Disciplina: Estatística
Professora: Danielle

Conceitos Básicos

Estatística: Parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para utilização dos mesmos na tomada de decisões. A estatística tem como objetivo o estudo dos fenômenos coletivos.

Processos estatísticos de abordagem:

- Estimação: é uma avaliação indireta de um parâmetro, com base em um estimador através do cálculo de probabilidades;
- Censo: é uma avaliação direta de um parâmetro, utilizando-se todos os componentes da população.

Dados: Os fatos e números que são coletados, analisados e interpretados.

Dados brutos: Seqüência de valores numéricos não organizados, obtidos diretamente da observação de um fenômeno coletivo.

Rol: Seqüência ordenada dos dados brutos.

Conjunto de dados: Todos os dados coletados em um determinado estudo.

Elementos: As entidades sobre as quais os dados são coletados.

Variável: Uma característica de interesse para os elementos.

Observação: O conjunto de medidas obtidas para um único elemento.

Dados qualitativos: Dados que fornecem rótulos ou nomes para uma característica de um elemento.

Variável qualitativa: Uma variável com dados qualitativos.

Dados quantitativos: Dados que indicam a quantidade de alguma coisa. Dados quantitativos são sempre numéricos.

Variável quantitativa: Uma variável com dados quantitativos.

Dados de seção transversal: Dados coletados no mesmo ou aproximadamente no mesmo ponto de tempo.

Dados de série histórica: Dados coletados em vários períodos de tempo sucessivos.

Estatística descritiva: Métodos tabulares, gráficos e numéricos usados para sintetizar dados.

População: Conjunto de todos os elementos de interesse em um determinado estudo.

Amostra: Um subconjunto da população.

Inferência estatística: Processo de utilizar dados obtidos a partir de uma amostra para fazer estimativas ou testar hipóteses sobre as características de uma população.

Introdução Geral à compreensão Estatística

Panorama Histórico:

Objetivos:

- Localizar a Estatística na linha do tempo;

- Relacionar a palavra estatística com suas origens etimológicas;

- Identificar situações práticas às quais a palavra estatística poderia ser aplicada com propriedade;

Origem da Estatística:

1. Embora a palavra Estatística ainda não existisse, há indícios de que em 3000 A. C. já se faziam censos na Babilônia, China e Egito (taxação e impostos);
2. A própria Bíblia leva-nos a essa recuperação histórica. O livro quarto do velho testamento começa com uma instrução a Moisés: fazer um levantamento dos homens de Israel aptos para guerrear;
3. Na época do Imperador César Augusto, saiu um edito para que se fizesse o censo em todo império Romano;

   Censo = censere que em latim significa taxar.

4. A palavra Estatística vende status (estado em latim). Sob essa palavra acumularam-se descrições e dados relativos ao Estado. Em 1085, Guilherme, o conquistador, ordenou que se fizesse um levantamento estatístico na Inglaterra. Esse levantamento deveria incluir informações sobre terras, proprietários, uso das terras, empregados, animais e serviria de base para o cálculo de impostos;
5. No século XVII ganhou destaque na Inglaterra a partir das tábuas de mortalidade, a aritmética política de John Graunt, que consistiu de exaustivas análises de nascimentos e mortes. Dessas análises resultou a conclusão (entre outras) de que a porcentagem de nascimentos de crianças do sexo masculino era ligeiramente superior à de crianças do sexo feminino;

51 % *meninos e 49% meninas*

As tábuas de mortalidade usadas pelas companhias de seguro originarm-se de estudos como esse.

6. O verbete "statistics" aparece na enciclopédia Britânica em 1797.

Por que é importante a Pesquisa Estatística e o uso da Estatística?
Os estágios da Pesquisa.

Por que o pesquisados usa Estatística?

Todos nós temos um pouco de cientistas. Quase que diariamente, temos "palpites" com relação a acontecimentos em nossas vidas, a fim de prever o que acontecerá em novas situações ou experiências.

Ex.:

- investir na bolsa de valores;

- votar em algum candidato que prometesse resolver problemas nacionais;

- jogar nos cavalos;

- tomar um remédio para resfriado;

- tentar adivinhar o que os examinadores irão perguntar nos exames.

Às vezes ganhamos; às vezes perdemos. Nem todas as previsões acabam se tornando realidade.

A natureza da pesquisa

Os cientistas têm idéias sobre a natureza da realidade (idéias que ele denomina hipótese) e testa suas idéias através de pesquisa sistemática.

Exemplos:

Hipótese: crianças socialmente isoladas assistem mais TV do que crianças bem integradas em seu grupo.

Pesquisa: Levantamento em que fossem feitas perguntas tanto à crianças socialmente isoladas quanto às bem integradas com relação ao n° de horas que elas passam diante do vídeo.

Hipótese: Família constituída por um só genitor (pai ou mãe ausente) gera maIOr delinqüência do que família constituída por ambos os genitores

Pesquisa: Entrevista de amostras de delinqüentes e não delinqüentes que procedem de famílias de ambos os tipos.

Por que testar hipóteses?

É de modo geral desejável - necessária até - testar nossas hipóteses sobre a natureza da realidade mesmo aquela que parecem verdadeiras, lógicas ou evidentes.

Os nossos *testes* de *bom* senso diário, geralmente baseiam-se em idéias pré-concebidas muito estreitas, se não tendenciosas, e em experiências pessoais, o que pode levar-nos a aceitar conclusões não-válidas a respeito da natureza dos fenômenos.

Exemplo: grande n° de soldados durante a 2^3 Guerra Mundial Hipóteses:

1. Homens de nível educacional mais alto apresentam maior quantidade de sintomas psiconeuróticos do que aqueles que haviam recebido menor instrução;
2. Homens procedentes do meio rural mantiveram-se mais bem-humorados durante o período de guerra do que os soldados recrutados nos meios urbanos;
3. A capacidade dos soldados procedentes do sul para suportar o clima quente das llhas dos Mares do Sul era maior que a dos soldados vindos do Norte;
4. À medida que a guerra prosseguia, os homens sentiam-se mais ansiosos para retomar aos E.D.A. do que após a derrocada dos alemães.

- Você teria sido capaz de prever *estes* resultados em suas experiências diárias?

- Você acha que valeu a pena testá-Ias?

Se você acredita que essas relações eram tão óbvias que qualquer teste teria sido dispensável, saiba que cada item representa a conclusão oposta ao que realmente se verificou.

Estágios da Pesquisa

1 - Hipótese

2 - Questionário ou esquema de uma entrevista

3 - Dados coletados

4 – Dados analisados em conjunto com as hipóteses iniciais

5 – Resultados são interpretados e comunicados

População e Amostra

1. Variáveis

A cada fenômeno corresponde um número de resultados possíveis. Por exemplo:

a) Para o fenômeno "estado civil" podemos ter como resultados possíveis: solteiro, casado, viúvo, divorciado, dentre outros.
b) Para o fenômeno "número de filhos" há um número de resultados possíveis expresso através de números naturais: 0, 1, 2, ..., n.
c) Para o fenômeno "estatura" temos uma situação diferente, pois, se levarmos em consideração a precisão da medida, os resultados podem

tomar um número infinito de valores numéricos dentro de um determinado intervalo.

Variável é, convencionalmente, o conjunto de resultados possíveis de um fenômeno ou as características que podem ser observadas (ou medidas) em cada elemento do conjunto todo que iremos pesquisar.

As variáveis podem ser:

a) **qualitativa:** quando seus valores são expressos por atributos ou qualidades: sexo (masculino, feminino); estado civil (solteiro, casado, divorciado, viúvo); nível de satisfação (muito satisfeito, satisfeito, pouco satisfeito, insatisfeito). Neste último exemplo, como existe uma ordenação do nível de satisfação nas quatro opções, dizemos que a variável é *qualitativa ordinal*.

b) **quantitativa:** quando os seus valores são expressos em números. Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de *variável contínua*; uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de *variável discreta*. Assim, o número de filhos de uma pessoa pode assumir qualquer um dos valores do conjunto 0, 1, 2, ..., n, mas nunca poderá assumir um valor como 2,5 ou 3,72. Logo, o número de filhos que uma determinada pessoa possui é uma *variável quantitativa discreta*. Já a altura dessas pessoas é uma variável quantitativa contínua, pois elas tanto podem medir 1,70m, como 1,75m, ou ainda 1,754m, só vai depender do valor da precisão da medida.

2. População e Amostra

Chamamos de *população* o conjunto de elementos que formam o universo de nosso estudo e que são passíveis de serem observados. Uma parte desses elementos é dita uma *amostra*. De maneira mais simplificada, conceituamos população como sendo o conjunto de todos os itens (pessoas, coisas, objetos) que interessam ao estudo de um fenômeno coletivo segundo alguma característica. Entendemos por amostra qualquer subconjunto não vazio desta população.

Por exemplo: no fenômeno coletivo eleição para governador no Estado de Rondônia, a população é o conjunto de todos os eleitores habilitados no Estado de

Rondônia. A amostra é um grupo de 1000 ou 2000 eleitores selecionados em todo o Estado.

Mas, para que as conclusões estejam corretas, é necessário garantir que a amostra seja representativa da população, isto é, a amostra deve possuir as mesmas características básicas da população, no que diz respeito ao fenômeno que desejamos pesquisar. É preciso, pois, que a amostra ou as amostras que vão ser usadas sejam obtidas por processos adequados.

3. Amostragem

A amostragem é naturalmente usada em nossa vida diária. Por exemplo, para verificar o tempero de um alimento em preparação, podemos provar (observar) uma pequena amostra (porção) deste alimento. Estamos fazendo uma amostragem, ou seja, extraindo do todo (população) uma parte (amostra), com o propósito de avaliarmos a qualidade de tempero do alimento.

Em grandes populações torna-se interessante à realização de uma amostragem, ou seja, a seleção de uma parte da população para ser observada. Para um leigo em Estatística, é surpreendente como uma amostra de 3.000 eleitores forneça um perfil bastante preciso sobre a preferência de todo o eleitorado, na véspera de uma eleição presidencial. Mas isto só é possível se esta amostra for extraída sob um rigoroso plano de amostragem, capaz de garantir a sua representatividade.

A amostragem é, então, nada mais que uma técnica para recolher amostras, que garante, tanto quanto possível, o acaso da escolha.

Dessa forma, cada elemento da população passa a ter a mesma chance de ser escolhido, o que garante à amostra o caráter de representatividade, e isto é muito importante, pois, como vimos, nossas conclusões relativas vão ser baseadas nos resultados obtidos nas amostras desta população.

Por que amostragem?

Citaremos quatro razões para o uso de amostragem em levantamentos de grandes populações

1) *Economia*: Em geral, torna-se bem mais econômico o levantamento de somente uma parte da população.

2) *Tempo*: Numa pesquisa eleitoral, a três dias de uma eleição presidencial, não haveria tempo suficiente para pesquisar toda a população de leitores do país, mesmo que houvesse recursos financeiros em abundância.
3) *Confiabilidade dos dados*: Quando se pesquisa um número reduzido de elementos, pode-se dar mais atenção aos casos individuais, evitando erros nas respostas.
4) *Operacionalidade*: É mais fácil realizar operações de pequena escala. Um dos problemas típicos nos grandes censos é o controle de entrevistadores.

Quando o uso de amostragem não é interessante?

Citaremos três situações em que pode não valer a pena a realização de amostragem:

1) *População pequena*: Se a população for pequena, para termos uma amostra capaz de gerar resultados precisos para os parâmetros da população, necessitamos de uma amostra relativamente grande (em torno de 80 % da população). Geralmente é mais relevante o tamanho absoluto da amostra do que a percentagem que ela representa na população.
2) *Características de fácil mensuração*: Talvez a população não seja tão pequena, mas a variável que se quer observar é de tão fácil mensuração que não compensa investir num plano de amostragem.
3) *Necessidade de alta precisão*: É o caso do censo demográfico, que é realizado a cada dez anos pelo IBGE. Nesta pesquisa são estudadas várias características da população brasileira que são fundamentais para o planejamento do país.

Para fazermos um plano de amostragem devemos ter bem definidos os objetivos da pesquisa, a população a ser amostrada, bem como os parâmetros que precisamos estimar para atingir os objetivos da pesquisa. Estudaremos, inicialmente, algumas formas de seleção dos elementos que irão compor a amostra. Posteriormente discutiremos a questão do tamanho da amostra.

3.1 Amostragem casual ou aleatória simples

Para a seleção de uma amostra aleatória simples precisamos ter uma lista completa dos elementos da população. Este tipo de amostragem consiste em selecionar a amostra através de um sorteio, sem restrições. É equivalente a um sorteio lotérico

Na prática, a amostragem casual pode ser realizada numerando-se a população de 1 a n e sorteando-se, a seguir, por meio de um dispositivo aleatório qualquer, k números dessa seqüência, os quais corresponderão aos elementos pertencentes à amostra.

A amostragem aleatória simples tem a seguinte propriedade: *qualquer subconjunto da população, com o mesmo número de elementos, tem a mesma probabilidade de fazer parte da amostra*. Em particular, temos que *cada elemento da população tem a mesma probabilidade de pertencer à amostra*.

Exemplo:
Seja uma população com N elementos. Uma forma de extrair uma amostra aleatória simples de tamanho n, sendo $n < N$, é identificar os elementos da população em pequenos pedaços de papel e retirar, ao acaso, n pedaços (considere que o sorteio seja feito sem reposição).
Quando o número de elementos da amostra é grande, esse tipo de sorteio torna-se muito trabalhoso. A fim de facilitá-lo, foi elaborada uma tabela — **Tabela de Números Aleatórios** — , construída de modo que os dez algarismos (de 0 a 9) são distribuídos ao acaso nas linhas e colunas.

3.2 Amostragem estratificada

A técnica da amostragem estratificada consiste em dividir a população em subgrupos, que são denominados estratos. Estes estratos devem ser internamente mais homogêneos do que a população toda, com respeito às variáveis em estudo. Por exemplo, para estudar o interesse dos funcionários, de uma grande empresa, em realizar um programa de treinamento, podemos estratificar esta população por nível de instrução, ou pelo nível hierárquico, ou ainda pelo setor de trabalho. Devemos escolher um critério de estratificação que forneça estratos bem homogêneos, com respeito ao que se está estudando. Neste contexto, um prévio conhecimento sobre a população em estudo é fundamental.

Sobre os diversos estratos da população, são realizadas seleções aleatórias, de forma independente. A amostra completa é obtida através da agregação das amostras de cada estrato.

3.2.1 Amostragem Proporcional Estratificada

Neste caso particular de amostragem estratificada, a proporcionalidade do tamanho de cada estrato da população é mantida na amostra. Por exemplo: se um estrato corresponde a 20 % do tamanho da população, ele também deve corresponder a 20% do tamanho da amostra.

A amostragem estratificada proporcional garante que cada elemento da população tem a mesma probabilidade de pertencer a amostra.

Exemplo:

Com o objetivo de levantar o estilo de liderança preferida pela comunidade de uma escola, vamos realizar um levantamento por amostragem. A população é composta por 10 professores, 10 servidores técnico-administrativos e 30 alunos. Supondo que a preferência, quanto ao estilo de liderança, possa ser relativamente homogênea dentro de cada categoria, vamos realizar uma amostragem estratificada proporcional por categoria, para obter uma amostra global de tamanho n = 10. A tabela seguinte mostra as relações de proporcionalidade:

Tabela 2.1 Cálculo do tamanho da amostra em cada estrato

ESTRATO	Proporção na População	Tamanho no subgrupo na Amostra
Professores	10/50 = 0,20 ou 20%	20% de 10 = 2
Servidores	10/50 = 0,20 ou 20 %	20% de 10 = 2
Alunos	30/50 = 0,60 ou 60%	60% de 10 = 6

Isto significa que a amostra deve conter 2 professores, 2 servidores e 6 alunos, num total de 10 pessoas. Para selecionar qual professor, ou qual servidor participará da amostra pode-se escolher o método do sorteio ou da tabela de números aleatórios.
Desde que, no problema em estudo, os estratos formam subgrupos mais homogêneos do que a população como um todo, uma amostra estratificada proporcional tende a gerar resultados mais precisos, quando comparada com uma amostra aleatória simples.

3.2.2 Amostragem Estratificada Uniforme:

Seleciona-se a mesma quantidade de elementos em cada estrato. No exemplo anterior, para selecionar uma amostra estratificada uniforme de, digamos, n = 12 indivíduos da comunidade da escola, devemos selecionar 4 indivíduos de cada categoria.

A amostragem estratificada uniforme costuma ser usada em situações em que o maior interesse é obter estimativas separadas para cada estrato, ou ainda, quando se deseja comparar os diversos estratos.
É importante observar que na fase de análise dos dados deve-se levar em conta o planejamento amostral utilizado.

3.3 Amostragem sistemática

Quando os elementos da população já se acham ordenados, não há necessidades de construir o 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 esse tipo de amostragem denominamos sistemática.
Assim, 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.

Exemplo:

Suponhamos uma rua contendo novecentos prédios, dos quais desejamos obter uma amostra formada de cinqüenta prédios. Podemos, neste caso, usar o seguinte procedimento: como $\frac{900}{50}$ = 18, escolhemos por sorteio casual um número de 1 a 18 (inclusive), o qual indicaria o primeiro elemento sorteado para a amostra; os demais elementos seriam periodicamente considerados de 18 em 18. Assim, se o número sorteado fosse o 4, tomaríamos, pelo lado direito da rua, o 4º prédio, o 22º, o 40º, até voltarmos ao início da rua, pelo lado esquerdo.

3.4 Amostragem de conglomerados

Ao contrário da amostragem estratificada, a amostragem de conglomerados tende a produzir uma amostra que gera resultados menos precisos, quando comparada com uma amostra aleatória simples de mesmo tamanho. Contudo, seu custo financeiro tende a ser bem menor.

Chamamos de conglomerado a um agrupamento de elementos da população. Por exemplo, considere o problema de selecionar uma amostra de domicílios de uma cidade. Podemos tomar as ruas como conglomerados.

Este tipo de amostragem consiste, num primeiro estágio, em selecionar conglomerados de elementos. Num segundo estágio, faz-se uma nova seleção, tomando amostras de elementos dos conglomerados extraídos no primeiro estágio. Todas as seleções devem ser aleatórias.

4. Exercícios

a) Considerando a população apresentada abaixo, extraia uma amostra aleatória simples de n = 10 funcionários. Use a segunda linha da tabela de números aleatórios que está em anexo.

POPULAÇÂO: funcionários da empresa

Aristóteles	Anastácia	Arnaldo	Bartolomeu	Bernardino
Cardoso	Carlito	Cláudio	Ermílio	Ercílio
Ernestino	Endevaldo	Francisco	Felício	Fabrício

Geraldo	Gabriel	Getúlio	Hiraldo	José de Souza
Joana	Joaquim	Joaquina	José da Silva	João
Josefa	Josefina	Maria José	Maria Cristina	Mauro
Paula	Paulo César			

b) Em uma escola existem 250 alunos, sendo 35 na 1ª série, 32 na 2ª, 30 na 3ª, 28 na 4ª, 35 na 5ª, 32 na 6ª, 31 na 7ª e 27 na 8ª. Obtenha uma amostra de 40 alunos e preencha o seguinte quadro:

Séries	**População**	**Cálculo Proporcional**	**Amostra**
1ª	35	$\frac{35 \times 40}{250} = 5,6$	6
2ª			
3ª			
4ª	28		
5ª			6
6ª			
7ª		$\frac{31 \times 40}{250} =$	
8ª			
Total	250	--	40

c) Classifique as variáveis em qualitativas ou quantitativas

I) Universo: alunos de uma escola
Variável: cor dos cabelos

II) Universo: casais residentes em uma cidade
Variável: número de filhos

III) Universo: as jogadas de um dado
Variável: o ponto obtido em cada jogada

IV) Universo: peças produzidas por certa máquina
Variável: número de peças produzidas por hora

V) Universo: peças produzidas por certa máquina
Variável: diâmetro externo

d) Diga quais das variáveis abaixo são discretas e quais são contínuas:

I) População: estação meteorológica de uma cidade
Variável: precipitação pluviométrica, durante o ano

II) População: Bolsa de valores de São Paulo
Variável: número de ações negociadas

III) População: Funcionários de uma empresa
Variável: salários

IV) População: segmento de reta
Variável: comprimento

V) População: Aparelhos produzidos em uma linha de montagem
Variável: número de defeitos por unidade.

e) Usando a primeira coluna da tabela de números aleatórios, extraia uma amostra aleatória simples de 4 (quatro) letras do alfabeto da língua portuguesa.

f) Com o objetivo de levantar o estilo de liderança preferida pela comunidade de uma escola, vamos realizar um levantamento por amostragem. A população é composta por 10 professores, 10 servidores técnico-administrativos e 30 alunos. Supondo que a preferência, quanto ao estilo de liderança, possa ser relativamente homogênea dentro de cada categoria, realize uma amostragem estratificada proporcional por categoria, para obter uma amostra global de tamanho n = 10. Utilize a primeira linha da tabela de números aleatórios para formar a amostra. Para resolver o problema considere a seguinte tabela de identificação:

POPULAÇÃO

Professores: P1 P2 P3 P4 P5 P6 P7 P8 P9 P10
Alunos:A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11 A12 A13 A14 A15 A16 A17 A18 A19 A20 A21 A22 A23 A24 A25 A26 A27 A28 A29 30
Servidores: S1 S2 S3 S4 S5 S6 S7 S8 S9 S10

Tabela de Números Aleatórios

9808624826	4524028404	4499908896	3909473407	3544131880
3318516232	4194150949	8943548581	8869541994	3754873043
8095100406	9638270774	2015123387	2501625298	9462461171
7975249140	7196128296	6986102591	7485220539	0038759579
1863332537	9814506571	3101024674	0545561427	7793891936
7402943902	7755732270	97790119	5252758021	8081451748
5417845611	8099337143	0533512969	5612719255	3604090324
1166449883	5207984827	5938171539	0997333440	8846123356
4832477928	3124964710	0229536870	3230757546	1502009994
6907494138	8763791976	3558404401	1051821615	0184876938
0918820097	3282539527	0422086304	8338987374	6427858044
9004585497	5198150654	9893881997	9187076150	6847664659
7318950207	4767725269	6229064464	2712467018	4136182760
7576896490	2097181749	9042912272	9537505871	9382343178
5401644056	6628131003	0068227398	2071453295	0770617813

Séries Estatísticas

1) Tabelas

Um dos objetivos da Estatística é sintetizar os valores que uma ou mais variáveis podem assumir, para que tenhamos uma visão global da variação dessa ou dessas variáveis. E isso ela consegue, inicialmente, apresentando esses valores em **tabelas** e **gráficos**, que irão nos fornecer rápidas e seguras informações a respeito das variáveis em estudo, permitindo-nos determinações administrativas e pedagógicas mais coerentes e científicas.

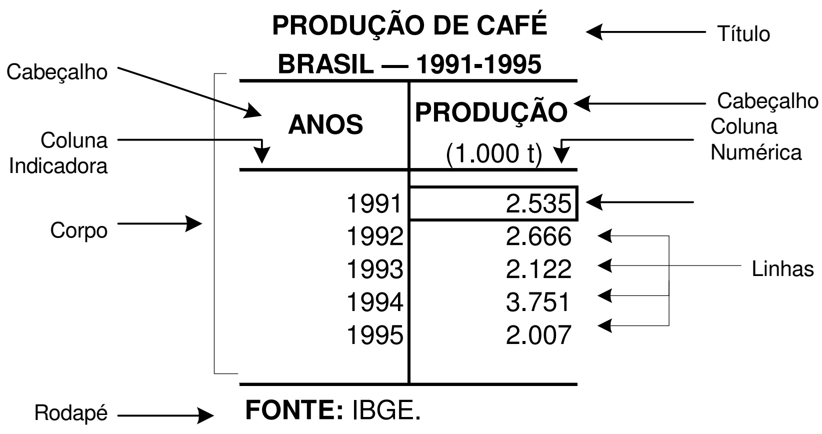
Tabela é um quadro que resume um conjunto de informações.

Uma tabela compõe-se de:

a. **Corpo:** conjunto de linhas e colunas que contem informações sobre a variável em estudo;
b. **Cabeçalho:** parte superior da tabela que especifica o conteúdo das colunas;
c. **Coluna indicadora:** parte da tabela que especifica o conteúdo das linhas;
d. **Linhas:** retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que inscrevem nos seus cruzamentos com as colunas;
e. **Casa ou célula:** espaço destinado a um só número;
f. **Título:** conjunto de informações, as mais completas possíveis, respondendo às perguntas: **O quê?, Quando?, Onde?,** localizado no topo da tabela;

Há ainda a considerar os elementos complementares da tabela, que são a fonte, as notas e as chamadas, colocados, de preferência, no seu rodapé.

Exemplo:



2) Séries estatísticas

Denominamos série estatística toda tabela que apresenta a distribuição de um conjunto de dados estatísticos em função da época, do local ou da espécie.
Daí pode inferir que numa série estatística observamos a existência de três elementos de fatores; o tempo, o espaço e a espécie.
Conforme varie um dos elementos da série, podemos classificá-la em histórica, geográfica e específica.

2.1) Séries Históricas, Cronológicas, Temporais ou marchas

Descrevem os valores da variável, em determinado local, discriminados segundo intervalos de tempos variáveis.

Exemplo:

PREÇO DO ACÉM NO VAREJO
SÃO PAULO — 1989-1994

ANOS	PREÇO MÉDIO (US$)
1989	2.535
1990	2.666
1991	2.122
1992	3.751
1993	2.007
1994	

FONTE: APA.

2.2) Séries Geográficas, Espaciais, Territoriais ou de Localização

Descrevem os valores da variável, em determinado instante, discriminados segundo regiões

Exemplo:

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.

2.3) Séries específicas ou Categóricas

Descrevem os valores da variável, em determinado tempo e local, discriminados segundo especificações ou categorias.

Exemplos:

REBANHOS BRASILEIROS
1992

ESPÉCIES	QUANTIDADE (1.000 cabeças)
Bovinos	154.440,8
Bubalinos	1.423,3
Equinos	549,5
Asininos	47,1
Muares	208,5
Suínos	34.532,2
Ovinos	19.995,9
Caprinos	12.159,6
Coelhos	6,1

FONTE: IBGE.

3) Séries Conjugadas

Muitas vezes temos a necessidade de apresentar, em uma única tabela, a variação de valores de mais de uma variável, ou seja, 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** (linha) e uma **vertical** (coluna).

Exemplo:

TERMINAIS TELEFÔNICOS EM SERVIÇO
1991-93

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

Podem existir, se bem que mais raramente, pela dificuldade de representação, séries compostas de três ou mais entradas.

4) Dados Absolutos e Dados Relativos

Os dados estatísticos resultantes da coleta direta da fonte, sem outra manipulação senão a contagem ou medida, são chamados **dados absolutos**.

Dados Relativos são o resultado de comparações por quociente (razões) que se estabelecem entre dados absolutos e têm por finalidade realçar ou facilitar as comparações entre quantidades.

Os dados relativos podem ser as percentagens, índices, coeficientes ou taxas.

4.1) As percentagens

O emprego da percentagem é de grande valia quando nosso objetivo é destacar a participação da parte no todo.

Exemplo:

MATRÍCULAS NAS ESCOLAS DAS CIDADES A E B - 1995

CATEGORIAS	CIDADE A		CIDADE B	
	Nº DE ALUNOS	%	Nº DE ALUNOS	%
1º grau	19.286	91,0	38.66	91,0
2º grau	1.681	7,9	3.399	8,0
3º grau	234	1,1	424	1,0
Total	21.201	100,0	42.483	100,0

4.2) Índices

Os índices são razões entre duas grandezas tais que uma não inclui a outra.

Exemplos:

1) Densidade Demográfica = $\dfrac{\text{população}}{\text{superfície}}$

2) Renda *per capita* = $\dfrac{\text{renda}}{\text{população}}$

3) Produção *per capita* = $\dfrac{\text{valor total da produção}}{\text{população}}$

4.3) Os coeficientes

Os coeficientes são razões entre o número de ocorrências e o número total.

Exemplos:

1. Coeficiente de Natalidade = $\dfrac{\text{nº de nascimentos}}{\text{população total}}$

2. Coeficiente de mortalidade = $\dfrac{\text{nº de óbitos}}{\text{população total}}$

3. Coeficiente de evasão escolar = $\dfrac{\text{nº de alunos evadidos}}{\text{nº inicial de matrículas}}$

4.4) As taxas

As taxas são os coeficientes multiplicados por uma potência de 10 (10, 100, 1000, ...) para tornar o resultado mais inteligível.

Exemplos:

1. Taxa de mortalidade = coeficiente de mortalidade x 1000
2. Taxa de natalidade = coeficiente de natalidade x 1000
3. Taxa de evasão escolar = coeficiente de evasão escolar x 100

Gráficos Estatísticos

O gráfico estatístico é uma forma de apresentação dos dados estatísticos, cujo objetivo é o de reproduzir, no investigador ou no público em geral, uma impressão mais rápida e viva do fenômeno em estudo, já que os gráficos falam mais rápido à compreensão que as séries.

1. Diagramas

Os diagramas são gráficos geométricos de, no máximo, duas dimensões; para sua construção, em geral, fazemos uso do sistema cartesiano.

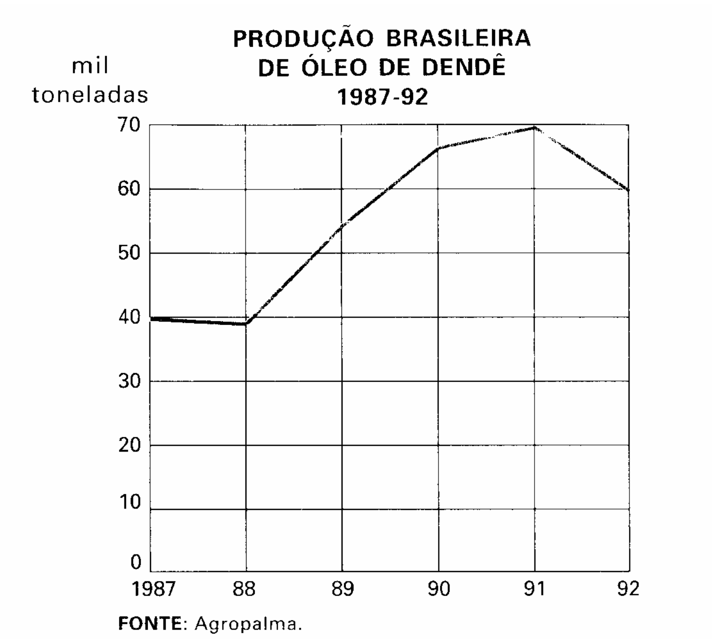
1.1. Gráfico em linha ou em curva

PRODUÇÃO BRASILEIRA DE ÓLEO DE DENDÊ
1987-92

ANOS	QUANTIDADE (1000 t)
1987	39,3
1988	39,1
1989	53,9
1990	65,1
1991	69,1
1992	59,5

FONTE: Agropalma.

Dada a série acima, vamos construir um gráfico de linha ou curva utilizando os anos como abscissas e as quantidades como ordenadas. Assim, um ano dado (x) e a respectiva quantidade (y) formam um par cartesiano (x, y), que pode ser representado num sistema cartesiano. Determinados, graficamente, todos os pontos da série, usando as coordenadas, ligamos todos esses pontos, dois a dois, por segmentos de reta, o que irá nos dar uma poligonal, que é o gráfico de linha ou em curva correspondente a série em estudo.



1.2. Gráfico em colunas ou em barras

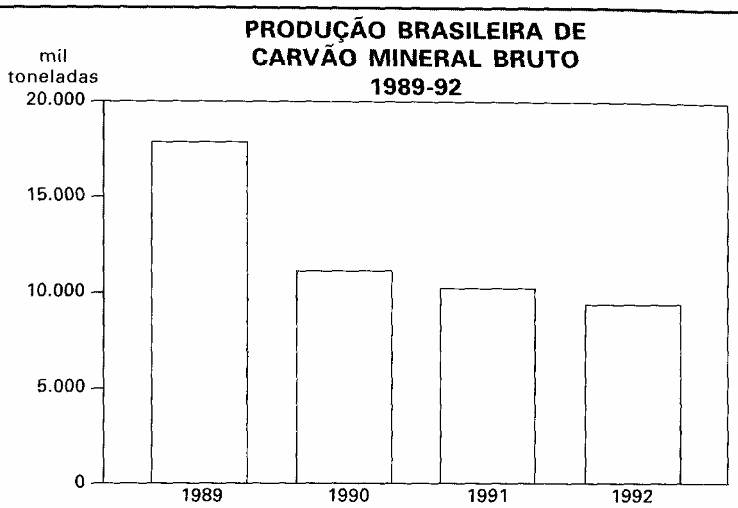
É a representação de uma série por meio de retângulos, dispostos verticalmente (em colunas) ou horizontalmente (em barras).

- **Gráfico em colunas**

PRODUÇÃO BRASILEIRA DE CARVÃO MINERAL BRUTO
1989-92

ANOS	QUANTIDADE PRODUZIDA (1000 t)
1989	18.196
1990	11.168
1991	10.468
1992	9.241

FONTE: Ministério da Agricultura



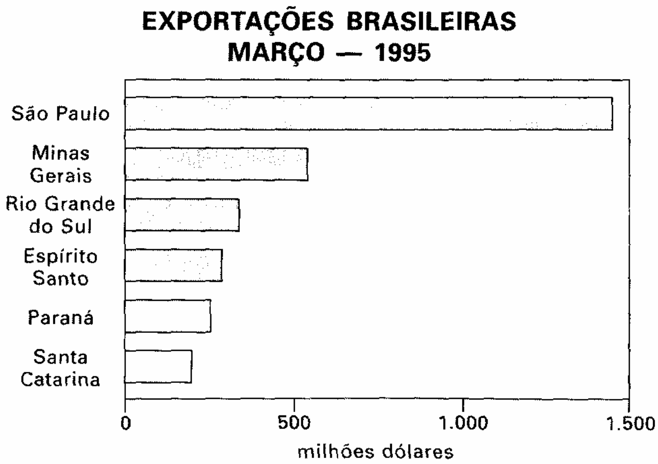
FONTE: Ministério da Agricultura.

- **Gráfico em barras**

**EXPORTAÇÕES BRASILEIRAS
MARÇO — 1995**

ESTADOS	VALOR (US$ milhões)
São Paulo	1.344
Minas Gerais	542
Rio Grande do Sul	332
Espírito Santo	285
Paraná	250
Santa Catarina	202

FONTE: SECEX.



FONTE: SECEX.

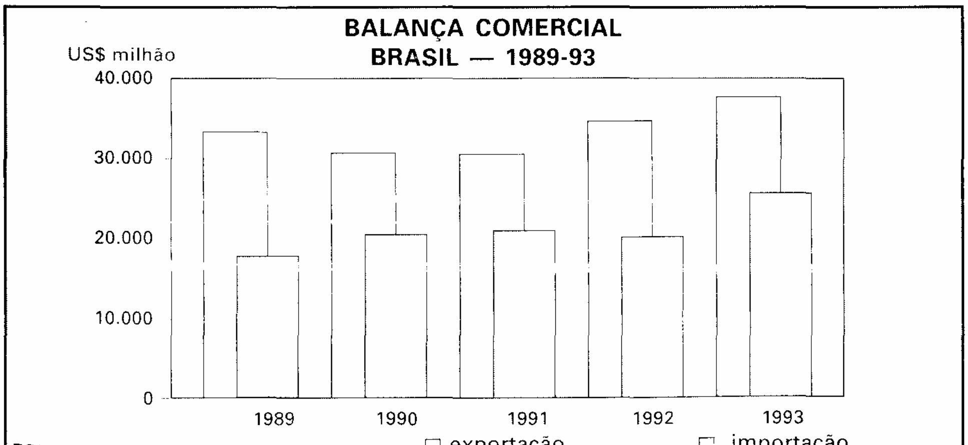
1.3. Gráfico em colunas ou em barras múltiplas

Esse tipo de gráfico é geralmente utilizado quando queremos representar, simultaneamente, dois ou mais fenômenos estudados com o propósito de comparação.

BALANÇA COMERCIAL DO BRASIL
1989-93

ESPECIFICAÇÕES	VALOR (US$ 1.000.000)				
	1989	1990	1991	1992	1993
Exportação (FOB)	34.383	31.414	31.620	35.793	38.783
Importação	18.263	20.661	21.041	20.554	25.711

FONTE: Ministério da Fazenda.



FONTE: Ministério da Fazenda.

1.2. Gráfico em setores

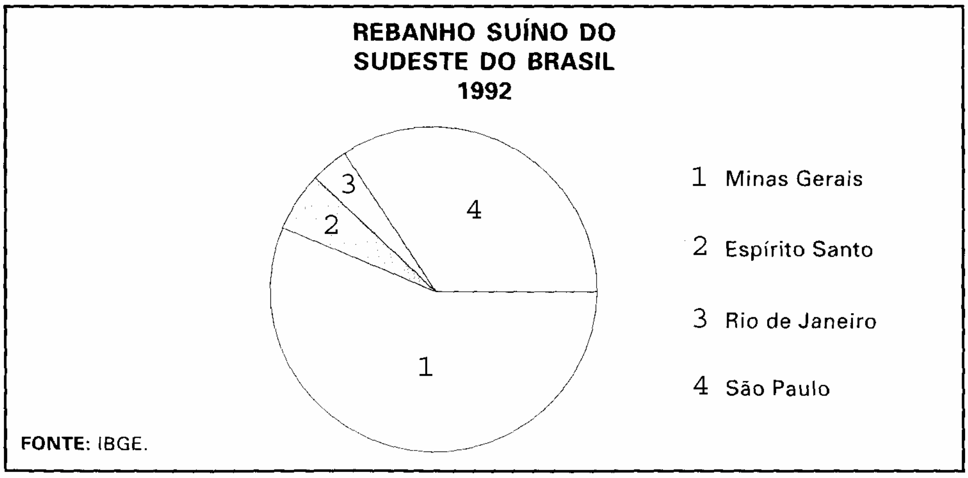
Este gráfico é construído com base em um círculo, e é empregado sempre que desejamos ressaltar a participação do dado no total. O Total é representado pelo círculo, que fica dividido em tantos setores quantas são as partes. Obtemos os setores por meio de regra de três simples e direta, lembrando que o total da série corresponde a 360º.

REBANHO SUÍNO DO SUDESTE DO BRASIL
1992

ESTADOS	QUANTIDADE (mil cabeças)
Minas Gerais	3.363,7
Espírito Santo	430,4
Rio de Janeiro	308,5
São Paulo	2.035,9
Total	6.138,5

FONTE: IBGE.

Logo temos:



2. Cartograma

Representação sobre uma carta geográfica.

Exemplo:

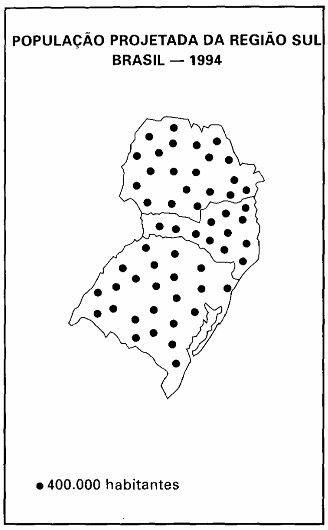
Dada a série:

POPULAÇÃO PROJETADA DA REGIÃO SUL DO BRASIL - 1994

ESTADOS	POPULAÇÃO (hab.)	ÁREA (km^2)	DENSIDADE
Paraná	8.651.100	199.324	43,4
Santa Catarina	4.767.800	95.318	50,0
Rio Grande do Sul	9.475.900	280.674	33,8

FONTE: IBGE.

Obtemos os seguintes diagramas:



3. Pictograma

O pictograma constitui um dos processos gráficos que melhor fala ao público, pela sua forma ao mesmo tempo atraente e sugestiva. A representação gráfica consta de figuras.

Exemplo:

Para a série:

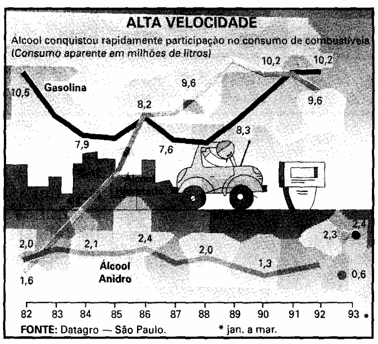
POPULAÇÃO DO BRASIL
1960-90

ANOS	HABITANTES (milhares)
1960	70.070,4
1970	93.139,0
1980	118.562,5
1990	155.822,4

FONTE: IBGE.

Temos a seguinte representação pictórica:

Outro exemplo:



Globo Rural, jul. 1993.

Exercícios – Séries e Gráficos Estatísticos

1. Considere a série estatística:

SÉRIES	ALUNOS MATRICULADOS	%
1ª	546	
2ª	328	
3ª	280	
4ª	120	
Total	1.274	

Complete-a, determinando as percentagens com uma casa decimal e fazendo a compensação, se necessário.

2. Verificou-se, em 1993, o seguinte movimento de importação d mercadorias: 14.839.804 t, oriundas da Arábia Saudita, no valor de US$ 1.469.104.000; 10.547.889 t, dos Estados Unidos, no valor de US$ 6.034.946.000; e 561.024 t, do Japão, no valor de US$ 1.518.843.000. Confeccione a série correspondente e classifique-a, sabendo que os dados acima foram fornecidos pelo Ministério da Fazenda.

3. Represente as séries abaixo usando:

a) Gráfico de linhas

COMÉRCIO EXTERIOR
BRASIL – 1984-93

ANOS	QUANTIDADE (1.000 t)	
	EXPORTAÇÃO	IMPORTAÇÃO
1984	141.737	53.988
1985	146.351	48.870
1986	133.832	60.597
1987	142.378	61.975
1988	169.666	58.085
1989	177.033	57.293
1990	168.095	57.184
1991	165.974	63.278
1992	167.295	68.059
1993	182.561	77.813

FONTE: Min. Indústria, Comércio e Turismo.

b) Gráfico de colunas

PRODUÇÃO BRASILEIRA DE PETRÓLEO BRUTO
1991-93

ANOS	QUANTIDADE (1.000 m^3)
1991	36.180,4
1992	36.410,5
1993	37.164,3

FONTE: Petrobrás.

c) Gráfico em barras

PRODUÇÃO DE OVOS DE GALINHA
BRASIL – 1992

REGIÕES	QUANTIDADE (1.000 dúzias)
Norte	57.297
Nordeste	414.804
Sudeste	984.659
Sul	615.978
Centro-Oeste	126.345

FONTE: IBGE.

d) Gráfico em setores

ÁREA TERRESTRE
BRASIL

REGIÕES	RELATIVA (%)
Norte	45,25
Nordeste	18,28
Sudeste	10,85
Sul	6,76
Centro-Oeste	18,86
Total	100,00

FONTE: IBGE.

Distribuição de Freqüência

1) Introdução

Uma das vantagens das tabelas estatísticas é a de condensar, de forma consistente, as informações necessárias ao estudo desejado. No caso específico das seriações, acontece normalmente que, ao coletar os dados referentes ao fenômeno objeto de estudo, o analista se defronta com valores que se repetem algumas vezes, sugerindo sua representação através de tabelas onde somente apareçam valores distintos uns dos outros. Essa providência favorece evidentemente uma análise e interpretação mais rápida da natureza e comportamento do fenômeno observado.

2) Dados Brutos

Feita a coleta, os dados originais ainda não se encontram prontos para análise, por não estarem numericamente organizados. Por essa razão, costuma-se chamá-los de **dados brutos**. Tomando-se, por exemplo, as alturas dos funcionários de uma empresa e anotando-se os resultados em uma lista da qual constem os nomes dos funcionários em ordem alfabética, ninguém garantirá que os valores correspondentes às alturas observarão uma determinada ordem numérica, crescente ou decrescente. Provavelmente os valores estarão desorganizados, uma vez que a ordem das alturas não corresponde necessariamente à ordem alfabética. Esta lista (tabela) de alturas é, portanto, uma lista de dados brutos. São aqueles valores a que se chegou apenas pela simples coleta, sem preocupação alguma quanto à ordenação.

A tabela 1.1 mostra um exemplo de dados coletados numa empresa qualquer. A tabela refere-se às estaturas de quarenta funcionários desta empresa. Os dados estão representados na forma bruta.

Tabela 1.1

Estaturas de 40 funcionários da empresa X									
166	160	161	150	162	160	165	167	164	160
162	161	168	163	156	173	160	155	164	168
155	152	163	160	155	155	169	151	170	164
154	161	156	172	153	157	156	158	158	161

A essa tabela, cujos elementos não foram ainda numericamente organizados, denominamos **tabela primitiva.**

Quando os vares estão dispostos de forma desorganizada, consegue-se obter pouca informação deles. Mesmo se quisermos observar qual a maior altura, ou a menor, teremos que observar todos os valores um a um.

Assim, conhecidos os valores de uma variável, é difícil formarmos uma idéia exata do comportamento do grupo como um todo, a partir dos dados não-ordenados.

3) Rol

A maneira mais simples de organizar os dados é através de uma certa ordenação (crescente ou decrescente). A tabela obtida após a ordenação dos dados recebe o nome de **rol.**

Essa classificação dos dados proporciona algumas vantagens concretas com relação à sua forma original. Ela possibilita visualizar, de forma bem ampla, as variações de alturas, uma vez que os extremos são percebidos de imediato. È possível observar, também se há ou não uma tendência de concentração de valores numa determinada faixa.

Apesar do rol propiciar ao analista mais informações e com menos esforço de concentração do que os dados brutos, ainda assim persiste o problema de a análise ter que se basear no total de observações individuais. O problema se agravará quando o número de dados for muito grande.

Ordenando os valores da tabela 1.1 formaremos uma nova tabela, 1.2, que recebe o nome de rol.

Tabela 1.2

Estaturas de 40 funcionários da empresa X

150	154	155	157	160	161	162	164	166	169
151	155	156	158	160	161	162	164	167	170
152	155	156	158	160	161	163	164	168	172
153	155	156	160	160	161	163	165	168	173

4) Distribuição de Freqüência

As tabelas de freqüência são representações nas quais os valores se apresentam em correspondência com suas repetições, evitando-se assim que eles apareçam mais de uma vez na tabela, como ocorre com o rol.

No exemplo que trabalhamos, a variável em questão, estatura, será observada e estudada muito mais facilmente quando dispusermos valores ordenados em uma coluna e colocarmos, ao lado de cada valor, o número de vezes que aparece repetido.

Denominamos **freqüência** o número de funcionários que fica relacionado a um determinado valor da variável. Obtemos, assim, uma tabela que recebe o nome de **distribuição de freqüências**. Isto pode ser observado na tabela 1.3:

Tabela 1.3[*]

ESTAT. (cm)	FREQ.
150	1
151	1
152	1
153	1
154	1
155	4
156	3
157	1
158	2
160	5
161	4
162	2
163	2
164	3
165	1
166	1
167	1
168	2
169	1
170	1
172	1
173	1
Total	40

- A tabela foi dividida em três partes para não ocupar muito espaço.

A tabela de freqüências proporciona uma apresentação esteticamente mais vantajosa dos dados, facilitando ainda a verificação do comportamento do fenômeno. É possível, por outro lado, com a utilização de uma tabela de freqüências, a obtenção de estatísticas (medidas) com menos cálculo, e, conseqüentemente, em menos tempo do que se esse trabalho fosse realizado a partir dos dados brutos.

Na tabela 1.3 estamos usando uma **variável discreta** para representar todas as alturas dos alunos pesquisados, pois colocamos em ordem crescente *todos os valores distintos* das alturas. Devemos optar por uma variável discreta na representação de uma série de valores quando o número de elementos distintos da série for pequeno.

Mas o processo dado é ainda inconveniente, já que exige muito espaço, mesmo quando o número de valores da variável (**n**) é de tamanho razoável. Sendo possível, a solução mais aceitável, pela própria natureza da **variável contínua**, é o agrupamento dos valores em vários *intervalos*. Em Estatística, preferimos chamar estes intervalos de intervalos de classes.

Chamando de **freqüência de uma classe** o número de valores da variável pertencente à classe, os dados da tabela 1.3 podem ser dispostos como na tabela 1.4, denominada **distribuição de freqüência com intervalos de classe**:

Tabela 1.4

Estaturas (cm)	**Freqüências**
150 |- 154	4
154 |- 158	9
158 |- 162	11
162 |- 166	8
166 |- 170	5
170 |- 174	3
Total	**40**

Quando os dados estão organizados em uma distribuição de freqüência, comumente denominamos de **dados agrupados**.

Ao agruparmos os valores da variável em classes, ganhamos em simplicidade, mas perdemos em pormenores. O que pretendemos com a construção dessa nova tabela é realçar o que há de essencial nos dados e, também, tornar possível o uso de técnicas analíticas para sua total descrição, até porque a Estatística tem por finalidade específica analisar o conjunto de valores, desinteressando-se de casos isolados.

Devemos optar por uma variável contínua na representação de uma série de valores quando o número de elementos distintos da série for grande.

5) Elementos de uma Distribuição de Freqüência

Para construir uma tabela de freqüências, é necessário conhecer alguns termos próprios e de uso corrente, bem como o procedimento técnico mais adequado. Esses termos serão listados a seguir.

5.1) Freqüência simples absoluta

A **freqüência simples absoluta** de uma classe ou de um valor individual é o número de observações correspondentes a essa classe ou a esse valor. A freqüência simples absoluta é simbolizada por f_i. Assim, em nosso exemplo temos:

$$f_1 = 4, \quad f_2 = 9, \quad f_3 = 11, \quad f_4 = 8, \quad f_5 = 5, \quad f_6 = 3$$

A soma de todas as freqüências é representada pelo símbolo de somatório:

$$\sum_{i=1}^{k} f_i \ \text{ e é evidente que } \ \sum_{i=1}^{k} \mathrm{f_i} = n$$

5.2) Classe:

Classe de freqüência, ou, simplesmente, classe, é cada um dos grupos de valores em que se subdivide a amplitude total do conjunto de valores observados da variável.

As classes são representadas simbolicamente por **i**, sendo i = 1, 2, 3, ..., k (onde k é o número total de classes da distribuição).

Assim, em nosso exemplo anterior, o intervalo 154 □ 158 define a segunda classe (i = 2). Como a distribuição é formada por seis classes, podemos afirmar que k = 6.

Note que na tabela 1.4 usamos para representar as classes, intervalos reais semiabertos à direita. Isto significa que o intervalo contém o menor número, mas não contém o maior.

Usamos isto, pois muitas vezes teremos que ajustar o último valor da última classe (ou o último intervalo de classe) para podermos trabalhar com intervalos iguais. Isto facilitará muitos nossos cálculos.

5.3) Limites de Classe

Os limites de classe são os extremos de cada classe. O menor número é o **limite inferior** da classe (l_i) e o maior número, o **limite superior** da classe (L_i).

Tomando a segunda classe novamente como exemplo, temos que:

$$l_2 = 154 \text{ e } L_2 = 158$$

5.4) Amplitude de um intervalo de classe

A **amplitude de um intervalo de classe** ou, simplesmente, **intervalo de classe** é a medida do intervalo que define a classe. Ela é obtida pela diferença entre os limites superior e inferior dessa classe e é indicada por h_i. Assim:

$$h_i = L_i - l_i$$

No mesmo exemplo da tabela 1.4, temos que:

$$h_2 = L_2 - l_2 \qquad \Rightarrow \qquad h_2 = 158 - 154 = 4 \qquad \Rightarrow \qquad h_2 = 4cm$$

Ou seja, a amplitude da segunda classe é igual a 4cm.

Na realidade, as classes não precisam necessariamente ter a mesma amplitude como no exemplo dado anteriormente. Porém, sempre que possível, devemos trabalhar com classes de mesma amplitude. Isto facilita sobremaneira os cálculos posteriores.

5.5) Amplitude total da distribuição

A **amplitude total da distribuição** (AT) é a diferença entre o limite superior da última classe (**limite superior máximo**) e o limite inferior da primeira classe (**limite inferior mínimo**):

$$AT = L(\text{máx.}) - l(\text{mín.})$$

No exemplo que apresentamos temos:

$$AT = 174 - 150 = 24 \Rightarrow AT = 24cm$$

5.6) Amplitude Amostral

Como foi dito anteriormente, pode acontecer casos em que o último limite é alterado para tornar possível que todas as amplitudes dos intervalos de classes sejam iguais. Neste caso a amplitude da amostra será diferente da amplitude total da distribuição.

Amplitude amostral (AA) é a diferença entre o valor máximo e o valor mínimo da amostra:

$$AA = x(max.) - x(mín.)$$

No nosso exemplo, temos:

$$AA = 173 - 150 = 23 \Rightarrow AA = 23cm$$

5.7) Ponto médio de uma classe

O **ponto médio de uma classe** (**x_i**) é o ponto que divide o intervalo de classe em duas partes iguais. Obtemos este valor da seguinte forma:

$$x_i = \frac{l_i + L_i}{2}$$

Assim, o ponto médio da segunda classe, no nosso exemplo, é:

$$x_2 = \frac{l_2 + L_2}{2} \quad \Rightarrow \quad x_2 = \frac{154 + 158}{2} = 156 \quad \Rightarrow \quad x_2 = 156cm$$

O ponto médio ou valor médio da classe é o valor que a representa, para efeito de cálculo de certas medidas.

Para obter os pontos médios das classes seguintes, ou das anteriores, basta acrescentar, ou subtrair, ao ponto médio da classe precedente a amplitude do intervalo de classe (se este for constante).

Podemos escrever a tabela 1.4 de outra forma, como é mostrado na tabela 1.5:

Tabela 1.5

i	Estaturas (cm)	f_i
1	150 |- 154	4
2	154 |- 158	9
3	158 |- 162	11
4	162 |- 166	8
5	166 |- 170	5
6	170 |- 174	3
	Total	$\sum f_i = 40$

5.8) Número de Classes

Intervalo de Classe

A primeira preocupação que temos, na construção de uma distribuição de freqüência, é a determinação do número de classes e, conseqüentemente, da amplitude e dos limites dos intervalos de classe. O número de classes a ser utilizado depende muito da experiência do pesquisador e das questões que ele pretende responder com a variável contínua.

Para a determinação do número de classes de uma distribuição podemos lançar mão da regra de Sturges, que nos dá o número de classes em função do número de valores da variável:

$$\boxed{i = 1 + 3{,}3 \times \log n}$$

onde: **i** é o número de classes

n é o número total de dados

Podemos usar também o critério da raiz:

$$i = \sqrt{n} \text{ onde n é o número total de elementos}$$

Como o número i de classes deve necessariamente um número inteiro e como dificilmente $\sqrt{n}$, é um número inteiro, deixa-se como opção para o valor i o valor inteiro mais próximo de $\sqrt{n}$, uma unidade a menos ou uma unidade a mais que este valor.

Exemplo:

n = 40 (total de funcionários do exemplo)

i é o número de classes procurado

Utilizando os dados do nosso exemplo para o caso do critério da raiz, temos que:

$i = \sqrt{n}$ ou

$i = \sqrt{40} \quad \Rightarrow i = 6{,}32$ `em termos de arredondamento temos que i = 6`

Decidido o número de classes que deve ter a distribuição, resta-nos resolver o problema da determinação da amplitude do intervalo da classe, e com isso, a escolha dos intervalos, os quais deverão ser tais que forneçam, na medida do possível, para pontos médios, números que facilitem os cálculos. Isto (amplitude) é possível obtermos dividindo-se a amplitude total amostral pelo número de classes:

$h = AA/i$ onde AA é a amplitude amostral

No nosso exemplo, temos:

$$AA = 173 - 150 = 23 \Rightarrow AA = 23\text{cm}$$

Quando o resultado não for exato devemos arredondá-lo para mais.

Utilizando o valor de *i* encontrado pelo método da raiz, basta encontrar o valor de h (a amplitude):

$$h = 23/6 \quad \Rightarrow \quad h = 3{,}8 = 4$$

Isto é, nosso exemplo terá seis classes cada uma com intervalos iguais a 4.

5.9) Tipos de Freqüência

Freqüências simples ou absolutas (f_i) de uma classe são o número de elementos da seqüência que são maiores ou iguais ao limite inferior desta classe e menores que o limite superior desta classe. A soma das freqüências simples é igual ao número total de dados.

Exemplo:

i	Estaturas (cm)	f_i
1	150 |- 154	4
2	154 |- 158	9
3	**158 |- 162**	**11**
4	162 |- 166	8
5	166 |- 170	5
6	170 |- 174	3
	Total	$\sum f_i = 40$

A freqüência simples da terceira classe é igual 11.

Freqüências relativas (fr_i) são os valores das razões entre as freqüências simples e a freqüência total:

$$fr_i = \frac{f_i}{\sum f_i}$$

Logo, a freqüência relativa da terceira classe, em nosso exemplo, é:

$$fr_3 = \frac{f_3}{\sum f_i} \quad \Rightarrow \quad fr_3 = \frac{11}{40} = 0{,}275 \quad \Rightarrow \quad fr_3 = 0{,}275 \quad \text{ou} \quad \mathrm{fr}_3 = 27{,}5\%$$

Note que este valor representa a participação percentual do elemento na série. Assim é possível fazer a interpretação: 27, 5% dos valores da série estão entre 158cm e 162cm.

É evidente que a soma de todas as freqüências relativas é igual a 1 ou 100%. Seu propósito é o de permitir a análise ou facilitar as comparações.

Freqüência acumulada (F_i) é o total das freqüências de todos os valores inferiores ao limite superior do intervalo de uma classe:

$$F_k = f_1 + f_2 + \ldots + f_k$$

Assim, se quisermos calcular a freqüência acumulada correspondente à terceira classe, teremos:

$$F_3 = \sum_{i=1}^{3} f_i = f_1 + f_2 + f_3 \quad \Rightarrow \quad F_3 = 4 + 9 + 11 \quad \Rightarrow F_3 = 24$$

O que significa existirem 24 alunos com estatura inferior a 162cm.

Freqüência acumulada relativa (Fr_i) de uma classe é a freqüência acumulada da classe, dividida pela freqüência total da distribuição.

Baseados na tabela de dados apresentada anteriormente (1.4), podemos montar uma nova tabela com as freqüências estudadas:

i	**Estaturas (cm)**	f_i	x_i	fr_i	F_i	Fr_i
1	150 l- 154	4	152	0,100	4	0,100
2	154 l- 158	9	156	0,225	13	0,325
3	158 l- 162	11	160	0,275	24	0,600
4	162 l- 166	8	164	0,200	32	0,800
5	166 l- 170	5	168	0,125	37	0,925
6	170 l- 174	3	172	0,075	40	0,100
		$\sum = 40$		$\sum = 1{,}000$		

Existem, ainda outros tipos de freqüência. Como é o caso da freqüência acumulada "acima de". A freqüência absoluta acumulada "acima de" uma classe ou de um valor individual representa o número de observações existentes além do valor ou da classe, incluindo no cálculo as observações correspondentes a esse valor ou a essa classe.

A freqüência relativa acumulada "acima de" classe ou valor individual é igual a soma da freqüência simples relativa dessa classe ou desse valor com as freqüências simples relativas das classes ou dos valores posteriores.

5.10) Distribuição de Freqüência sem Intervalo de Classe

Quando se trata de variável discreta de variação relativamente pequena, cada valor pode ser tomado como um intervalo de classe (intervalo degenerado) e, nesse caso, a distribuição é chamada **Distribuição sem intervalos de classe**.

Exercícios

1) As notas obtidas por 50 alunos de uma classe foram:

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

a) Complete a distribuição de freqüência abaixo:

i	Notas	x_i	f_i
1	0 |- 2	1	1
2	2 |- 4		
3	4 |- 6		
4	6 |- 8		
5	8 |- 10		
			$\sum f_i = 50$

b) Agora, responda:

1) Qual a amplitude amostral?
2) Qual a amplitude da distribuição?
3) Qual o número de classes da distribuição?
4) Qual o limite inferior da quarta classes?
5) Qual o limite superior da classe de ordem 2?
6) Qual a amplitude do segundo intervalo de classe?

c) Complete:

1) $h_3 =$ 3) $l_1 =$ 5) $x_2 =$

2) n = 4) $L_3 =$ 6) $f_5 =$

2. Construa a distribuição de freqüências para a série abaixo que representa uma amostra dos salários de 25 funcionários selecionados em uma empresa.

Classe	**Salários R$**	**Número de funcionários** f_i
1	1.000,00 |- 1.200,00	2
2	1.200,00 |- 1.400,00	6
3	1.400,00 |- 1.600,00	10
4	1.600,00 |- 1.800,00	5
5	1.800,00 |- 2.000,00	2

3. Complete o quadro de distribuição de freqüências.

Classe	**Int. Classe**	f_i	f_{ri} %	F_i	F_{ri} %
1	6 |- 10	1			
2	10 |- 14		25		
3	14 |- 18			14	
4	18 |- 22				90
5	22 |- 26	2			

4. Construa a distribuição de freqüências para a série abaixo que representa o número de acidentes em determinado cruzamento observados por dia, durante 40 dias.

Número de acidentes por dia - x_i	**Número de dias -** f_i
0	30
1	5
2	3
3	1
4	1

Estatística Aplicada
Livro: Introdução Ilustrada à Estatística – Sérgio F. Costa – pág. 77 a 88.

Medidas de Variabilidade

Conjunto de medidas (estatísticas) que medem as oscilações de uma variável.
Ex: variância, desvio-padrão.
Ex:

Conjunto A: 8,9,10,8,6,11,7,13.
Conjunto B: 7,3,10,6,5,13,18,10.

A: Total de acertos:72 B: Total de acertos: 72
Total de atiradores:8 Total de atiradores:8

Qual dos grupos de atiradores é mais estável? Em qual dos grupos a variação entre o desempenho é menor?

$\bar{x}$ A = 9 $\bar{x}$ B = 9 acertos

Concluir que os conjuntos são igualmente estáveis? Não dá.

```
Conjunto A

Acertos variam
de 6 a 13
```

```
Conjunto B

Acertos
variam de 3
a 18
```

Variação total: 13 - 6 = 7 18 – 3 = 15 acertos

Variância:
- Subtrair de cada valor a média
- Elevar cada diferença ao quadrado
- Somar os quadrados
- Dividir a soma dos quadrados pelo número de parcelas

A				B		
xi	xi-$\bar{x}$	(xi-$\bar{x}$)2		yi	yi-$\bar{y}$	(yi-$\bar{y}$)2
8	8-9 = -1	1		7	7-9 = -2	4
9	9-9 = 0	0		3	3-9 = -6	36
10	10-9 = 1	1		10	10-9 = 1	1
8	8-9 = -1	1		6	6-9 = -3	9
6	6-9 = -3	9		5	5-9 = -4	16
11	11-9 = 2	4		13	13-9 = 4	16
7	7-9 = -2	4		18	18-9 = 9	81
13	13-9 = 4	16		10	10-9 = 1	1
72	0	36		72	0	164

36/8 = 4,5 acertos2 164/8 = 20,5 acertos2
variância = s2(x) variância = s2(y)

s(x) = $\sqrt{4,5}$ = 2,1 acertos   s(y) = $\sqrt{20,5}$ = 4,5 acertos

Os acertos distam (variam) da média de 2,1 acertos (em média)

Os acertos variam em torno da média de 4,5 acertos

- Quanto maior a variância, maior a heterogeneidade entre os grupos.
- Quanto maior a variância, maior o desvio-padrão.

  s(x) = 2,1 < s(y) = 4,5

  O grupo A é mais estável, mais homogêneo que o grupo B.

Variância e desvio-padrão – dados agrupados

Ex 1:

x (cm)	fi	xi	xi- $\bar{x}$	(xi- $\bar{x}$)2	(xi- $\bar{x}$)2. fi
140 |- 145	3	142.5	-17.375	301.89	905.67
145 |- 150	5	147.5	-12.375	153.14	765.7
150 |- 155	2	152.5	-7.375	54.39	108.78
155 |- 160	7	157.5	-2.375	5.64	39.48
160 |- 165	14	162.5	2.625	6.89	96.47
165 |- 170	6	167.5	7.625	58.14	348.84
170 |- 175	0	172.5	12.625	159.39	0
175 |- 180	1	177.5	17.625	310.64	310.64
180 |- 185	2	182.5	22.625	511.89	1023.78
	40				3599.36

Média = $\frac{\sum x_i f_i}{\sum f_i} = \frac{6395}{40} = 159.875$

S2(x) = $\frac{3599.36}{40} = 89.984$

s(x) = 9.49 cm

Ex 2:

Considere os seguintes conjuntos de dados:

x:70,70,70,70,70 $\bar{x}$ = 70

y: 68,69,70,71,72 $\bar{x}$ = 70

z: 5,15,50,120,160 $\bar{x}$ = 70

x é mais homogêneo que y e z e y é mais homogêneo que z.

DP x:

xi	xi- $\bar{x}$	(xi- $\bar{x}$)2
70	0	0
70	0	0
70	0	0
70	0	0
70	0	0

Variância = 0/5 = 0
DP = 0
Não existe variabilidade

DP y

yi	yi- $\bar{y}$	(yi- $\bar{y}$)2
68	-2	4
69	-1	1
70	0	0
71	1	1
72	2	4
		10

Variância = 10/5 = 2
DP = $\sqrt{2}$ = 1,4142

DP z

zi	zi- $\bar{z}$	(zi- $\bar{z}$)2
5	-65	4225
15	-55	3025
50	-20	400
120	50	2500
160	90	8100
		18250

Variância = 18250/5 = 3650
DP = 60,42

Ex 3:

Estaturas (cm)	fi	xi	xi.fi	xi-$\bar{x}$	(xi-$\bar{x}$)2	(xi-$\bar{x}$)2. fi
150 |- 154	4	152	608	-9	81	324
154 |- 158	9	156	1404	-5	25	225
158 |- 162	11	160	1760	-1	1	11
162 |- 166	8	164	1312	3	9	72
166 |- 170	5	168	840	7	49	245
170 |- 174	3	172	516	11	121	363
	40		6440			1240

Média = $\frac{\sum x_i f_i}{\sum f_i} = \frac{6440}{40} = 161$

s2(x) = $\frac{1240}{40} = 31$

s(x) = 5.57 cm

Lista de Exercícios

1. As taxas de juros recebidas por 10 ações durante um certo período foram (%) 2,59; 2,64; 2,62; 2,57; 2,55; 2,61; 2,50; 2,63; 2,64. Calcule a média e o desvio-padrão.
2. Para facilitar um projeto de ampliação da rede de esgoto de um região de uma cidade, as autoridades tomaram uma amostra de tamanho 50 dos 270 quarteirões que compõem a região. Foram encontrados os seguintes números de casas por quarteirão:

2	2	3	10	12	14	15	15	16	16
18	18	20	21	22	22	23	24	25	25
26	27	29	29	30	32	36	42	44	45
45	46	48	52	58	59	61	61	61	65
66	66	68	75	78	80	89	90	92	97

a) agrupe os dados em uma tabela de freqüência;
b) determine a média e o desvio-padrão;
c) explique o significado dessas medidas.

3. O departamento pessoal de uma certa firma fez um levantamento dos salários dos 120 funcionários do setor administrativo, obtendo os resultados:

Faixa salarial	fi
0 |- 2	30
2 |- 4	48
4 |- 6	24
6 |- 10	18

Calcule a média e o desvio-padrão.

4. Para se estudar o desempenho de duas companhias corretoras de ações, selecionaram-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação computam-se o percentual de lucro durante um período fixo de tempo.

Corretora A				Corretora B		
45	60	54		57	55	58
62	55	70		50	52	59
38	48	64		59	55	56
55	56	55		61	52	53
54	59	48		57	57	50
65	55	60		55	58	54
				59	51	56

Quem tem maior variabilidade? Corretora A ou B?

5. Um grupo de 85 moças tem estatura média de 160,6 cm com desvio-padrão de 5,97 cm. O outro grupo de 125 moças tem estatura média de 161,9 cm com desvio-padrão de 6,01 cm. Qual grupo é mais homogêneo?