em geral. Note que a amostra é constituída de assinantes do UOL, que responderam voluntariamente à pergunta.

Comparação entre Média, Mediana e Moda.

	Definição	Vantagens	Limitações	Quando usar
Média	Soma de todos os valores dividido pelo total de elementos do conjunto.	1 – Reflete cada valor 2 – Possui propriedades matemáticas atraentes	1 – É influenciada por valores extremos.	1 – Deseja-se obter a medida de posição que possui a maior estabilidade. 2 – Houver necessidade de um tratamento algébrico posterior.
Mediana	Valor que divide o conjunto em duas partes iguais.	1 – Menos sensível a valores extremos do que a média.	1 – Difícil de determinar para grande quantidade de dados.	1 – Deseja-se obter o ponto que divide o conjunto em partes iguais. 2 – Há valores extremos que afetam de uma maneira acentuada a média. 3 – A variável em estudo é salário.
Moda	Valor mais freqüente	1 – Valor “típico”. Maior quantidade de valores concentrados neste ponto.	1 – Não se presta à análise matemática. 2 – Pode não haver moda para certos conjuntos de dados.	1 – Deseja-se obter uma medida rápida e aproximada de posição. 2 – A medida de posição deve ser o valor mais típico da distribuição.

Referências Bibliográficas

STEVENSON, William J., **Estatística aplicada à administração**, Ed. Harbra, São Paulo, 1981.

CRESPO, Antônio Arnot, **Estatística Fácil**, Ed. Saraiva, São Paulo, 1998.

MAGALHÃES, Marcos Nascimento; LIMA, Antônio Carlos Pedroso de, **Noções de Probabilidade e Estatística**, 5ª ed., Editora da Universidade de São Paulo, São Paulo, 2002.

VIEIRA, Sônia, **Princípios de Estatística**, Ed. Pioneira Thomson Learning, São Paulo, 2003.

1ª LISTA DE EXERCÍCIOS DE ESTATÍSTICA I

- 1-** O que é Estatística?
- 2-** Quais são as três áreas principais da Estatística?
- 3-** Defina os termos “amostra” e “população”.
- 4-** Cite as fases do trabalho estatístico.
- 5-** O que é coleta de dados?
- 6-** Quais são as principais razões da amostragem?
- 7-** Para ser útil, que características deve ter uma amostra?
- 8-** Para que serve a crítica dos dados?
- 9-** O que é tratar dados?
- 10-** Como podem ser apresentados os dados?
- 11-** As conclusões, as inferências pertencem a que parte da Estatística?
- 12-** Cite três ou mais atividades do planejamento empresarial em que a Estatística se faz necessária.
- 13-** Para as situações descritas a seguir, identifique a população e a amostra correspondente. Discuta a validade do processo de inferência estatística para cada um dos casos.
 - a) Para avaliar a eficácia de uma campanha de vacinação no Estado de São Paulo, 200 mães de recém nascidos, durante o primeiro semestre de um dado ano e em uma dada maternidade em São Paulo, foram perguntadas a respeito da última vez em que vacinaram seus filhos.
 - b) Uma amostra de sangue foi retirada de um paciente com suspeita de anemia.
 - c) Para verificar a audiência de um programa de TV, 563 indivíduos foram entrevistados por telefone com relação ao canal em que estavam sintonizados.
 - d) A fim de avaliar a intenção de voto para presidente dos brasileiros. 122 pessoas foram entrevistadas em Brasília.
- 14-** A parcela da população conveniente escolhida para representá-la é chamada de:
 - a) Variável
 - b) Rol
 - c) Amostra
 - d) Dados brutos
 - e) Nada podemos afirmar, porque a informação é incompleta.

2ª LISTA DE EXERCÍCIOS DE ESTATÍSTICA I

1- Classifique cada uma das variáveis abaixo em qualitativa (nominal / ordinal) ou quantitativa (discreta / contínua).

- a) Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (sim ou não são possíveis respostas para essa variável)
- b) Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos, além de *não sei*).
- c) Perda de peso de maratonistas na Corrida de São Silvestre, em quilos.
- d) Intensidade da perda de peso de maratonistas na Corrida de São Silvestre (leve, moderada, forte)
- e) Grau de satisfação da população brasileira com relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
- f) Cor dos cabelos dos alunos de uma escola
- g) Precipitação pluviométrica, durante um ano em uma estação meteorológica de Vitória.
- h) Raça dos alunos de uma certa escola.
- i) Número de ações negociadas na bolsa de valores de São Paulo.
- j) Número de filhos de casais residentes em Vila Velha.
- K) Salário dos funcionários de uma empresa.
- l) Diâmetro externo de peças produzidas por certa máquina.
- m) Número de peças produzidas por hora por certa máquina.
- n) Índice de liquidez das indústrias de Cariacica.
- o) Pontos obtidos em cada jogada de um dado.
- p) Sexo dos filhos dos casais residentes em Viana.
- q) Número de exemplares dos livros da biblioteca da Univila.

2- Ao nascer, os bebês são pesados e medidos, para se saber se estão dentro das tabelas de peso e altura esperados. Estas duas variáveis são:

- a) Qualitativas
- b) Ambas discretas
- c) ambas contínuas
- d) contínua e discreta, respectivamente.
- e) discreta e contínua, respectivamente.

3ª LISTA DE EXERCÍCIOS DE ESTATÍSTICA I

1- Suponha que um questionário foi aplicado aos alunos do 3º período do curso de administração da UNIVILA fornecendo as seguintes informações:

Id: Identificação do aluno

Turma: Turma em que o aluno foi alocado (A ou B)

Sexo: F se feminino, M se masculino.

Idade: Idade em anos

Alt: Altura em metros

Peso: Peso em quilogramas

Filhos: Número de filhos na família

Fuma: Hábito de fumar, sim ou não.

Toler: Tolerância ao cigarro:

(I) Indiferente, (P) Incomoda pouco e (M) Incomoda muito.

Exerc: Horas de atividade física, por semana.

Cine: Número de vezes que vai ao cinema por semana.

OpCine: Opinião a respeito das salas de cinema na cidade:

(B) Regular a boa e (M) muito boa.

TV: Horas gastas assistindo TV, por semana.

OpTV: Opinião a respeito da qualidade da programação na TV:

(R) Ruim, (M) média, (B) Boa e (N) não sabe.

Com base na tabela dos dados brutos abaixo, faça uma tabela de distribuição de frequência e um gráfico adequado para cada variável apresentada.

Informações de questionário estudantil - dados brutos													
Id	Turma	Sexo	Idade	Altura	Peso	Filhos	Fuma	Toler.	Exer.	Cine	OpCine	TV	OpTV
1	A	F	17	1,60	50,5	2	NÃO	P	0	1	B	16	R
2	A	F	18	1,69	55,0	1	NÃO	M	0	1	B	7	R
3	A	M	18	1,85	72,8	2	NÃO	P	5	2	M	15	R
4	A	M	25	1,85	80,9	2	NÃO	P	5	2	B	20	R
5	A	F	19	1,58	55,0	1	NÃO	M	2	2	B	5	R
6	A	M	19	1,76	60,0	3	NÃO	M	2	1	B	2	R
7	A	F	20	1,60	58,0	1	NÃO	P	3	1	B	7	R
8	A	F	18	1,64	47,0	1	SIM	I	2	2	M	10	R
9	A	F	18	1,62	57,8	3	NÃO	M	3	3	M	12	R
10	A	F	17	1,64	58,0	2	NÃO	M	2	2	M	10	R
11	A	F	18	1,72	70,0	1	SIM	I	10	2	B	8	N
12	A	F	18	1,66	54,0	3	NÃO	M	0	2	B	0	R
13	A	F	21	1,70	58,0	2	NÃO	M	6	1	M	30	R
14	A	M	19	1,78	68,5	1	SIM	I	5	1	M	2	N
15	A	F	18	1,65	63,5	1	NÃO	I	4	1	B	10	R
16	A	F	19	1,63	47,4	3	NÃO	P	0	1	B	18	R
17	A	F	17	1,82	60,0	1	NÃO	P	3	1	B	10	N
18	A	M	18	1,80	85,2	2	NÃO	P	3	4	B	10	R
19	A	F	20	1,60	54,5	1	NÃO	P	3	2	B	5	R
20	A	F	18	1,68	52,5	3	NÃO	M	7	2	B	14	M
21	A	F	21	1,70	60,0	2	NÃO	P	8	2	B	5	R
22	A	F	18	1,65	58,5	1	NÃO	M	0	3	B	5	R
23	A	F	18	1,57	49,2	1	SIM	I	5	4	B	10	R
24	A	F	20	1,55	48,0	1	SIM	I	0	1	M	28	R
25	A	F	20	1,69	51,6	2	NÃO	P	8	5	M	4	N
26	A	F	19	1,54	57,0	2	NÃO	I	6	2	B	5	R
27	B	F	23	1,60	63,0	2	NÃO	M	8	2	M	5	R
28	B	F	18	1,62	52,0	1	NÃO	P	1	1	M	10	R
29	B	F	18	1,57	49,0	2	NÃO	P	3	1	B	12	R
30	B	F	25	1,65	59,0	4	NÃO	M	1	2	M	2	R
31	B	F	18	1,61	52,0	1	NÃO	P	2	2	M	6	N
32	B	M	17	1,71	73,0	1	NÃO	P	1	1	B	20	R
33	B	F	17	1,65	56,0	3	NÃO	M	2	1	B	14	R
34	B	F	17	1,67	58,0	1	NÃO	M	4	2	B	10	R
35	B	M	18	1,73	97,0	1	NÃO	M	7	1	B	25	B
36	B	F	18	1,60	47,0	1	NÃO	P	5	1	M	14	R
37	B	M	17	1,70	95,0	1	NÃO	P	10	2	M	12	N
38	B	M	21	1,85	84,0	1	SIM	I	6	4	B	10	R
39	B	F	18	1,70	60,0	1	NÃO	P	5	2	B	12	R
40	B	M	18	1,73	73,0	1	NÃO	M	4	1	B	2	R
41	B	F	17	1,70	55,0	1	NÃO	I	5	4	B	10	B
42	B	F	23	1,45	44,0	2	NÃO	M	2	2	B	25	R
43	B	M	24	1,76	75,0	2	NÃO	I	7	0	M	14	N
44	B	F	18	1,68	55,0	1	NÃO	P	5	1	B	8	R
45	B	F	18	1,55	49,0	1	NÃO	M	0	1	M	10	R
46	B	F	19	1,70	50,0	7	NÃO	M	0	1	B	8	R
47	B	F	19	1,55	54,5	2	NÃO	M	4	3	B	3	R
48	B	F	18	1,60	50,0	1	NÃO	P	2	1	B	5	R
49	B	M	17	1,80	71,0	1	NÃO	P	7	0	M	14	R
50	B	M	18	1,83	86,0	1	NÃO	P	7	0	M	20	B

2- Complete as tabelas abaixo, usando como população os alunos de sua turma.

a) Variáveis Qualitativas Nominais

Distribuição dos alunos do 3º período de administração 2004/1

Sexo	F	fr
Masculino		
Feminino		
Total		

FONTE: Secretaria da Faculdade UNIVILA

b) Variáveis Qualitativas Ordinais

Distribuição dos alunos do 3º período de administração 2004/1

Segundo o grau de interesse pelo cinema nacional

Grau de interesse	F	fr
MUITO		
RAZOÁVEL		
POUCO		
NADA		
Total		

c) Variáveis Quantitativas Discretas

Distribuição dos alunos do 3º período de
Administração 2004/1 segundo o número de irmãos

Número de irmãos	F	F
0		
1		
2		
3		
Total		

e) Variáveis quantitativas contínuas

Distribuição dos alunos do 3º período de administração 2004/1 segundo a altura

Amplitude Total = $\text{máx} - \text{min} =$

Número de Classes: Raiz quadrada de $n =$

Amplitude das classes = AT: n de classes =

3 – As notas obtidas por 50 alunos de uma classe foram:

1	2	2	2	2	2	3	3	3	3
3	3	3	4	4	4	4	4	4	5
5	5	5	5	5	5	6	6	6	6
6	6	6	6	6	7	7	7	7	7
7	8	8	8	8	8	8	9	9	9

Complete a distribuição de frequência abaixo:

Notas	f	fr	F
0 – 2			
2 – 4			
4 – 6			
6 – 8			
8 – 10			
Total			

4 – Conhecidas notas de 50 alunos:

84	68	33	52	47	73	68	61	73	77
74	71	81	91	65	55	57	35	85	88
59	80	41	50	53	65	76	85	73	60
67	41	78	56	94	35	45	55	64	74
65	94	66	48	39	96	89	98	42	54

Obtenha uma distribuição de frequência e construa um histograma para representar estas notas.

5 – Os resultados do lançamento de um dado 50 vezes foram os seguintes:

6	5	2	6	4	3	6	2	6	5
1	6	3	3	5	1	3	6	3	4
5	4	3	1	3	5	4	4	2	6
2	2	5	2	5	1	3	6	5	1
5	6	2	4	6	1	5	2	4	3

Forneça uma distribuição de frequência e construa um gráfico de barras para essa variável.

6 – Considerando as notas de um teste de inteligência aplicado a 100 alunos:

64	78	66	82	74	103	78	86	103	87
73	95	82	89	73	92	85	80	81	90
78	86	78	101	85	98	75	73	90	86
86	84	86	76	76	83	103	86	84	85
82	90	83	81	85	72	81	96	81	85
68	96	86	70	72	74	84	99	81	89
71	73	63	105	74	98	78	78	83	96
95	94	88	62	91	83	98	93	83	76
94	75	67	95	108	98	71	92	72	73

Forneça uma distribuição de frequência e construa um gráfico adequado para esta variável.

7 – A tabela abaixo apresenta as vendas diárias de um aparelho elétrico, durante um mês, por uma firma comercial.

14	12	11	13	14	13
12	14	13	14	11	12
12	14	10	13	15	11
15	13	16	17	14	14

Forneça uma distribuição de frequência e construa um gráfico adequado para esta variável.

8 – Complete a tabela abaixo:

classes	f	fr	F
0 - 8	4		
8 - 16	10		
16 - 24	14		
24 - 32	9		
32 - 40	3		
Total	40		

9 – A tabela abaixo apresenta uma distribuição de frequência das áreas de 400 lotes:

ÁREAS (m ²)	300	400	500	600	700	800	900	1000	1100	1200
Nº de lotes	14	46	58	76	68	62	48	22	6	

Com referencia a essa tabela determine:

- A amplitude total.
- A amplitude de cada classe.
- A frequência da quarta classe.
- A frequência relativa da sexta classe.
- A frequência acumulada da quinta classe.
- O Nº de lotes cuja área não atinja 700m².
- O Nº de lotes cuja área atinja e ultrapasse 800m².
- A percentagem dos lotes cuja área não atinja 600m².
- A percentagem dos lotes cuja área seja maior ou igual a 900m².
- A percentagem dos lotes cuja área é de 500m² no mínimo, mas inferior a 1000m².
- A classe do 72º lote.
- Até que classe estão incluídos 60% dos lotes.

10 – A distribuição abaixo indica o Nº de acidentes ocorridos com 70 motoristas de uma empresa de ônibus:

Nº ACIDENTES	0	1	2	3	4	5	6	7
Nº MOTORISTAS	20	10	16	9	6	5	3	1

Determine:

- O número de motoristas que não sofreram nenhum acidente.
- O número de motoristas que sofreram pelo menos 4 acidentes.
- O número de motoristas que sofreram menos de 3 acidentes.
- O número de motoristas que sofreram no mínimo 3 e no máximo 5 acidentes.
- A percentagem dos motoristas que sofreram no máximo 2 acidentes.

11 – Represente graficamente cada uma das tabelas abaixo:

a)

Produção Brasileira de Petróleo Bruto 1991 - 1993

Anos	Quantidade (1.000 m ³)
1991	36.180,4
1992	36.410,5
1993	37.164,3

Fonte: Petrobrás

b)

Entrega de Gasolina para Consumo 1988 - 1991

Anos	Volume (1.000 m ²)
1988	9.267,7
1989	9.723,1
1990	10.121,3
1991	12.345,4

Fonte: IBGE

c)

**Produção de Ovos de Galinha
Brasil - 1992**

Regiões	Quantidade (1.000 dúzias)
Norte	57297
Nordeste	414.804
Sudeste	984.659
Sul	615.978
Centro-Oeste	126.345

Fonte: IBGE

d)

**Produção de Veículos de Autopropulsão-
Brasil-1993**

Tipos	Quantidade
Automóveis	1.100.278
Comerciais Leves	224.387
Comerciais Pesados	66.771

Fonte: ANFAVEA

e)

Área Terrestre Brasil

Regiões	Relativa (%)
Norte	45,25
Nordeste	18,28
Sudeste	10,85
Sul	6,75
Centro-Oeste	18,86

Fonte: IBGE

f)

**Produção de Ferro Gusa
Brasil - 1993**

Unidade de Federação	Produção (1.000t)
Minas gerais	12.888
Espírito Santo	3.174
Rio de Janeiro	5.008
São Paulo	2.912

Fonte: Instituto Brasileiro de Siderurgia

g)

**Salário dos funcionários do
Supermercado Peg-Pag**

Salários (R\$)	f
500 - 700	8
700 - 900	20
900 - 1100	7
1100 - 1300	5
1300 - 1500	2
1500 - 1700	1
1700 - 1900	1
Total	44

h)

Área de 400 lotes da Praia da Costa

Áreas (m ²)	Nº de lotes
300 - 400	14
400 - 500	46
500 - 600	58
600 - 700	76
700 - 800	68
800 - 900	62
900 - 1000	48
1000 - 1100	22
1100 - 1200	6

Sugestão: Para as letras a), b), c) e d) use gráfico de barras. Para as letras e) e f) use gráfico de setores e para as demais letras use um histograma.

12 – Uma variável qualitativa com três valores (A, B, C) foi observada em cada um dos vinte elementos de uma amostra. Obteve-se:

B	A	A	C	A	B	B	B	A	A
C	A	B	B	A	C	B	B	B	A

- Determine a frequência de cada classe.
- Calcule a frequência relativa de cada classe.
- Construa um gráfico de barras.
- construa um gráfico de setores.

13 – Quinze pacientes de uma clinica de ortopedia foram entrevistados quanto ao número de meses previsto para a fisioterapia, se haverá (S) ou não (N) seqüelas após o tratamento e o grau de complexidade da cirurgia realizada: Alto (A), médio (M) ou baixo (B). Os dados são apresentados na tabela abaixo:

Pacientes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Fisioterapia	7	8	5	6	4	5	7	7	6	8	6	5	5	4	5
Seqüelas	S	S	N	N	N	S	S	N	N	S	S	N	S	N	N
Cirurgia	A	M	A	M	M	B	A	M	B	M	B	B	M	M	A

- a) Classifique cada uma das variáveis
b) Para cada variável, construa uma distribuição de frequência e faça uma representação gráfica.
c) Para o grupo de pacientes que não ficaram com seqüela, faça um gráfico de barras para a variável Fisioterapia. Você acha que essa variável se comporta de modo diferente nesse grupo?

14 – Um grupo de estudantes do ensino médio foi submetido a um teste de matemática resultando em:

Nota	F
0 2	14
2 4	28
4 6	27
6 8	11
8 10	4
Total	84

- a) Construa um histograma
b) Se a nota mínima para passar é 5, qual será a porcentagem de aprovação?

15 – Uma pesquisa com usuários de transporte coletivo na cidade de São Paulo indagou sobre os diferentes tipos usados nas suas locações diárias. Dentre ônibus, metrô e trem, o número de diferentes meios de transporte utilizados foi o seguinte:

2	3	2	1	2	1	2	3	1	1
1	2	3	1	1	1	1	2	1	1
2	2	1	2	1	2	3	1	2	3

- a) Organize uma tabela de frequência.
b) Faça uma representação gráfica.
c) Admitindo que essa amostra represente bem o comportamento do usuário paulistano, você acha que a porcentagem dos usuários que utilizam mais de um tipo de transporte é grande?

16 – Um novo medicamento para cicatrização está sendo testado e um experimento é feito para estudar o tempo (em dias) de completo fechamento em cortes provenientes de cirurgia. Uma amostra em 30 cobaias forneceu os valores:

15	17	16	15	17	14	17	16	16	17
15	18	14	17	15	14	15	16	17	18
17	15	16	14	18	18	16	15	14	18

- a) Organize uma tabela de frequência.
b) Que porcentagem das observações estão abaixo de 16?
c) Classifique como *rápida* as cicatrizações iguais ou inferior a 15 dias e como *lenta* as demais. Faça um gráfico de setores indicando as porcentagens para cada classificação.

17 – Um questionário foi aplicado aos 10 funcionários do setor de contabilidade de uma empresa fornecendo os dados apresentados na tabela.

Funcionário	Curso (completo)	Idade	Salário (R\$)	Anos de Empresa
1	superior	34	1.100,00	5
2	superior	43	1.450,00	8
3	médio	31	960,00	6
4	médio	37	960,00	8
5	médio	24	600,00	3
6	médio	25	600,00	2
7	médio	27	600,00	5
8	médio	22	450,00	2

9	fundamental	21	450,00	3
10	fundamental	26	450,00	3

- Classifique cada uma das variáveis.
- Faça uma representação gráfica para a variável Curso.
- Discuta a melhor forma de construir a tabela de frequência para a variável Idade. Construa uma representação gráfica.
- Repita o item (c) para a variável salário.
- Considerando apenas os funcionários com mais de três anos de casa, descreva o comportamento da variável Salário.

18 – Um grupo de pedagogos estuda a influência da troca de escolas no desempenho de alunos do ensino fundamental. Como parte do levantamento realizado, foi anotado o número de escolas cursadas pelos alunos participantes da pesquisa.

Escolas Cursadas	F
1	46
2	57
3	21
4	15
5	4
Total	143

- Qual é a porcentagem dos alunos que cursaram mais de uma escola?
- Construa o gráfico de barras.
- Classifique os alunos em dois grupos segundo a rotatividade: *alta* para alunos com mais de 2 escolas e *baixa* para os demais. Obtenha a tabela de frequência dessa variável.

19 – O tempo de utilização de caixas eletrônicos depende de cada usuário e das operações efetuadas. Foram coletadas 26 medidas desse tempo (em minutos):

1,1	1,2	1,7	1,5	0,9	1,3	1,4	1,6	1,7	1,6	1,0	0,8	1,5
1,3	1,7	1,6	1,4	1,2	1,2	1,0	0,9	1,8	1,7	1,5	1,3	1,5

- Organize uma tabela de frequência agrupando os dados em classes.
- Construa um histograma.

20 – Foram feitas medidas em operários da construção civil a respeito da taxa de hemoglobina no sangue (em gramas/cm³):

11,1	12,2	11,7	12,5	13,9	12,3	14,4	13,6	12,7	12,6
11,3	11,7	12,6	13,4	15,2	13,2	13,0	16,9	15,8	14,7
13,5	12,7	12,3	13,5	15,4	16,3	12,3	15,2	13,7	14,1

- Construa uma tabela de frequência.
- Construa o histograma.
- Taxas abaixo de 12 ou acima de 16 são consideradas alteradas e requerem acompanhamento médico. Obtenha a tabela de frequência da variável *Acompanhamento Médico* com dois valores sim ou não.

21 – O valor médio de comercialização da saca de milho de 60 quilos na Bolsa de Cereais é apresentado abaixo, em reais, para os últimos 40 meses.

6,1	6,2	6,7	6,5	6,9	6,3	7,4	7,6	7,7	7,6
7,3	7,7	7,6	7,4	7,2	7,2	7,3	7,6	7,5	7,4
7,5	7,7	8,2	8,3	8,1	8,1	8,1	7,9	7,8	7,4
7,5	7,6	7,5	7,6	7,4	7,3	7,4	7,5	7,5	7,4

- Organize os dados em uma tabela de frequência.
- Construa o histograma.

22 – O número de gols marcados no último campeonato da Federação Paulista de Futebol pelos 20 clubes participantes nos seus 38 jogos é uma variável com os seguintes valores:

Clube	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Gols	32	42	73	35	79	57	37	52	35	25	55	70	42	41	68	66	74	29	47	53

- Classifique a variável. Você acha razoável construir uma tabela de frequência de acordo com a classificação dada?
- Construa uma tabela de frequência agrupando as observações em intervalos de classes.
- Obtenha o histograma.
- Que porcentagem dos clubes marcaram mais de 38 gols?

23 – Uma nova ração foi fornecida a suínos recém desmamados e deseja-se avaliar sua eficiência. A ração tradicional dava um ganho de peso ao redor de 3,5 Kg em um mês. A seguir, apresentamos os dados referentes ao ganho, em quilos, para essa nova ração, aplicada durante um mês em 200 animais nas condições acima.

- Construa o histograma.
- Você acha que a nova ração é mais eficiente que a tradicional? Justifique.

Ganho	F
1,0 – 2,0	45
2,0 – 3,0	83
3,0 – 4,0	52
4,0 – 5,0	15
5,0 – 6,0	4
6,0 – 7,0	1
Total	200

24 – num estudo sobre rotatividade de mão de obra na indústria, anotou-se o número de empregos nos últimos 3 anos para operários especializados e não especializados.

- Construa o diagrama de barras correspondente a cada tabela usando a frequência relativa.
- Junte as informações das duas tabelas em uma só e obtenha um diagrama de barras da rotatividade de mão de obra na indústria sem diferenciar a especialização.]
- Você acha que os trabalhadores especializados trocam menos de emprego? Justifique.

Não Especializados	
Empregos	f
1	106
2	222
3	338
4	292
5	164
Total	1122

Especializados	
Empregos	f
1	210
2	342
3	109
4	91
5	35
Total	787

25 – Como parte de uma avaliação médica em uma certa universidade, foi medida a frequência cardíaca dos alunos do primeiro ano. Os dados são apresentados em seguida.

- Obtenha o histograma.
- Frequências cardíacas que estejam abaixo de 62 ou acima de 92 requerem acompanhamento médico. Qual a porcentagem de alunos nessas condições?
- Uma frequência ao redor de 72 batidas por minuto é considerada padrão. Você acha que de modo geral esses alunos se encaixam nesse caso?

Frequência Cardíaca	f
60 – 65	11
65 – 70	35
70 – 75	68
75 – 80	20

80 - 85	12
85 - 90	10
90 - 95	1
95 - 100	3
Total	160

26 – Um exame vestibular para uma faculdade tem 80 questões, sendo 40 de português e 40 de matemática. Para os 20 melhores classificados apresentamos o número de acertos em cada disciplina, em ordem decrescente do total de pontos.

- Organize uma tabela de frequência para cada variável.
- Faça uma representação gráfica das tabelas obtidas em (a).
- Construa a tabela de frequência da variável *Total de pontos*.
- Comente sobre a afirmação: Os aprovados são melhores em português do que em matemática.

Aluno	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Port.	35	35	34	32	31	30	26	26	24	23	23	12	11	20	17	12	14	20	8	10
Mat.	31	29	27	28	28	26	30	28	25	23	21	32	31	20	21	25	20	13	23	20

27 – Vinte baterias para automóveis de uma certa marca foram testadas quanto à sua vida útil. O teste simula a utilização da bateria, acelerando seu desgaste de modo a criar uma réplica da situação real. Os resultados da durabilidade, em meses, são apresentados a seguir:

- Construa o histograma
- Se a amostra acima for considerada representativa do desempenho dessa marca de bateria, quantas, em 1000 fabricadas, serão repostas pelo fabricante, se ele oferece uma garantia de 6 meses?
- Se o fabricante vende cada bateria por 20% acima do preço e custo, em 1000 baterias fabricadas, descartadas as que repõe, quanto será seu lucro por bateria em função do preço de custo?

Durabilidade	f_r
0 - 3	0,02
3 - 6	0,05
6 - 9	0,15
9 - 12	0,25
12 - 15	0,30
15 - 20	0,23
Total	1,00

4ª LISTA DE EXERCÍCIOS DE ESTATÍSTICA I

1- Uma escola de 1º grau abriga 124 alunos. Obtenha uma amostra representativa correspondente a 15% da população.

Sugestão: Use a 8ª coluna, a partir da 1ª linha, da Tabela de Números Aleatórios, de cima para baixo.

2 – Em uma escola há 80 alunos. Obtenha uma amostra de 12 alunos.

Sugestão: Use a 2ª coluna, a partir da 4ª linha, da Tabela de Números Aleatórios, de cima para baixo.

3 – Uma população é formada por 140 notas resultantes da aplicação de um teste de inteligência:

62	129	95	123	81	93	105	95	96	80	87	110	139	75
123	60	72	86	108	120	57	113	65	108	90	137	74	106
109	84	121	60	128	100	72	119	103	128	80	99	149	85
77	91	51	100	63	107	76	82	110	63	131	65	114	103
104	107	63	117	116	86	115	62	122	92	102	113	74	78
69	116	82	95	72	121	52	80	100	85	117	85	102	106
94	84	123	42	90	91	81	116	73	79	98	82	69	102
100	79	101	98	110	95	69	77	91	95	74	90	134	94
79	92	73	83	74	125	101	82	71	75	101	102	78	108
125	56	86	98	106	72	117	89	99	86	82	57	106	90

Obtenha uma amostra Sistemática formada de 26 elementos, tomando, inicialmente, a primeira linha da esquerda para a direita.

4 – O diretor de uma escola, na qual estão matriculados 280 meninos e 320 meninas, desejoso de conhecer as condições de vida extra-escolar de seus alunos e não dispondo de tempo para entrevistar todas as famílias, resolveu fazer um levantamento, por amostragem, em 10% dessa clientela. Obtenha, para esse diretor os elementos componentes da amostra.

5 – A cidade de Margaridinha apresenta o seguinte quadro relativo às suas escolas de 1º grau:

ESCOLAS	Nº DE ESTUDANTES	
	MASCULINO	FEMININO
A	80	95
B	102	120
C	110	92
D	134	228
E	150	130
F	300	290
Total	876	955

Obtenha uma amostra proporcional estratificada de 120 estudantes.

6 – É dada uma população construída pelas 12 primeiras letras do alfabeto. Explique o que você faria para obter uma amostra sistemática de três elementos.

7 – É dada uma população infinita constituída apenas por números 2, 4 e 6. escreva todas as amostras causais possíveis de tamanho dois. Note que são possíveis amostras de números iguais e que são diferentes amostras de mesmos números, mas em ordem diferente.

8 – Pretende-se obter uma amostra dos alunos de uma universidade para estimar a proporção que tem trabalho remunerado. Qual é a população em estudo? Qual é o parâmetro que se quer estimar? Você acha que se obterá uma boa amostra dos alunos no restaurante universitário? No ponto de ônibus mais próximo? Nas portas das salas de aula? Ou você tem outra alternativa melhor?

9 – Para estimar o número médio de pessoas em um domicílio, um pesquisador obteve uma amostra sistemática de 1000 domicílios. No entanto, mesmo fazendo varias visitas, o entrevistador não encontrou pessoas em 147 deles. O pesquisador obteve então uma segunda amostra e quando o entrevistador completou a visita aos 147 domicílios que compunham a amostra de 1000, analisou os dados. Havia sido contadas 3087 pessoas. O pesquisador considerou então que o numero médio de pessoas em domicílio é 3,1. O que você acha?

10 – Um fiscal da Vigilância Sanitária precisa verificar se as farmácias da cidade estão cumprindo um novo regulamento. A cidade tem 33 farmácias, mas como a fiscalização demanda muito tempo, o fiscal resolveu optar por uma amostragem. Para escolher a amostra, o fiscal estratificou a população de farmácias de acordo com o volume de vendas. Existem 3 farmácias de uma grande cadeia, 10 de cadeias menores e 20 farmácias pequenas, de proprietários locais. O fiscal decide visitar as três farmácias da grande cadeia, quatro das cadeias menores e três farmácias pequenas. O cumprimento do regulamento, evidentemente desconhecido do fiscal, esta apresentado na tabela a seguir. Com base nessa tabela,

- Sorteie uma amostra estratificada para o local, de acordo com o que ele planejou;
- Estime, com base na amostra, a proporção de farmácias que estão cumprindo o regulamento;
- Com base nos dados da população, estime o parâmetro;
- Você obteve uma boa estimativa?

Distribuição dos dados sobre cumprimento a um regulamento, segundo o estrato.			
Estrato A		Estrato B	
1. sim	1. não	1. sim	11. sim
2. sim	2. sim	2. sim	12. sim
3. não	3. não	3. não	13. não
	4. não	4. sim	14. sim
	5. sim	5. não	15. sim
	6. não	6. não	16. não
	7. sim	7. não	17. não
	8. não	8. sim	18. não
	9. não	9. não	19. sim
	10. sim	10. não	20. sim

