

ELEMENTOS DE ESTATÍSTICA

Este texto constitui instrumento valioso para aqueles que precisam interpretar estatísticas e se revela obrigatório para os que precisam fazer tabelas, desenhar gráficos e calcular médias, correlações e proporções.

Diferentemente dos textos de Estatística que pressupõem conhecimentos básicos na matéria, este livro foi escrito para os que estão iniciando o estudo da Estatística. Os conceitos são expostos aqui de forma didática, sempre acompanhados de exemplos claros, extraídos da vida real.

Partindo inicialmente da definição de Estatística, aborda dados estatísticos, população e amostra, técnicas de amostragem e tendenciosidade da amostra. Em seguida trata da organização de dados em tabelas, da apresentação gráfica de dados, tabelas de distribuição de frequências, das medidas de tendência central, das medidas de dispersão, da correlação e regressão linear simples. Finalmente aborda a confiabilidade e números-índices.

Para a verificação da aprendizagem foi apresentado, no final dos capítulos, um conjunto de exercícios com as respostas no final do texto.

BIOGRAFIA SOBRE OS AUTORES

SONIA VIEIRA é professora titular na Unicamp. Tem diversos livros publicados, dentre os quais se destacam *Introdução à bioestatística*, *Metodologia científica aplicada à área de saúde e estatística*, *Estatística: introdução ilustrada*, este último publicado pela Atlas.

RODOLFO HOFFMANN é professor titular na USP. É conhecido por seus trabalhos na área de distribuição de renda. Publicou os seguintes livros: *Administração da empresa agrícola*, *Análise de regressão: uma introdução à econometria*, *Estatística para economistas*.

INDICAÇÃO

Texto básico para disciplinas introdutórias de Estatística, dos cursos de Administração, Educação, Psicologia, Biblioteconomia e Ciências Sociais. Livro-texto para a disciplina ESTATÍSTICA no ensino de 2º Grau.

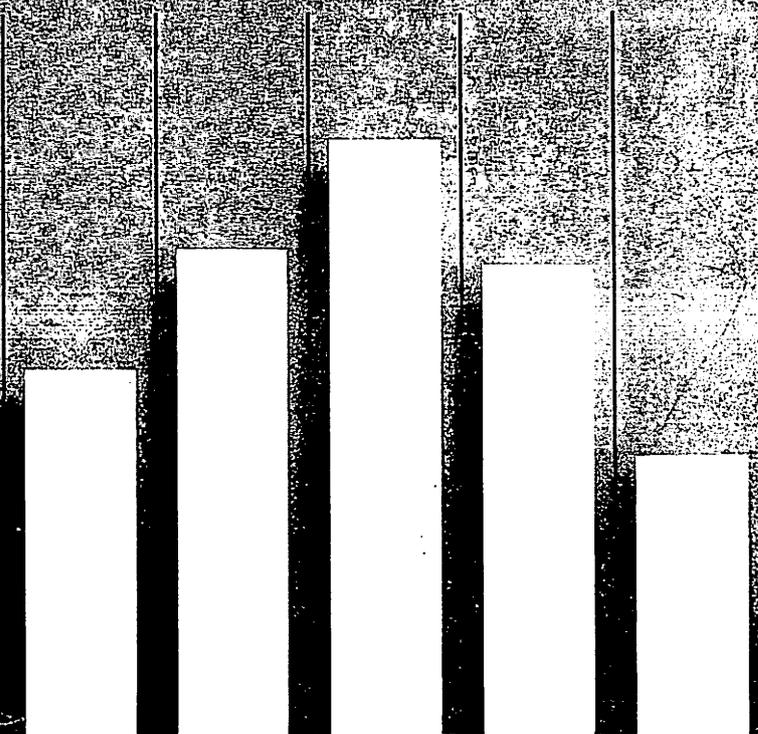
publicação atlas

ISBN 85-224-0

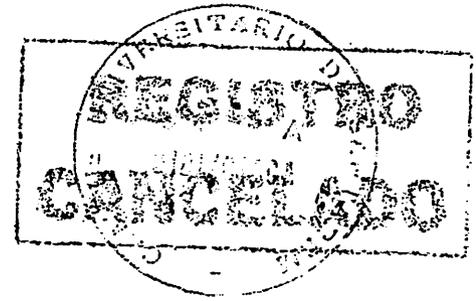
39.7
V67e
2d.
EX04

Sonia Vieira
Rodolfo Hoffmann

ELEMENTOS DE ESTATÍSTICA

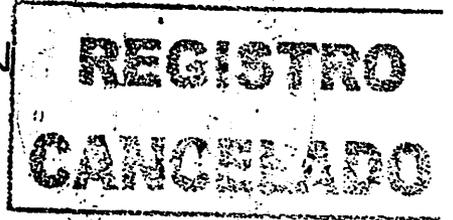


atlas



**ELEMENTOS
DE
ESTADÍSTICA**

SONIA VIEIRA
RODOLFO HOFFMANN



ELEMENTOS DE ESTATÍSTICA

2.^a Edição



EDITORIA ATLAS S.A.

Rua Conselheiro Nébias, 1384 (Campos Elísios)
Caixa Postal 7186 — Tel.: (011) 221-9144 (PABX)
01303 São Paulo (SP)

SÃO PAULO
EDITORA ATLAS S.A. — 1990

1986 by EDITORA ATLAS S.A.
Rua Conselheiro Nébias, 1384 (Campos Elísios)
Caixa Postal 7186 — Tel.: (011) 221-9144 (PABX)
01303 São Paulo (SP)

d. 1986; 2. ed. 1990

N 85-224-0178-0

Impresso no Brasil/Printed in Brazil

Depósito legal na Biblioteca Nacional, conforme Decreto nº 1.825, de 20 de dezembro de 1907.

TODOS OS DIREITOS RESERVADOS — É proibida a reprodução total ou parcial, de qualquer forma ou por qualquer meio, salvo com autorização, por escrito, do Editor.

Ilustrado por
Sílvia Ferreira Leite

UFOPA - BIBLIOTECA	
CAMPUS	
REGISTRO	4935/30 EX. 010
CLASSE	
CUTTER	

Dados de Catalogação na Publicação (CIP) Internacional
(Câmara Brasileira do Livro, SP, Brasil)

Vieira, Sônia.
Elementos de estatística / Sonia Vieira, Rodolfo Hoffmann.
— 2. ed. — São Paulo : Atlas, 1990

Bibliografia.
ISBN 85-224-0178-0

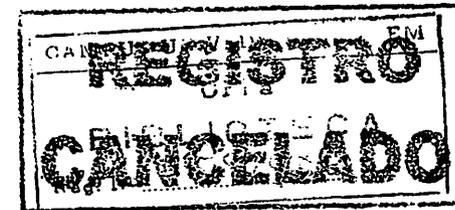
1. Estatística 2. Estatística matemática I. Hoffmann, Rodolfo,
1942 — II. Título.

89-2063

CDD-001.422
-519.5

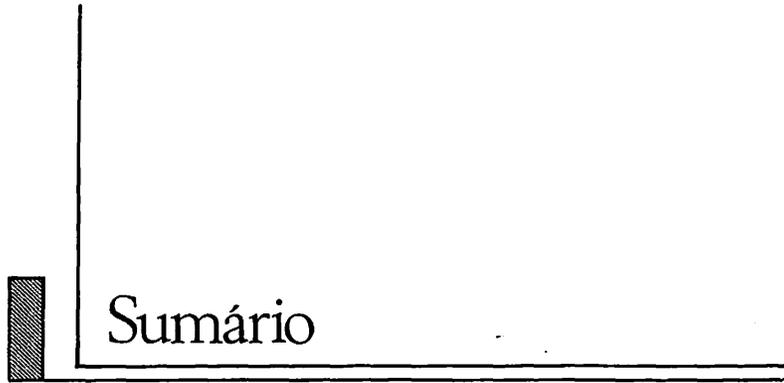
Índices para catálogo sistemático:

1. Estatística : Método 001.422
2. Estatística matemática 519.5
3. Método estatístico 001.422



Os autores agradecem a:

- Ronaldo Wada, pela ilustração do texto.
- Mauro Eduardo Del Grossi, Ivany A. Lombardo, Sueli D. de O. Soliani e Ivany do C. G. Gerola, pelas críticas e sugestões.
- Sonia, Baby e Elisa, pelos serviços datilográficos.



Sumário

1 INTRODUÇÃO, 13

- 1.1 Dados estatísticos, 14
- 1.2 Apuração de dados, 14
- 1.3 População e amostra, 16
- 1.4 Técnicas de amostragem, 17
 - 1.4.1 Amostragem casual simples, 17
 - 1.4.2 Amostragem sistemática, 18
 - 1.4.3 Amostragem estratificada, 19
- 1.5 Tendenciosidade da amostra, 19
- 1.6 Exercícios, 19

2 ORGANIZAÇÃO DE DADOS EM TABELAS, 21

- 2.1 Elementos das tabelas, 22
- 2.2 Normas para a apresentação de tabelas, 25
- 2.3 Séries estatísticas, 28
- 2.4 Exercícios, 30

3 APRESENTAÇÃO GRÁFICA DE DADOS, 33

- 3.1 Gráfico de barras, 35
- 3.2 Gráfico de linhas, 39
- 3.3 Gráfico retangular de composição, 41
- 3.4 Gráfico de setores, 44
- 3.5 Gráfico polar, 45
- 3.6 Gráficos comparativos, 47
- 3.7 Exercícios, 51

4 TABELAS DE DISTRIBUIÇÃO DE FREQUÊNCIAS, 55

- 4.1 Organização da tabela de distribuição de frequências, 55
- 4.2 Organização de dados em classes, 57
- 4.3 Frequência relativa, 61
- 4.4 Frequência acumulada, 62
- 4.5 Frequência relativa acumulada, 63
- 4.6 Apresentação gráfica, 64
 - 4.6.1 Histograma, 64
 - 4.6.2 Polígono de frequências, 67
- 4.7 Exercícios, 69

5 MEDIDAS DE TENDÊNCIA CENTRAL, 73

- 5.1 Média aritmética, 74
 - 5.1.1 Média aritmética de dados apresentados em tabelas de distribuição de frequências, 75
- 5.2 Média ponderada, 78
- 5.3 Média geométrica, 79
- 5.4 Média harmônica, 80
- 5.5 Moda, 81
 - 5.5.1 A moda de dados organizados em classes, 82

5.6 Mediana, 84

- 5.6.1 A mediana de dados apresentados em tabelas de distribuição de frequências, 87

5.7 Quartis e decis, 88

5.8 Observações, 90

5.9 Exercícios, 90

6 MEDIDAS DE DISPERSÃO, 93

6.1 Amplitude, 94

6.2 Variância, 95

- 6.2.1 Variância de dados agrupados, 99

6.3 Desvio-padrão, 101

6.4 Coeficiente de variação, 102

6.5 Exercícios, 104

7 CORRELAÇÃO E REGRESSÃO LINEAR SIMPLES, 107

7.1 Coeficiente de correlação, 108

7.2 Cuidados na interpretação do coeficiente de correlação, 113

7.3 Reta de regressão, 115

7.4 Interpolação e extrapolação, 118

7.5 Exercícios, 121

8 PROBABILIDADE, 123

8.1 Conceito de probabilidade, 123

8.2 Probabilidade condicional, 124

- 8.2.1 Eventos independentes, 125

8.3 Teorema do produto, 126

8.4 Teorema da soma, 128

- 8.5 Freqüência relativa, 129
- 8.6 Algumas taxas demográficas, 131
- 8.7 Exercícios, 132

9 NÚMEROS-ÍNDICES, 135

- 9.1 Índice relativo de preço, 135
- 9.2 Índice de custo de vida, 137
- 9.3 Cálculo de valores reais ou deflacionados, 140
- 9.4 Exercícios, 144

Respostas aos exercícios, 147

Referências Bibliográficas, 159

1

Introdução

A palavra *estatística* está associada à idéia de “coleção de números”. A estatística trata, sem dúvida alguma, da organização e da apresentação de contagens e de medições. Dentro deste conceito — de que estatística é coleção de números — cabe lembrar que no Brasil existem contagens e medições sobre assuntos tão diversos como índices de audiência em programas de televisão e valores de taxa de câmbio do dólar no mercado paralelo.

Mas a estatística não deve ser vista como simples “coleção de números”. As informações numéricas são obtidas com a finalidade de acumular informação para a tomada de decisão. Assim, por exemplo, é com base em índices de audiência que uma emissora de televisão pode tirar um programa do ar ou modificar o final previsto de uma novela; é com base em pesquisas de opinião que um candidato a cargo eletivo pode modificar certas atitudes ou mesmo o modo de se vestir; é com base na quantidade vendida de um produto em determinado período que um supermercado pode aumentar ou diminuir o estoque desse produto, e assim por diante.

Muitas decisões são tomadas com base em informações numéricas. No entanto — é preciso considerar —, os números não “falam por si mesmos”. Precisam ser organizados, discutidos e interpretados. Então, a estatística pode ser vista como um conjunto de técnicas que ajudam na tomada de decisão, quando prevalecem condições de incerteza. Por esta razão, todo profissional — seja ele administrador, executivo, cientista ou professor — deve adquirir algum conhecimento de estatística para tomar uma decisão consciente quando tiver, à sua disposição, informações numéricas, ou seja, dados estatísticos. Mas o que são os dados estatísticos?

1.1 DADOS ESTATÍSTICOS

Toda informação devidamente coletada e registrada, quer seja na forma de contagem ou de medição, é um dado estatístico. Todo dado se refere a uma variável. Não existe interesse em coletar mais de um valor quando se trata de uma constante. Assim, não existe interesse em coletar informações sobre analfabetismo entre secundaristas, porque todos os estudantes do secundário são alfabetizados.

Existem três tipos de variável:

- nominal ou qualitativa;
- ordinal;
- cardinal ou quantitativa.

Tem-se uma *variável nominal* ou *variável qualitativa* quando os indivíduos ou elementos são classificados em determinado número de categorias mutuamente exclusivas. Por exemplo, alfabetização é uma variável nominal porque determinada pessoa só pode pertencer a uma de duas categorias mutuamente exclusivas: ou ela é alfabetizada ou é analfabeta. Outro exemplo de variável nominal é a classificação dos brasileiros de acordo com o Estado onde nasceram: um indivíduo é ou baiano, ou paulista, ou cearense etc.

Denomina-se *variável ordinal* aquela que pode ser distribuída em determinado número de categorias mutuamente exclusivas, mas tais categorias se apresentam segundo uma ordem lógica. Então, grau de instrução é uma variável ordinal porque uma pessoa pode ser classificada em uma das seguintes categorias: 1. sem instrução; 2. com primário incompleto; 3. com primário completo; 4. com secundário incompleto; 5. com secundário completo; 6. com superior incompleto; 7. com superior completo.

As categorias de uma variável ordinal podem ser indicadas por números, de acordo com a ordem. É preciso deixar claro, no entanto, que neste caso apenas a ordem do número tem interesse. Não podem ser feitas operações aritméticas com estes números. Assim, no exemplo dado, duas vezes primário completo não é igual a curso superior incompleto.

Denomina-se *variável cardinal* ou *variável quantitativa* aquela que é expressa através de um número, com todas as suas propriedades. Por exemplo, a renda de uma pessoa é uma variável cardinal porque é expressa numericamente, em cruzados e, nesse caso, não há dúvida de que uma renda de 12 mil é três vezes maior do que uma renda de 4 mil.

1.2 APURAÇÃO DE DADOS

Os dados são registrados em fichas. Essas fichas normalmente contêm, além dos dados de interesse, diversas outras informações, como nome

e endereço de quem forneceu os dados, nome do entrevistador etc. Então, terminada a fase de coleta de dados, é preciso proceder à *apuração*, isto é, reunir os dados.

Quando os dados se referem a uma variável nominal, ou a uma variável ordinal, a apuração se reduz a simples contagem, isto é, a apuração consiste em estabelecer o número de elementos que caem em cada categoria. Para entender o procedimento, imagine que um professor quer saber a proporção de alunos de cada sexo que se candidataram a determinado exame. O professor deve então tomar todas as fichas de inscrição e escrever, numa folha de papel:

Masculino

Feminino

Depois o professor deve examinar cada ficha de inscrição e fazer um traço, na linha correspondente a um dos sexos, toda vez que a ficha de inscrição indicar que o aluno é desse sexo. No exemplo dado em seguida, cada traço representa um aluno e cada quadrado, cortado pela diagonal, representa cinco alunos.

Masculino ☑☑☑☑☑ = 22

Feminino ☑☑☑☑ = 18

Também pode ser adotada outra forma de notação. No exemplo dado em seguida, cada traço representa um aluno e cada conjunto de traços, cortados em horizontal, representa cinco alunos.

Masculino ||| ||| ||| ||| || = 22

Feminino ||| ||| ||| || = 18

Se os dados se referem a uma variável cardinal, a apuração exige que se anote cada valor observado. Para entender o procedimento, imagine que um professor de educação física quer apurar os dados de estatura dos alunos de uma classe. Para isso o professor precisa tomar todas as fichas biométricas e anotar, em uma folha de papel, o nome (ou o número) de cada aluno e a respectiva estatura, como mostra o exemplo dado em seguida.

Nome	Estatura
Antônio	1,68 m
Bráulio	1,72 m
.....
.....
Tadeu	1,59 m

1.3 POPULAÇÃO E AMOSTRA

Na linguagem comum, população significa o conjunto de habitantes de um território, de um país, de uma região, de uma cidade etc. Em Estatística, no entanto, o termo *população* tem significado bem mais geral. Entende-se por *população* o conjunto de elementos que têm, em comum, determinada característica.

As populações podem ser finitas como, por exemplo, os alunos matriculados em determinada escola, ou infinitas como, por exemplo, os resultados obtidos quando se joga um dado sucessivamente. Existem populações que, embora finitas, são consideradas infinitas para qualquer finalidade prática. Como exemplo, imagine o número de peixes que há no mar. Esse número é, em dado instante, matematicamente finito, mas tão grande que pode ser considerado infinito para qualquer finalidade prática.

Entende-se por *amostra* qualquer conjunto de elementos, retirado da população, desde que esse conjunto seja não-vazio e tenha menor número de elementos do que a população. Imagine uma população constituída por quatro elementos, isto é, por:

Ana, Pedro, Luís, Clara.

Dessa população podem ser retiradas:

a) quatro amostras de três elementos cada uma:

Ana; Pedro; Luís

Ana; Luís; Clara

Ana; Pedro; Clara

Pedro; Luís; Clara

b) seis amostras de dois elementos cada uma:

Ana; Pedro

Ana; Clara

Pedro; Clara

Ana; Luís

Pedro; Luís

Luís; Clara

c) quatro amostras de um elemento cada uma:

Ana

Luís

Pedro

Clara

Quando são coletados dados sobre todos os elementos da população, diz-se que foi feito um recenseamento. No Brasil, os recenseamentos gerais são feitos por um órgão oficial, a Fundação Instituto Brasileiro de Geografia e Estatística — IBGE —, que tem sede na cidade do Rio de Janeiro. Denomina-se *censo* o conjunto de dados obtidos através de recenseamento. Existem censos demográficos (que descrevem as caracterís-

ticas da população), censos agropecuários, censos industriais, censos comerciais e de serviços etc. No Brasil foram feitos nove recenseamentos gerais, isto é, em 1872, 1890, 1900, 1920, 1940, 1950, 1960, 1970 e 1980.

O Recenseamento Geral de 1980, a exemplo dos anteriores, atendeu às determinações da ONU (Organização das Nações Unidas) para se integrar ao Censo das Américas. Essas determinações visam garantir a uniformidade de conceitos utilizados nos censos das nações americanas.

É preciso deixar claro, porém, que nem sempre é possível coletar dados de toda a população. Um censo pode revelar-se impossível por razões de economia. Assim, por exemplo, para determinar o índice de audiência de um programa de rádio, que atinge todo o Brasil, seria absurdo realizar um censo, devido ao custo muito elevado. Então, os índices de audiência de programas de rádio e de televisão são calculados com base em amostras.

Também não se usam censos nos casos de estudos demorados, que exigem atenção cuidadosa do pesquisador. Considere, como exemplo, um estudo que tenha a finalidade de obter uma descrição pormenorizada de como uma família gasta sua renda, isto é, um estudo sobre orçamento familiar. Nesse tipo de estudo é preciso coletar grande número de dados de cada família. É impraticável coletar dados de todas as famílias brasileiras ou, até mesmo, de todas as famílias de uma cidade. Então, os estudos de orçamento familiar são, necessariamente, feitos com base em amostras.

Estes exemplos servem para lembrar que boa parte do conhecimento se baseia no estudo de amostras. No entanto, a validade de uma amostra depende da técnica utilizada para sua obtenção.

1.4 TÉCNICAS DE AMOSTRAGEM

Uma amostra só é representativa da população quando é constituída por elementos selecionados de acordo com uma técnica conhecida. Existem diversas técnicas de amostragem, mas as mais conhecidas são:

- amostragem casual simples;
- amostragem sistemática;
- amostragem estratificada.

1.4.1 Amostragem casual simples

No processo de *amostragem casual simples*, todos os elementos da população têm igual probabilidade de serem selecionados para constituir

a amostra. Como exemplo, imagine que o gerente de uma indústria pretenda obter uma amostra dos empregados. Para isso, o gerente pode conferir um número a cada empregado e sortear aqueles que irão constituir a amostra.

Para fazer o sorteio, o gerente pode usar uma urna que contenha fichas numeradas, de zero a nove, inclusive. Então o gerente retira uma ficha da urna, ao acaso, e anota o número, que dará o primeiro dígito do número do empregado que será selecionado. Depois o gerente recoloca a ficha na urna, mistura e retira outra ficha, que dará o segundo dígito do número do empregado que será selecionado. O gerente deve repetir esse processo até que sejam obtidos todos os dígitos da numeração utilizada.

Se a indústria tiver, por exemplo, 900 empregados, devem ser sorteados três dígitos para definir cada empregado que será selecionado para constituir a amostra. Então, no primeiro sorteio pode ocorrer, por exemplo:

Primeira ficha retirada:	4
Segunda ficha retirada:	7
Terceira ficha retirada:	4

O resultado do primeiro sorteio indica que foi selecionado, para a amostra, o empregado de número 474. O gerente deve repetir esse processo até completar a amostra. Evidentemente, serão desprezados números iguais a zero (0,0,0), números já sorteados (já estão na amostra) e números maiores do que 900 (não existem nessa população).

1.4.2 Amostragem sistemática

No processo de *amostragem sistemática*, os elementos são selecionados para a amostra por um sistema preestabelecido. Assim, por exemplo, imagine que um gerente de vendas quer obter uma amostra dos clientes cadastrados de sua firma. Se as fichas dos clientes estiverem organizadas por ordem alfabética, o gerente obterá uma amostra sistemática se escolher a quinta de cada cinco fichas, ou a décima, de cada dez etc.

O processo de obtenção de uma amostra sistemática é bem mais simples do que o processo de obtenção de uma amostra casual simples nos casos em que a população está organizada. Assim, para obter uma amostra dos assinantes de um jornal é mais fácil localizar os nomes dos assinantes no arquivo e, de cada 20, tirar, por exemplo, o vigésimo do que dar um número a cada assinante e fazer sorteio até completar uma amostra que corresponderia a 1/20 da população.

Nas pesquisas domiciliares também se usa amostragem sistemática. Nesses casos é comum estabelecer um ponto de partida ao acaso em cada bairro e, a partir daí, percorrer as ruas usando determinado sistema como, por exemplo, de cada dez domicílios selecionar, para a amostra, o décimo.

1.4.3 Amostragem estratificada

Usa-se o processo de *amostragem estratificada* quando a população se apresenta dividida em estratos, isto é, quando a população está dividida em grupos distintos. Como exemplo, imagine que o diretor de uma faculdade quer obter uma amostra da comunidade acadêmica. O diretor precisa considerar que a faculdade é constituída por três grupos distintos de pessoas, isto é, professores, funcionários e alunos.

Então, para obter uma amostra mais representativa da comunidade acadêmica, o diretor deve selecionar uma amostra dentro de cada estrato, isto é, uma amostra dos professores, uma amostra dos funcionários e uma amostra dos alunos e depois reunir essas três amostras em uma só, constituindo então uma amostra estratificada.

1.5 TENDENCIOSIDADE DA AMOSTRA

Muitas vezes as amostras são constituídas por pessoas que devem executar algum tipo de tarefa, como responder perguntas, preencher um questionário ou até mesmo testar um produto. Algumas pessoas se recusam a cooperar. Nesses casos é preciso ter muito senso crítico para avaliar se a amostra efetivamente utilizada é representativa da população. Sempre é possível que a amostra obtida seja *tendenciosa* ou *viciada*, isto é, não representativa da população.

O senso crítico ainda é mais importante quando as amostras são constituídas por voluntários. Muitas vezes, o procedimento usado para solicitar voluntários conduz à formação de amostras tendenciosas. Como exemplo, imagine que um professor de educação física peça que três alunos da turma se apresentem como voluntários para apostar uma corrida. Ora, é bastante razoável imaginar que, neste caso, se apresentarão como voluntários apenas os alunos que sabem ser bons corredores. Então, os três alunos não constituirão uma amostra representativa da turma, mas uma amostra tendenciosa ou viciada.

1.6 EXERCÍCIOS

1. Classifique as seguintes variáveis como nominal, ordinal ou cardinal: idade, sexo, renda familiar, religião, cor, tempo disponível para estudo, atividades recreativas de interesse.
2. Dê dois exemplos de variável nominal, dois exemplos de variável ordinal e dois exemplos de variável cardinal.

3. Dada uma população com três elementos, A, B e C, identifique todas as amostras possíveis que podem ser obtidas dessa população.
4. Dada uma população com seis elementos, A, B, C, D, E e F, explique como você faria para obter, dessa população, uma amostra casual simples com três elementos.
5. Descreva uma forma de obter uma amostra sistemática com dez elementos de uma população com 100 elementos.
6. Explique a forma de obter uma amostra estratificada dos empregados de uma firma, considerando que existem empregados de escritório, empregados de oficina e representantes da firma.
7. Imagine que se pretenda fazer um levantamento de opinião pública para verificar se as pessoas são contra ou a favor do uso gratuito de ônibus pelos idosos. Pense em duas maneiras distintas de fazer a pergunta — uma que induza a resposta positiva e outra que induza a resposta negativa.
8. Para verificar que as pessoas tendem a dar algum tipo de resposta, mesmo quando desconhecem o assunto, pergunte para diversas pessoas se elas conhecem uma revista, ou um jornal, ou um ator que na realidade não existe. Pergunte, por exemplo: “Você já ouviu as músicas de Linda Junk?” Discuta a proporção de respostas positivas.
9. Muitas vezes as pessoas concordam ou discordam de uma frase em função não da frase em si, mas do autor da frase. Para verificar essa hipótese pergunte, para diferentes pessoas escolhidas ao acaso: Você concorda com Ruy Barbosa quando ele disse: “É preciso lutar por nossas idéias, até as últimas conseqüências.” Depois pergunte para outro grupo de pessoas: Você concorda com Hitler quando ele disse: “É preciso lutar por nossas idéias, até as últimas conseqüências.” Compare a proporção de respostas positivas.
10. Se uma moeda for jogada 20 vezes, os resultados constituirão uma amostra de tamanho 20 da população infinita de resultados que podem ocorrer quando se joga uma moeda. Obtenha três amostras desse tipo, isto é, jogue uma moeda 20 vezes e depois repita a experiência mais duas vezes. Compare a proporção de caras obtidas nas três amostras.

2

Organização de Dados em Tabelas

Os dados estatísticos são apresentados em tabelas. Como exemplo, observe a tabela 2.1.

Tabela 2.1

População residente no Brasil, segundo o sexo, de acordo com o censo demográfico de 1980

Sexo	Frequência	Percentual
Masculino	59 123 361	49,7
Feminino	59 879 345	50,3
Total	119 002 706	100,0

FONTE: IBGE (1982).

A simples observação da tabela 2.1 mostra que as pessoas que moravam no Brasil em 1980 (população residente) foram divididas em duas categorias: as de sexo masculino e as de sexo feminino. Havia 59 123 361 homens e 59 879 345 mulheres. Portanto, em 1980 a população residente no Brasil era constituída por mais mulheres do que homens. No entanto, a proporção de homens e mulheres era praticamente a mesma. Afinal, 49,7% das pessoas eram do sexo masculino e 50,3% eram do sexo feminino. A tabela 2.1 também mostra que, em 1980, a população brasileira residente era constituída por aproximadamente 120 milhões de pessoas ou — mais exatamente — 119 002 706 pessoas.

Todas essas observações foram feitas através do simples exame dos dados organizados na tabela 2.1. Então este exemplo serviu para mostrar que as tabelas apresentam, de maneira clara e precisa, dados estatísticos. Mas como se faz uma tabela?

2.1 ELEMENTOS DAS TABELAS

As tabelas são constituídas pelos seguintes elementos: título, corpo, cabeçalho e coluna indicadora.

Título: explica o tipo de dado que a tabela contém. Como exemplo, observe o título da tabela 2.1:

População residente no Brasil, segundo o sexo,
de acordo com o censo demográfico de 1980

A simples leitura desse título deixa claro que estão apresentados, na tabela, o número de pessoas de sexo masculino e o número de pessoas de sexo feminino, que constituíam a população residente no Brasil em 1980. O título deve ser colocado no alto da tabela, antes dos dados.

Corpo da Tabela: é formado pelos dados, distribuídos em linhas e colunas. Como exemplo, observe o corpo da tabela 2.1, mostrado em seguida.

59 123 361	49,7
59 879 345	50,3
119 002 706	100,0

Cada cruzamento de uma linha com uma coluna constitui uma casa ou célula. Então, o número 59 123 361, que está no cruzamento da primeira linha com a primeira coluna, é uma casa da tabela.

Cabeçalho: especifica a informação apresentada em cada coluna. Observe o cabeçalho da tabela 2.1, destacado em seguida.

Sexo	Frequência	Percentual
------	------------	------------

Este cabeçalho deixa claro que na primeira coluna estão indicados os dois sexos. Na segunda coluna é dada a frequência, ou seja, o número de pessoas de cada sexo; na terceira coluna está apresentada a porcentagem de pessoas de cada sexo.

Coluna indicadora: identifica o tipo de informação que cada linha contém. Como exemplo, observe a coluna indicadora da tabela 2.1, apresentada em seguida.

Masculino
Feminino
Total

É fácil verificar que a primeira linha da tabela 2.1 apresenta dados sobre sexo masculino, a segunda linha apresenta dados sobre sexo feminino

e a terceira linha apresenta dados sobre ambos os sexos, ou seja, apresenta o total.

Além de título, corpo, cabeçalho e coluna indicadora, as tabelas ainda podem conter os seguintes elementos: fonte, notas e chamadas.

Fonte: é a entidade responsável pelo fornecimento dos dados. Observe a tabela 2.1. Como indica a fonte, os dados apresentados nesta tabela foram fornecidos pela Fundação Instituto Brasileiro de Geografia e Estatística, IBGE. A fonte é dada no rodapé da tabela. Não se indica a fonte nos casos em que a tabela é apresentada pelo próprio pesquisador, ou pelo próprio grupo de pesquisadores, ou pela própria instituição que obteve os dados. Também não se indica a fonte quando os dados são fictícios, isto é, quando os dados são simulados, ou foram inventados, por determinada razão. Por razões didáticas, isto é, para esclarecer certos conceitos de Estatística, muitas vezes são usados dados fictícios. É claro que, nesses casos, é preciso informar, no texto ou na tabela, que os dados não são reais, mas não é preciso colocar essa informação como fonte da tabela.

Notas: são informações de natureza geral que servem para esclarecer o conteúdo das tabelas ou para explicar o método utilizado no levantamento dos dados. As notas são colocadas no rodapé da tabela, logo após a fonte, se houver. Como exemplo, observe a tabela 2.2. A nota apresentada nessa tabela esclarece que parte das informações levantadas no censo de 1950 foi perdida.

Tabela 2.2

População brasileira presente, com 15 anos e mais,
segundo a alfabetização, de acordo com
o censo demográfico de 1950

Alfabetização	Frequência	Percentual
Sabem ler e escrever	14 916 779	49,3
Não sabem ler e escrever	15 272 632	50,5
Sem declaração	60 012	0,2
Total	30 249 423	100,0

FORNTE: IBGE (1977).

NOTA: Excluídas 31.960 pessoas recenseadas nos Estados de Minas Gerais (10.481), São Paulo (7.588) e Paraná (13.911), cujas declarações não foram apuradas por extravio do material de coleta.

Chamadas: são informações de natureza específica que servem para explicar ou conceituar determinados dados. Como exemplo, observe os dados fictícios apresentados na tabela 2.3.

Tabela 2.3

Distribuição dos alunos do curso colegial "Maria de Lima Barbosa", segundo o hábito de fumar e o sexo

Hábito	Sexo	
	Masculino	Feminino
Fumantes (1)	50	60
Não-Fumantes	432	439

(1) Também foram considerados fumantes os alunos que diziam fumar apenas eventualmente.

A chamada apresentada na tabela 2.3 explica que foram considerados fumantes não apenas os alunos que se declararam fumantes, mas também aqueles que declararam fumar "de vez em quando".

As chamadas devem obedecer às seguintes regras:

- toda chamada deve ser indicada por algarismo arábico, entre parênteses, embora seja comum a indicação de chamada por asteriscos. O algarismo arábico que indica a chamada deve ser escrito à esquerda da casa, quando a chamada é feita no corpo da tabela, e à direita da coluna indicadora, quando a chamada é feita nessa coluna;
- se houver mais de uma chamada na mesma tabela, elas devem ser numeradas sucessivamente, de cima para baixo e da esquerda para a direita. Como exemplo, observe a tabela 2.4. Nessa tabela, que apresenta dados fictícios, existem duas chamadas, distribuídas em linhas diferentes e em colunas diferentes;
- As chamadas são colocadas no rodapé da tabela, em ordem numérica e separadas por pontos;
- quando a tabela ocupa várias páginas, as chamadas devem ser apresentadas na página em que aparecem.

Tabela 2.4

Preço médio de um produto durante uma semana, em 12 supermercados

75	(2) 50	81
78	77	79
(1) 73	77	78
80	76	80

(1) Este valor se refere apenas aos três primeiros dias da semana porque nos demais o produto não estava disponível. (2) Este preço foi mantido durante a semana porque o produto estava em "oferta".

É preciso deixar claro, neste ponto, que toda tabela tem, necessariamente, título e corpo. Os outros componentes podem até não existir, como acontece na tabela 2.4, que não apresenta nem cabeçalho, nem coluna indicadora. Também é preciso deixar claro que a existência, ou não, de determinado componente depende apenas da natureza do problema, nunca do gosto de quem faz a tabela.

2.2 NORMAS PARA A APRESENTAÇÃO DE TABELAS

Na apresentação de tabelas devem ser observadas algumas normas:

- as tabelas devem ser delimitadas, no alto e em baixo, por traços horizontais. Esses traços podem ser mais fortes do que os traços feitos no interior da tabela;
- as tabelas não devem ser delimitadas, à direita e à esquerda, por traços verticais;
- o cabeçalho deve ser delimitado por traços horizontais;
- podem ser feitos traços verticais no interior da tabela, separando as colunas;
- as tabelas devem ter significado próprio, isto é, devem ser entendidas mesmo quando não se lê o texto em que estão apresentadas;
- as tabelas devem ser numeradas com algarismos arábicos. Pode ser adotada a numeração progressiva por seções. Então, a tabela 2.3, por exemplo, seria a terceira tabela da segunda seção;
- a finalidade principal das tabelas é apresentar dados numéricos. Então não devem ser feitas tabelas que exibam mais casas sem números do que casas com números;
- a tabela deve ser colocada no texto em posição tal que não exija, para a leitura, rotação da página. Caso isso não seja possível, a tabela deve ser escrita em posição tal que exija, para a leitura, rotação da página em sentido horário;
- quando dois ou mais tipos de informação tiverem sido agrupados em um só conjunto, esse conjunto entra na tabela sob a denominação "outros". Veja o exemplo de dados fictícios apresentado na tabela 2.5. Na especificação "outros" está incluído todo o material disponível na biblioteca que não foi catalogado como livro, apostila ou periódico;

Tabela 2.5

Material disponível na biblioteca da Escola de 2.º grau "João de Melo", para a consulta dos alunos, no ano de 1985

Especificação	Número
Livros	120
Apostilas	60
Periódicos	10
Outros	5

- j) as tabelas podem apresentar dados obtidos através de perguntas ou de entrevistas. Nesses casos, se parte das pessoas não respondeu a determinada pergunta, essa informação deve ser apresentada na tabela sob a especificação "sem declaração". Como exemplo, observe a tabela 2.6;

Tabela 2.6

População residente no Brasil, com cinco anos e mais, segundo a alfabetização, de acordo com o censo demográfico de 1980

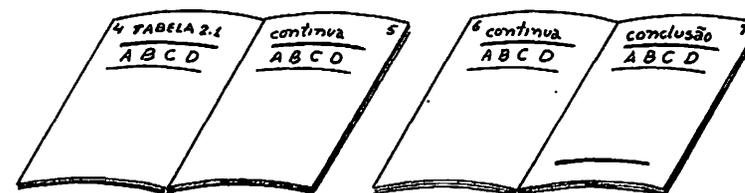
Alfabetização	Frequência	Percentual
Sabem ler e escrever	69 703 993	68,0
Não sabem ler e escrever ..	32 731 347	31,9
Sem declaração	143 666	0,1

FONTE: IBGE (1983).

- l) são comuns os traços verticais no interior do cabeçalho, para separar as especificações. Também podem ser feitas linhas pontilhadas na coluna indicadora, se na coluna são usadas palavras. Essas linhas vão até o traço que indica o início da primeira coluna do corpo da tabela. Esta norma deve ser seguida por órgãos ou instituições oficiais, mas não precisa ser obedecida na apresentação de tabelas em trabalhos acadêmicos ou escolares;
- m) nenhuma célula da tabela deve ficar em branco. Toda célula deve apresentar um número ou um sinal, conforme a convenção:
- ... O dado é desconhecido, podendo o fenômeno existir ou não existir.
 - O dado não existe.
 - 0; 0,0; 0,00 O dado existe, mas seu valor é inferior à metade da unidade de medida adotada na tabela.
 - x Dado omitido, a fim de evitar a individualização da informação.

Esta convenção, adotada pelo IBGE, é recomendada para a organização de tabelas apresentadas por órgãos ou instituições oficiais, mas não precisa ser necessariamente seguida quando se apresentam tabelas em trabalhos acadêmicos ou escolares;

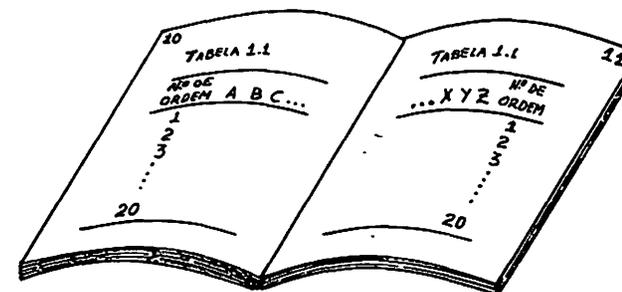
- n) as tabelas muito longas precisam ser apresentadas em duas ou mais páginas. Nesses casos, o cabeçalho deve ser repetido em todas as páginas, mas o título é escrito apenas na primeira. Nas demais páginas escreve-se, em lugar do título, "continua" e na última escreve-se "conclusão." Só deve ser feito o traço inferior, que delimita a tabela, na última página. Observe o esquema dado em seguida: na página 4 está apresentado o título, nas páginas 5 e 6 está escrito "continua" e, na última página, "conclusão":



- o) as tabelas com muitas linhas e poucas colunas ficam com melhor aspecto quando as colunas são organizadas em duas ou mais partes, escritas lado a lado. Essas partes são separadas por dois traços verticais. Nesses casos, o cabeçalho deve indicar o conteúdo das colunas em todas as partes, como mostra o esquema dado em seguida:

Idade	Peso	Idade	Peso	Idade	Peso

- p) as tabelas com muitas colunas precisam ocupar duas páginas que se confrontam como, por exemplo, as páginas 10 e 11. Para facilitar a leitura, todas as linhas devem receber um número de ordem. O número de ordem deve ser escrito na primeira coluna da página à esquerda e na última coluna da página à direita. Veja o esquema dado em seguida:



- q) o total é geralmente apresentado na última linha, entre dois traços horizontais, como mostra a tabela 2.2, embora também possa ser apre-

sentado na primeira linha. O IBGE apresenta os dados de todo o Brasil na primeira linha, com a entrada "Brasil", como mostra a tabela 2.7.

Tabela 2.7

Domicílios particulares permanentes, segundo a situação, no Brasil, em 1980

Situação	Frequência	Percentual
Brasil	25 210 639	100,0
Urbana	17 770 981	70,5
Rural	7 439 658	29,5

FONTE: IBGE (1983).

2.3 SÉRIES ESTATÍSTICAS

As tabelas também servem para apresentar séries estatísticas. Existem três tipos básicos de séries estatísticas: cronológica, geográfica e categórica.

Série cronológica: é feita para apresentar dados observados ao longo do tempo. Como exemplo, veja a tabela 2.8. Quando se constrói uma série cronológica é preciso observar rigorosamente a ordenação de dias, de meses, de anos. Assim, numa série cronológica, 1980 antecede 1981, nunca o contrário.

Tabela 2.8

Produção de automóveis no Brasil no período 1972-1984

Ano	Número de automóveis
1972	415 164
1973	467 575
1974	531 839
1975	542 203
1976	526 944
1977	464 505
1978	535 531
1979	547 626
1980	600 706
1981	406 016
1982	475 112
1983	576 356
1984	536 755

FONTE: Fundação Getúlio Vargas (1986).

Série geográfica: é feita para apresentar dados de diferentes regiões geográficas. Quando se constrói uma série geográfica deve-se escrever países segundo o continente; municípios e cidades segundo as unidades da federação; distritos e vilas segundo o município. Como exemplo, observe a série geográfica apresentada na tabela 2.9:

Tabela 2.9

Casamentos registrados no Brasil, segundo a região, em 1983

Região	Casamentos registrados
Norte	30 485
Nordeste	211 162
Sudeste	408 271
Sul	160 566
Centro-Oeste	55 706
Total	866 190

FONTE: IBGE (1984).

Série categórica: é feita para apresentar dados que se distribuem em diferentes categorias. Como exemplo, veja a tabela 2.10.

Tabela 2.10

População residente no Brasil, segundo a cor, de acordo com o censo demográfico de 1980

Cor	Frequência	Percentual
Branca	64 540 467	54,2
Preta	7 046 906	5,9
Amarela	672 251	0,6
Parda	46 233 531	38,8
Sem declaração	517 897	0,4
Total	119 011 052	100,0

FONTE: IBGE (1983).

NOTA: A pequena diferença entre o total desta tabela e o total da tabela 2.1 se deve ao fato de os dados apresentados aqui terem sido determinados pelo IBGE a partir de uma amostra.

Podem ser feitas combinações de dois tipos de série. As combinações de dois tipos de série recebem o nome de *séries mistas*. Assim, os dados de diferentes regiões ao longo do tempo constituem uma série mista geográfico-cronológica. Os dados de diferentes regiões, distribuídos em diferentes categorias, constituem uma série mista geográfico-categórica. Um exemplo deste tipo de série mista é apresentado na tabela 2.11. Os dados de diferentes categorias, ao longo do tempo, constituem uma série mista categórico-cronológica.

Tabela 2.11

População residente no Brasil por sexo, segundo a região, de acordo com o censo demográfico de 1980

Região	Sexo	
	Masculino	Feminino
Norte	2 989 906	2 890 800
Nordeste	17 043 570	17 771 869
Sudeste	25 735 943	26 001 205
Sul	9 534 287	9 497 703
Centro-Oeste	3 839 127	3 706 642
Total	59 142 833	59 868 219

FONTE: IBGE (1983).

2.4. EXERCÍCIOS

1. Construa uma tabela para mostrar que no curso do Colégio "Dr. J. Alves" o número de alunos matriculados nas 1.ª, 2.ª e 3.ª séries era, respectivamente, 40, 35 e 29 em 1984 e 42, 36 e 32 em 1985.
2. Construa uma tabela para mostrar que, do total de 40 alunos matriculados por série, nas quatro primeiras séries da Escola de 1.º Grau "D. Alcina", em 1985, foram reprovados 8, 5, 4 e 4, respectivamente nas 1.ª, 2.ª, 3.ª e 4.ª séries. Apresente na tabela as frequências e as frequências relativas.
3. Construa uma tabela para mostrar que, segundo o IBGE (Anuário Estatístico do Brasil — 1984), o pessoal docente no país era em número de 942 955, 960 905 e 967 975, em 1981, 1982 e 1983 respectivamente.

4. Faça uma tabela para mostrar que o número de estabelecimentos públicos de ensino de 1.º grau, no Brasil, era de 182 836 em 1981, 187 567 em 1982 e 180 983 em 1983 e que, nesses mesmos anos, o número de estabelecimentos particulares de ensino de 1.º grau era de 9 793, 10 463 e 9 934 respectivamente. Os dados têm como fonte o IBGE (Anuário Estatístico do Brasil — 1984).
5. De acordo com o IBGE (Anuário Estatístico do Brasil — 1977), a altura média, em centímetros, e o peso médio, em quilogramas, de crianças com idades entre 3 e 10 anos, inclusive, na região Sul do Brasil em 1975 era, em cada idade, respectivamente:

Altura média: 94,1; 100,8; 107,4; 113,1; 118,0; 124,1; 129,3; 133,7.

Peso médio: 14,4; 16,0; 18,0; 19,9; 21,6; 24,1; 26,5; 29,0.

Faça uma tabela para mostrar altura média e peso médio, segundo a idade.

6. Imagine que foi obtida a opinião de 1 000 pessoas a respeito da liberação de determinado filme para exibição em televisão. Dessas 1 000 pessoas, 432 mostravam-se favoráveis, 322 eram contrárias, 122 não quiseram declarar a opinião e as restantes disseram não ter opinião. Mostre esses dados numa tabela. Apresente também as frequências relativas.
7. Imagine que, das 1 000 pessoas entrevistadas cujas respostas foram apresentadas no exercício anterior, 500 eram homens e 500 eram mulheres. Do total de homens, 289 mostravam-se favoráveis, 120 eram contrários, 78 não quiseram declarar a opinião e os restantes disseram não ter opinião. Construa uma tabela para apresentar a distribuição das respostas segundo o sexo.
8. Jogue um dado 30 vezes e anote os resultados. Construa depois uma tabela para apresentar os resultados obtidos. Discuta então a possibilidade de o dado ser viciado.
9. Pergunte o time favorito de futebol para cada um de seus colegas de classe. Depois faça uma tabela para apresentar os dados.
10. Pergunte para pessoas de ambos os sexos qual é o tipo de programa que elas preferem ver na televisão. Faça uma tabela para apresentar os dados, relativos a cada sexo.

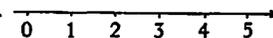
3

Apresentação Gráfica de Dados

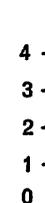
Os dados estatísticos, apresentados em tabelas, também podem ser expostos em gráficos. Desde que não haja necessidade de grande precisão, os gráficos dão, melhor do que as tabelas, visão de tendências e ajudam a interpretar um fenômeno.

Na apresentação gráfica de dados são necessários alguns cuidados:

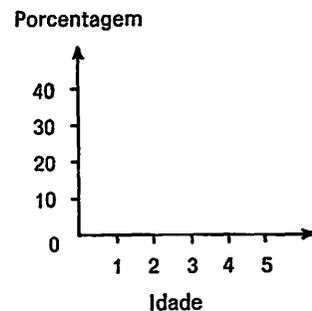
- a) todo gráfico deve ter título e escala, para que possa ser interpretado sem que haja necessidade de esclarecimentos adicionais no texto;
- b) o título do gráfico pode ser escrito acima ou abaixo do gráfico. O IBGE escreve o título acima do gráfico;
- c) no eixo das abscissas, a escala cresce da esquerda para a direita e é escrita embaixo do eixo, como mostra o esquema:



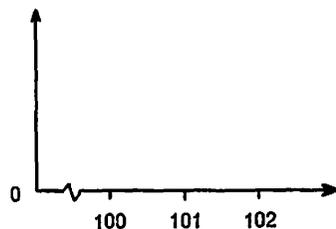
- d) no eixo das ordenadas, a escala cresce de baixo para cima e é escrita à esquerda do eixo, como mostra o esquema:



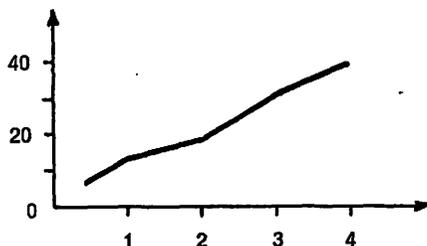
- e) podem ser feitas setas para indicar a direção dos eixos;
- f) as variáveis representadas em cada eixo devem ser claramente identificadas. No eixo das ordenadas escreve-se o nome da variável na extremidade do eixo. No eixo das abscissas escreve-se o nome da variável embaixo da escala, como mostra o esquema:



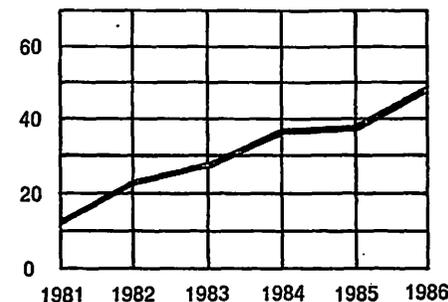
- g) a escala deve iniciar-se na origem do sistema de eixos cartesianos. Quando os valores iniciais dos dados são muito altos (como, por exemplo, 100, 101 etc.), deve ser feita uma interrupção no eixo, com indicação clara da posição do zero. Veja o esquema. Esta recomendação não se aplica nos casos em que se apresentam datas, em qualquer dos eixos.



- h) o sistema de eixos cartesianos e as linhas auxiliares devem ter traçado mais leve do que a parte do gráfico que se pretende evidenciar:



- i) para facilitar a leitura, podem ser feitas linhas auxiliares. Nesses casos, o gráfico é feito dentro de um retângulo, como mostra o esquema:



- j) os gráficos podem exibir, em rodapé, a fonte, isto é, a instituição, o pesquisador, ou o grupo de pesquisadores que forneceu o gráfico ou os dados que permitiram a construção do gráfico.

3.1 GRÁFICO DE BARRAS

O gráfico de barras é usado para apresentar séries cronológicas, geográficas e categóricas. Para entender como se faz um gráfico de barras, primeiro observe a série cronológica apresentada na tabela 3.1.

Tabela 3.1

População presente no Brasil, segundo o ano do censo demográfico

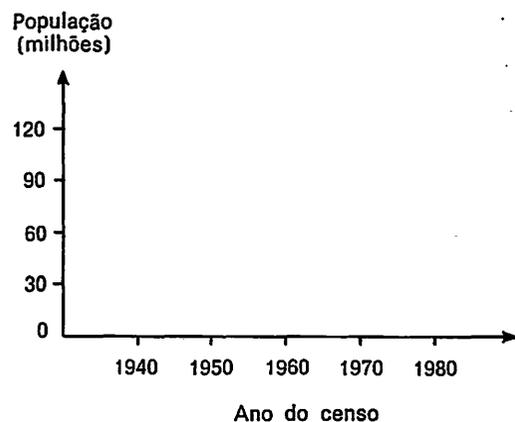
Ano do Censo Demográfico	População
1940	41 236 315
1950	51 944 397
1960	70 191 370
1970 (1)	93 139 037
1980 (1)	119 002 706

FONTE: IBGE (1984).

(1) População residente.

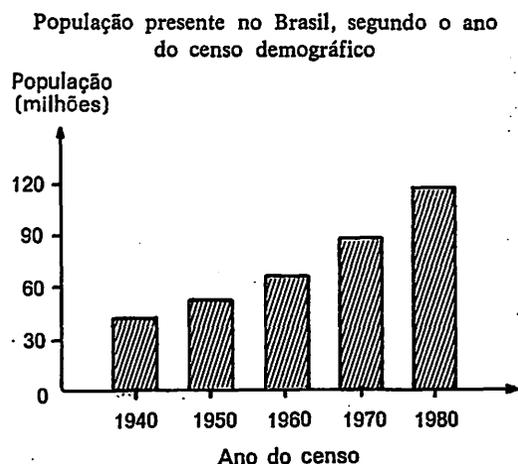
Para apresentar os dados da tabela 3.1 em um gráfico de barras siga os seguintes passos:

- trace o sistema de eixos cartesianos;
- escreva os anos do censo no eixo das abscissas; a população será apresentada no eixo das ordenadas, como mostra o esquema:



- construa barras retangulares para representar a população em cada ano de censo. As barras terão bases de mesma largura, mas a altura de cada barra será dada pela população no ano do censo;
- coloque título na figura.

Figura 3.1



A apresentação gráfica de séries geográficas e categóricas, através de gráficos de barras, é feita da mesma forma. Observe os dados da série geográfica apresentados na tabela 3.2. As regiões podem ser apresentadas no eixo das abscissas; o número de pessoas de cada região pode ser apre-

sentado no eixo das ordenadas, como mostra a figura 3.2. Para facilitar a leitura, foram feitas linhas auxiliares.

Tabela 3.2

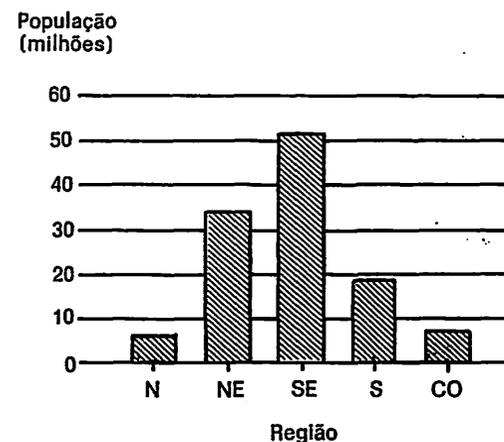
População residente no Brasil, segundo a região, de acordo com o censo demográfico de 1980

Região	População
Norte	5 880 268
Nordeste	34 812 356
Sudeste	51 734 125
Sul	19 031 162
Centro-oeste	7 544 795

FONTE: IBGE (1984).

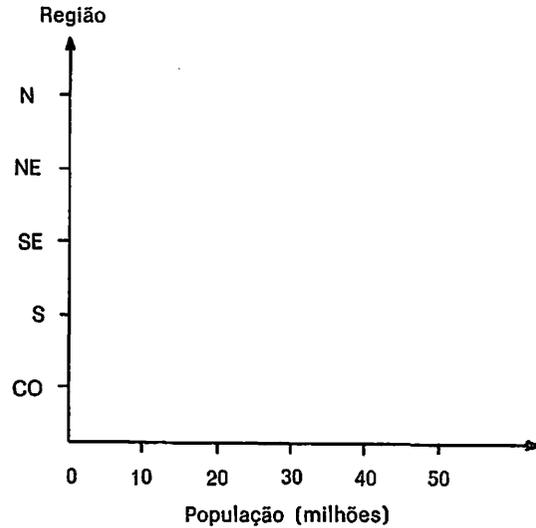
Figura 3.2

População residente no Brasil, segundo a região, de acordo com o censo demográfico de 1980.



É preciso deixar claro, no entanto, que no caso dos dados da tabela 3.2 as barras também podem ser construídas em posição horizontal. Para isso:

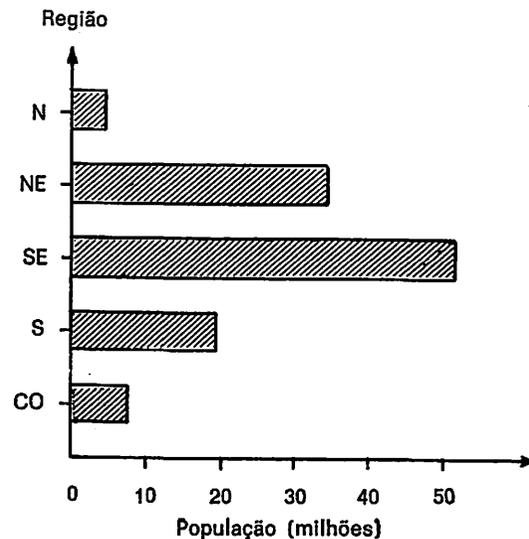
- trace o sistema de eixos cartesianos;
- apresente as regiões no eixo das ordenadas e a população no eixo das abscissas;



- c) construa barras retangulares com bases iguais e comprimento dado pelo número de pessoas de cada região;
- d) escreva o título.

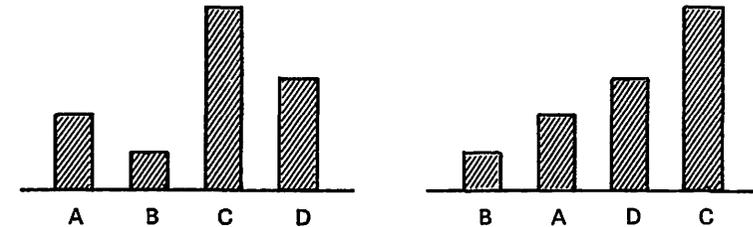
Figura 3.3

População residente no Brasil, segundo a região, de acordo com o censo demográfico de 1980



É mais comum a apresentação das barras em posição vertical, como foi feito na figura 3.2. No entanto, as barras em posição horizontal facilitam a identificação das categorias, principalmente nos casos em que essas categorias têm nomes muito longos. Isto porque os nomes podem ser escritos em posição horizontal e por extenso, dispensando o uso de abreviaturas e de legendas.

Os dados de séries cronológicas devem ser apresentados em seqüência. Se, porém, os dados são de uma série categórica, ou mesmo de uma série geográfica, a apresentação das barras em ordem crescente (ou decrescente) ajuda na comparação da importância relativa das categorias. Como mostra o esquema, é mais fácil interpretar a figura apresentada à direita do que a figura apresentada à esquerda.

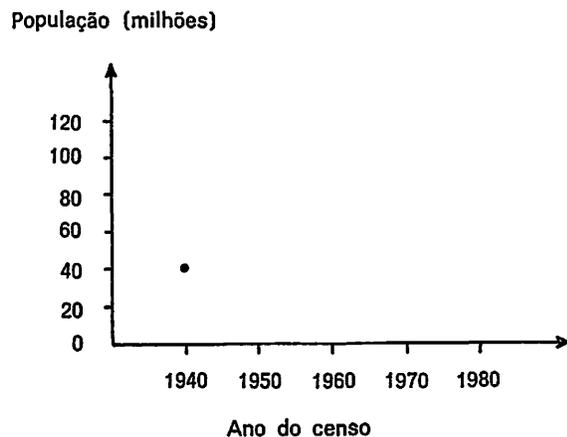


Os números e as palavras devem ser escritos fora das barras. Nos casos em que os nomes são muito extensos recomenda-se usar uma legenda explicativa.

3.2 GRÁFICO DE LINHAS

O gráfico de linhas é usado para apresentar séries cronológicas. Observe os dados da tabela 3.1. Para fazer o gráfico de linhas para esse exemplo:

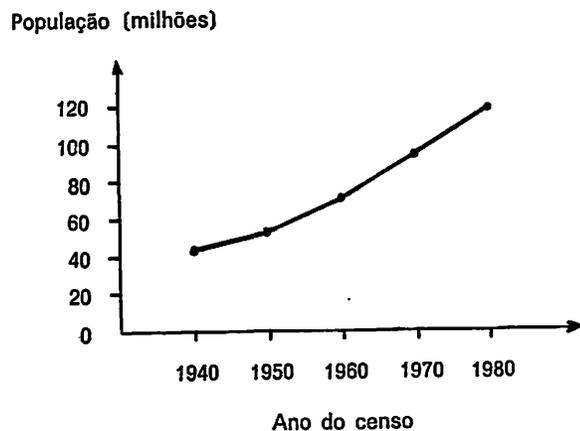
- a) trace o sistema de eixos cartesianos;
- b) no eixo das abscissas coloque os anos do censo. No eixo das ordenadas apresente os dados da população;
- c) faça corresponder um ponto para cada par de valores da série. Por exemplo, para o par (1940, 41 milhões) faça um ponto, como mostra o esquema;



- d) una os pontos por segmentos de reta;
- e) coloque título na figura:

Figura 3.4

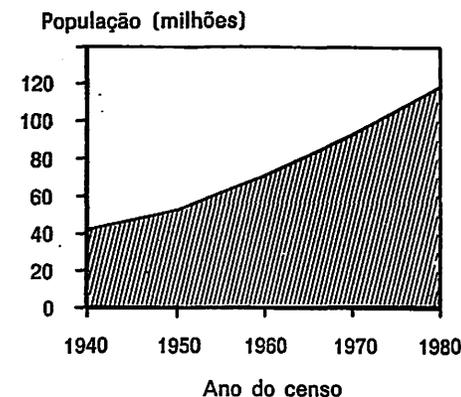
População presente no Brasil, segundo o ano do censo demográfico



O aspecto visual do gráfico de linhas melhora se a figura for fechada e hachurada. Observe a figura 3.5, que apresenta os mesmos dados expostos na figura 3.4. Note que a figura 3.5 tem melhor aspecto.

Figura 3.5

População presente no Brasil, segundo o ano do censo demográfico.



3.3 GRÁFICO RETANGULAR DE COMPOSIÇÃO

O gráfico retangular de composição é usado para evidenciar a composição percentual de uma amostra ou de uma população. Para entender como se faz um gráfico retangular de composição observe os dados fictícios apresentados na tabela 3.3.

Tabela 3.3

Distribuição de uma amostra de 50 alunos universitários, segundo o interesse em aprender computação

Interesse	Número de alunos	Porcentagem
Têm	25	50
Não têm	15	30
Não sabem	10	20
Total	50	100

Para fazer um gráfico retangular de composição:

- a) trace um retângulo, de comprimento conveniente, para representar o total. No caso do exemplo, o total corresponde a 50 alunos. Para representar esse total foi feito um retângulo com 7 cm de comprimento:



- b) se 50 alunos são representados por 7 cm, os 25 alunos que têm interesse em aprender computação são representados por um comprimento x , tal que:

$$\frac{7}{50} = \frac{x}{25}$$

Então:

$$x = \frac{7 \cdot 25}{50} = 3,5$$

Da mesma forma, os 15 alunos que não têm interesse em aprender computação são representados, no retângulo, pelo comprimento y , tal que:

$$\frac{7}{50} = \frac{y}{15}$$

Donde:

$$y = \frac{7 \cdot 15}{50} = 2,1$$

Os 10 alunos que ainda não sabem se têm, ou não têm, interesse em aprender computação devem ser representados no gráfico, pelo comprimento z , tal que:

$$\frac{7}{50} = \frac{z}{10}$$

Donde:

$$z = \frac{7 \cdot 10}{50} = 1,4$$

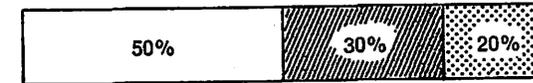
- c) feitos os cálculos, marque os valores obtidos dentro do retângulo, separando as partes por traços:



- d) faça um tracejado diferente em cada parte, para melhor distingui-las;
e) coloque título e legenda no gráfico. Para maior clareza, indique, no retângulo, as porcentagens correspondentes a cada categoria:

Figura 3.6

Distribuição de uma amostra de 50 alunos universitários, segundo o interesse em aprender computação



□ têm ▨ não têm ▩ não sabem

3.4 GRÁFICO DE SETORES

O gráfico de setores é usado para comparar proporções. Para entender como se faz um gráfico de setores, primeiro observe os dados fictícios apresentados na tabela 3.4.

Tabela 3.4

Distribuição dos alunos da 1.ª série do Grupo Escolar "Maria Conceição", segundo a situação em relação às notas

Situação	Frequência	Percentual
Promovido	12	50,0
Em recuperação	8	33,3
Retido	4	16,7
Total	24	100,0

A comparação das proporções, no caso deste exemplo e de outros similares, fica muito fácil se for construído um gráfico de setores. Para isso:

- a) trace uma circunferência. A área do círculo representará o total, isto é, os 24 alunos, ou 100%;
b) lembre-se de que uma circunferência tem 360°. Então, se aos 24 alunos correspondem 360°, aos 12 alunos promovidos corresponderá um setor cujo ângulo x é dado por:

$$\frac{360}{24} = \frac{x}{12}$$

Então:

$$x = \frac{12 \cdot 360}{24} = 180^\circ$$

Da mesma forma, aos oito alunos em recuperação corresponde um setor cujo ângulo y é dado por:

$$\frac{360}{24} = \frac{y}{8}$$

Donde:

$$y = \frac{8 \cdot 360}{24} = 120^\circ$$

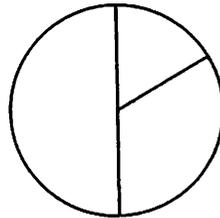
Os quatro alunos retidos são representados por um setor de ângulo z tal que:

$$\frac{360}{24} = \frac{z}{4}$$

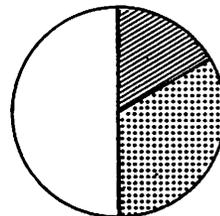
Donde:

$$z = \frac{4 \cdot 360}{24} = 60^\circ$$

- c) marque os valores dos ângulos calculados na circunferência e trace os raios, separando os setores:



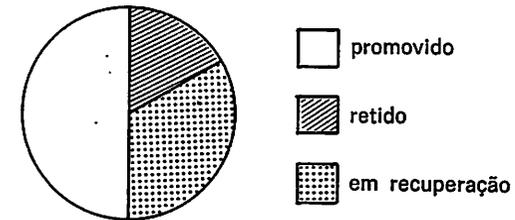
- d) faça um tracejado diferente em cada setor, para facilitar a distinção:



- e) coloque título e legenda no gráfico.

Figura 3.7

Distribuição dos alunos da 1.ª série do Grupo Escolar "Maria Conceição", segundo a situação em relação às notas



3.5 GRÁFICO POLAR

O *gráfico polar* é usado para representar dados que variam ao longo do tempo, desde que esse tempo se limite a determinado período. São feitos gráficos polares para representar a variação da temperatura ao longo do dia, a quantidade vendida de determinado produto ao longo da semana, a variação da chuva ao longo do ano etc. Para mostrar como se constrói um gráfico polar serão utilizados os dados apresentados na tabela 3.5.

Tabela 3.5

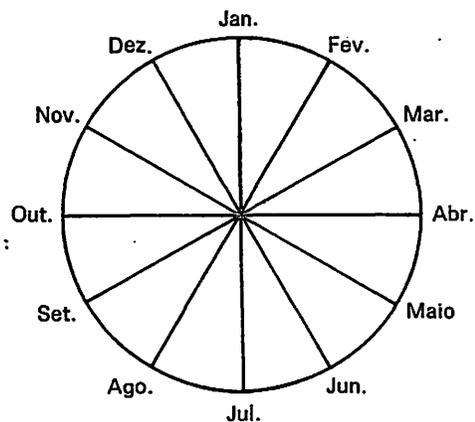
Altura total da precipitação pluviométrica, segundo o mês, em Manaus, no ano de 1983

Mês	Altura total (em mm)
Janeiro	32,6
Fevereiro	107,5
Março	418,2
Abril	192,4
Mai	171,5
Junho	99,4
Julho	134,1
Agosto	134,6
Setembro	99,2
Outubro	195,8
Novembro	69,5
Dezembro	372,1

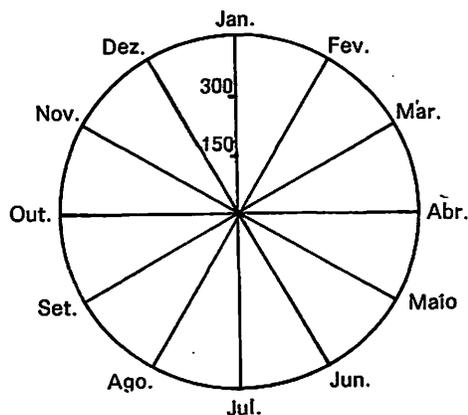
FONTÉ: IBGE (1984).

Para apresentar os dados da tabela 3.5 em gráfico polar devem ser seguidos os passos:

- trace uma circunferência;
- divida a circunferência em 12 partes iguais;
- separe essas partes por raios e faça cada raio representar um mês, como mostra o esquema:



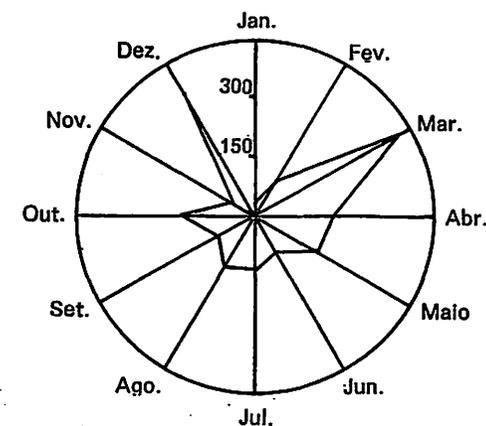
- faça uma escala:



- marque um ponto em cada raio para representar o valor observado naquele mês. Note, por exemplo, que o valor observado em março é 418,2. Una os pontos por segmentos de reta. Não una dezembro a janeiro, porque depois de dezembro de 1983 viria janeiro de 1984, não janeiro de 1983, apresentado na figura;
- Coloque título no gráfico:

Figura 3.8

Altura total, em milímetros, da precipitação pluviométrica, segundo o mês, em Manaus, em 1983



3.6 GRÁFICOS COMPARATIVOS

Muitas vezes são desenhados dois gráficos, lado a lado, para melhor estabelecer a comparação de um fenômeno. Outras vezes, os dados obtidos em situações distintas são sobrepostos na mesma figura, para evidenciar a comparação. Os gráficos individuais são desenhados da maneira comum, mas são colocados juntos para facilitar a comparação.

Observe a série cronológica apresentada na tabela 3.6. Para a apresentação gráfica desses dados pode ser feito o gráfico de barras, apresentado na figura 3.9, ou o gráfico de linhas, apresentado na figura 3.10. Nessas duas figuras se estabelece uma comparação entre população urbana e população rural.

Tabela 3.6

População residente no Brasil, urbana e rural, segundo o ano do censo demográfico

Ano do censo demográfico	População	
	Urbana	Rural
1940 (1)	12 880 182	28 356 133
1950 (1)	18 782 891	33 161 506
1960	31 303 034	38 767 423
1970	52 084 984	41 054 053
1980	80 436 409	38 566 297

FONTE: IBGE (1984).
(1) População presente.

Figura 3.9

População residente no Brasil, urbana e rural, segundo o ano do censo demográfico

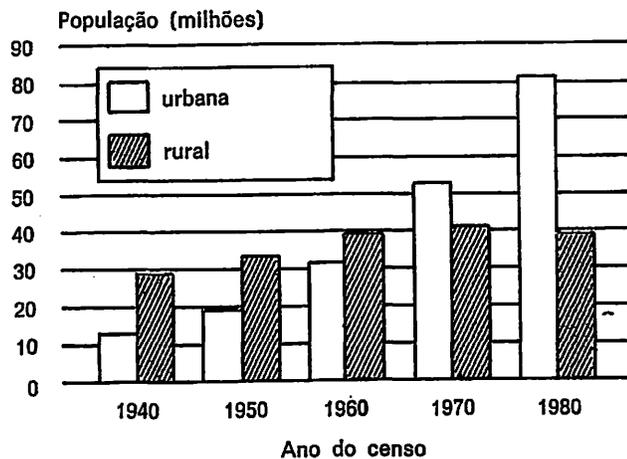
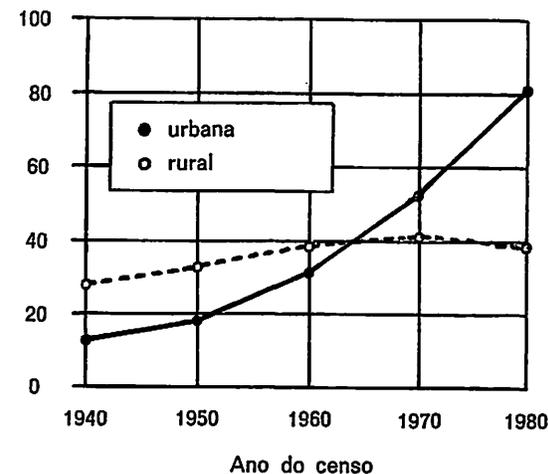


Figura 3.10

População residente no Brasil, urbana e rural, segundo o ano do censo demográfico.

População (milhões)



As figuras 3.9 e 3.10 mostram, melhor do que a tabela 3.6, que as populações urbana e rural têm, no Brasil, tendências diferentes de crescimento. Enquanto a população das cidades aumentou rapidamente, a população do campo cresceu cada vez mais lentamente até 1970 e diminuiu na última década.

Outro exemplo de gráficos comparativos pode ser obtido a partir da tabela 3.7. Para mostrar a composição percentual da população, em relação ao analfabetismo, tanto pode ser feito um gráfico retangular de composição, como um gráfico de setores. Como é dada a composição percentual em três décadas diferentes, para a comparação colocam-se os gráficos lado a lado, como mostram as figuras 3.11 e 3.12.

Tabela 3.7

População residente no Brasil, com 15 anos e mais, segundo a alfabetização e o ano do censo demográfico

Censo	Alfabetização	População	Percentual
1960 (1)	Sabem ler e escrever	24 259 284	60,2
	Não sabem ler e escrever	15 964 852	39,6
	Sem declaração	54 466	0,1
	Total	40 278 602	100,0
1970	Sabem ler e escrever	35 586 771	65,9
	Não sabem ler e escrever	18 146 977	33,6
	Sem declaração	274 856	0,5
	Total	54 008 604	100,0
1980	Sabem ler e escrever	54 793 268	74,5
	Não sabem ler e escrever	18 716 847	25,5
	Sem declaração	31 828	0,0
	Total	73 541 943	100,0

FONTE: IBGE (1984).
(1) População presente.

Figura 3.11

População residente no Brasil, com 15 anos e mais, segundo a alfabetização e o ano do censo demográfico

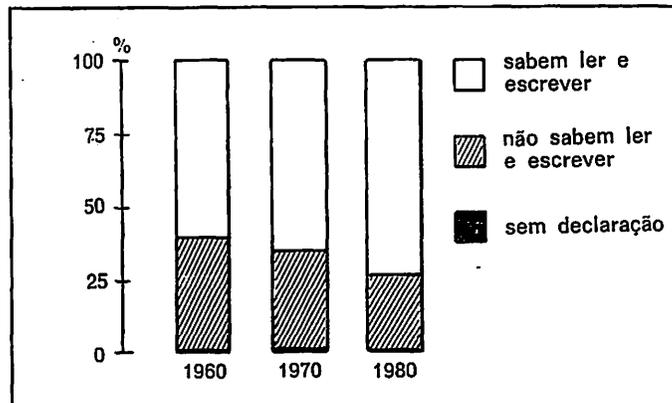
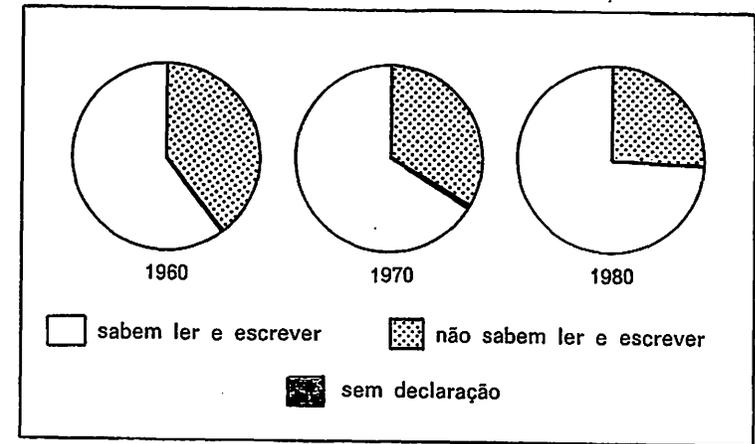


Figura 3.12

População residente no Brasil, com 15 anos e mais, segundo a alfabetização e o ano do censo demográfico



As figuras 3.11 e 3.12 mostram, claramente, que a porcentagem de analfabetos no Brasil diminuiu nas últimas décadas. No entanto, dado o crescimento da população, o número de analfabetos aumentou no período, conforme mostra a tabela 3.7.

3.7 EXERCÍCIOS

1. Faça um gráfico de barras para apresentar os valores de densidade demográfica, segundo a região.

Tabela 3.9

Densidade demográfica no Brasil, segundo a região, de acordo com o censo demográfico de 1980

Região	Densidade (hab./km ²)
Norte	1,65
Nordeste	22,57
Sudeste	56,31
Sul	33,86
Centro-Oeste	4,01

FONTE: IBGE (1984).

- Imagine que se perguntou a 1 000 pessoas se elas acreditavam em horóscopos. Dessas 1 000 pessoas, 488 disseram acreditar, 292 disseram não acreditar, 120 disseram que tinham dúvidas e as restantes expressaram opiniões diversas. Faça um gráfico de barras para apresentar esses dados.
- Construa um gráfico de linhas para mostrar que, na Escola de 1.º Grau "D. Pedro II", a taxa de evasão escolar (porcentagem de alunos que abandonaram a escola) foi 12,1; 11,3; 10,7; 15,0; 14,7 e 10,5 em 1980, 1981, 1982, 1983, 1984 e 1985, respectivamente.
- Faça um gráfico de linhas para apresentar o crescimento em altura de crianças do sexo masculino.

Altura média de crianças do sexo masculino residentes no Rio de Janeiro em 1975, segundo a idade

Idade (anos)	Altura média (cm)
7	119,7
8	124,4
9	129,3
10	134,1
11	139,2
12	143,2

FONTE: IBGE (1977).

- Faça gráficos retangulares de composição para apresentar os valores de mortalidade proporcional, segundo a idade, nas cinco regiões do Brasil. Note que é preciso fazer um gráfico retangular de composição para cada região. A figura terá então cinco gráficos, um para cada região.

Mortalidade proporcional por região do Brasil, segundo os grupos de idade, de acordo com o censo demográfico de 1980

Grupos de idade	Região				
	Norte	Nordeste	Sudeste	Sul	Centro-Oeste
Méno de 1 ano ..	31,7	34,7	20,4	18,2	21,6
De 1 a 4 anos ...	8,1	7,9	3,1	3,4	4,8
De 5 a 19 anos ..	6,5	4,1	3,6	4,3	6,1
De 20 a 49 anos ..	19,3	14,2	19,7	18,0	24,7
De 50 anos e mais .	34,4	39,1	53,2	56,1	42,8
Total	100,0	100,0	100,0	100,0	100,0

FONTE: IBGE (1984).

- De acordo com o IBGE (Anuário Estatístico do Brasil — 1984), das 43 235 712 pessoas economicamente ativas no Brasil em 1980, 31 392 986 eram homens e 11 842 726 eram mulheres. Faça um gráfico de setores para apresentar a população economicamente ativa, segundo o sexo.
- Desenhe gráficos polares para apresentar os valores de temperatura segundo o mês, nas cidades do Rio de Janeiro e de São Paulo. Note que é preciso fazer dois gráficos polares, um para cada cidade.

Temperatura do ar, segundo o mês, nas cidades de Rio de Janeiro e São Paulo, em 1983

Mês	Temperatura (°C)	
	Rio de Janeiro	São Paulo
Janeiro	26,6	22,5
Fevereiro	27,0	23,1
Março	26,0	21,2
Abril	24,4	20,7
Maiio	24,0	19,2
Junho	21,5	19,7
Julho	21,6	17,6
Agosto	20,6	17,0
Setembro	20,6	15,9
Outubro	22,6	18,8
Novembro	24,4	21,4
Dezembro	25,2	21,8

FONTE: IBGE (1984).

- Pergunte aos seus colegas de classe qual é a disciplina, entre as ministradas no semestre, que eles mais apreciam. Faça um gráfico de barras para mostrar os resultados.
- Pergunte a pessoas de ambos os sexos se elas assistem ou não a determinado programa de televisão. Depois faça gráficos para mostrar os resultados.
- Jogue uma moeda 20 vezes e anote a proporção de caras. Repita essa experiência mais duas vezes. Depois faça gráficos para mostrar os resultados observados nas três experiências.

4

Tabelas de Distribuição de Frequências

Muitas vezes é preciso apresentar dados numéricos provenientes de grande número de indivíduos ou mesmo de toda a população. A leitura da tabela torna-se então cansativa ou, até mesmo, impossível. No entanto, se os dados se referem a uma única variável como, por exemplo, número de filhos, podem ser apresentados em uma *tabela de distribuição de frequências*. A apresentação de dados, nesse tipo de tabela, permite apreensão rápida do assunto em estudo.

4.1 ORGANIZAÇÃO DA TABELA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

Para entender como se organiza uma tabela de distribuição de frequências observe a tabela 4.1.

Tabela 4.1

Número de filhos vivos, com idade inferior a 18 anos, por empregado da firma "José Souza e irmãos"

2	2	2	1	3
3	0	1	3	3
2	1	0	2	2
0	3	4	4	1
1	4	1	1	6
4	1	1	2	2
2	2	2	0	1
5	3	4	0	3
2	1	2	0	6

Na tabela 4.1 estão apresentados valores fictícios que se referem aos 45 empregados da firma "José Souza e irmãos". Esses valores são o número de filhos, com idade inferior a 18 anos, de cada um dos empregados. Os números são, evidentemente, inteiros. Existem muitos valores iguais, isto é, vários empregados não têm filhos vivos com menos de 18 anos, vários outros têm um filho vivo com menos de 18 anos, e assim por diante. Seria mais fácil observar os dados apresentados na tabela 4.1 se esses dados estivessem apresentados em uma tabela de distribuição de frequências. Para organizar essa tabela:

- a) escreva cada valor observado uma única vez, em coluna e em ordem crescente. No exemplo foram observados valores de 0 até 6, isto é, o número de filhos vivos com menos de 18 anos, por empregado, variou entre 0 e 6. Então escreva:

Número de filhos
0
1
2
3
4
5
6

- b) conte o número de vezes que cada valor se repete. Como mostra a tabela 4.1, o número zero apareceu seis vezes, o número 1 apareceu onze vezes etc. Então escreva o número de vezes que cada valor se repete — ou seja, a frequência — em outra coluna, à direita dos valores observados e na respectiva linha:

Número de filhos	Frequência
0	6
1	11
2	13
3	7
4	5
5	1
6	2

- c) desenhe a tabela de acordo com as normas apresentadas no capítulo 2 e coloque o título.

Tabela 4.2

Distribuição do número de filhos vivos, com idade inferior a 18 anos, por empregado da firma "José Souza e irmãos", em 1985

Número de filhos	Frequência
0	6
1	11
2	13
3	7
4	5
5	1
6	2

A tabela 4.2 é menor do que a tabela 4.1. No entanto, a tabela 4.2 mostra, claramente, que na firma em estudo são comuns empregados com um ou dois filhos vivos com menos de 18 anos e são pouco frequentes empregados com mais de quatro filhos vivos com menos de 18 anos.

4.2 ORGANIZAÇÃO DE DADOS EM CLASSES

Muitas vezes os dados têm precisão maior do que a necessária. Para entender o que significam dados com precisão maior do que a necessária basta lembrar que a idade das pessoas é declarada em anos completos, mas nas pesquisas é frequente referência a faixas de idade. Assim, é comum referência a crianças com menos de dois anos (uma faixa de idade), a pessoas com mais de 40 anos (outra faixa de idade), e assim por diante.

Os dados com precisão maior do que a necessária são apresentados em tabelas de distribuição de frequências. Para entender como se organizam dados nesse tipo de tabelas imagine que foram obtidas as idades de um grupo de pessoas. Pode haver interesse em organizar os dados de idade por faixas, como, por exemplo, de 0 a 4 anos, de 5 a 9 anos, de 10 a 14 anos etc.

Essas faixas recebem o nome genérico de *classes*. Na classe de 0 a 4 anos devem ser incluídas desde as crianças que acabaram de nascer até aquelas que estão na véspera de completar 5 anos. Então a classe de 0 a 4 anos cobre um *intervalo* de 5 anos, isto é, as idades 0, 1, 2, 3 e 4. As outras classes também cobrem um intervalo de 5 anos. Assim, a classe de 5 a 9 anos inclui as idades 5, 6, 7, 8 e 9, ou seja, 5 anos, a classe de

10 a 14 anos inclui as idades 10, 11, 12, 13 e 14, ou seja, 5 anos, e assim por diante. Então, neste exemplo, os intervalos de classe são de 5 anos.

É conveniente, mas não é obrigatório, estabelecer intervalos iguais para todas as classes. Então, quando se organizam dados de idade em classes, é comum escolher intervalos iguais, como, por exemplo, de 0 a 4, de 5 a 9, de 10 a 14, de 15 a 19 etc. Entretanto, também podem ser escolhidos intervalos diferentes, como, por exemplo, de 0 a 4, de 5 a 9, de 10 a 14, de 15 a 19 e, daí em diante, intervalos de 10 anos, isto é, de 20 a 29, de 30 a 39 etc., como mostra o esquema dado em seguida:

INTERVALOS DE CLASSE	
Iguais	Diferentes
de 0 a 4 anos	de 0 a 4 anos
de 5 a 9 anos	de 5 a 9 anos
de 10 a 14 anos	de 10 a 14 anos
de 15 a 19 anos	de 15 a 19 anos
de 20 a 24 anos	de 20 a 29 anos
de 25 a 29 anos	de 30 a 39 anos etc.
de 30 a 34 anos	
de 35 a 39 anos etc.	

Os valores que delimitam os intervalos de classe são denominados *extremos de classe*. É importante deixar claro, quando se escreve um intervalo, se os dados, com valores iguais aos dos extremos de classe, são, ou não, incluídos na classe. É comum escrever as classes da seguinte forma:

0 — 4
5 — 9 etc.

Entretanto, esta forma de apresentação nem sempre é clara. Quando se usa este tipo de apresentação podem surgir dúvidas do tipo "Em que classe deve ser colocada uma criança de 4 anos e meio?" No exemplo em discussão ficou estabelecido que seriam incluídas, na classe de zero a 4 anos, desde as crianças que acabaram de nascer até as crianças que estavam na véspera de completar 5 anos. No caso de valores de idade, esta convenção é clara porque coincide com a convenção social. Ninguém diz ter 30 anos, antes de ter completado 30 anos. Então uma criança com 4 anos e meio seria incluída na classe de 0 a 4 anos, e não na seguinte.

De qualquer forma, evitam-se dúvidas quando se usa outro tipo de indicação para os intervalos de classe. Assim, as classes de 0 a 4 anos, de 5 a 9 anos etc. poderiam ser indicadas na forma:

0 |— 5
5 |— 10 etc.

ou na forma:

[0 — 5)
[5 — 10) etc.

Estas duas maneiras de indicar intervalo são muito claras porque mostram que o intervalo é fechado à esquerda e aberto à direita. Então estão incluídos na classe 0 |— 5, ou [0 — 5), os valores iguais ao *extremo inferior da classe*, que é zero (o intervalo é fechado à esquerda). Não estão incluídos na classe 0 |— 5, ou [0 — 5), os valores iguais ao *extremo superior da classe*, que é 5 (o intervalo é aberto à direita). A indicação de intervalo fechado é dada pelo traço vertical do símbolo |—, ou pelo colchete. A indicação de intervalo aberto é dada pela ausência do traço vertical no símbolo |—, ou por parênteses.

Terminada a organização das classes, é preciso determinar o *ponto médio* de cada classe. O ponto médio é a média aritmética dos dois extremos de classe. Então para a classe 0 |— 5 o ponto médio é

$$\frac{0 + 5}{2} = 2,5$$

e para a classe 5 |— 10 o ponto médio é

$$\frac{5 + 10}{2} = 7,5$$

Calculados os pontos médios é preciso estabelecer as *freqüências de classe*, isto é, o número de pessoas que caem em cada classe. Para isso basta contar quantas pessoas têm de 0 a 4 anos, quantas têm de 5 a 9 anos, quantas têm de 10 a 14 anos etc. Obtidas as freqüências de classe, pode ser construída a tabela de distribuição de freqüências.

Como exemplo da maneira de apresentar uma tabela de distribuição de freqüências, observe a tabela 4.3. Os dados apresentados nessa tabela são bastante informativos. É fácil ver, por exemplo, que a maioria dos alcoólatras crônicos da amostra começou a ingerir bebidas alcoólicas entre 15 e 25 anos. Esta observação se explica pelo fato de que as freqüências maiores (34 e 28) correspondem às classes 15 |— 20 e 20 |— 25, respectivamente. A tabela 4.3 também mostra que o hábito de ingerir bebidas alcoólicas pode começar muito cedo (dois alcoólatras iniciaram o hábito antes dos 10 anos), mas também pode iniciar em idade madura (quatro alcoólatras iniciaram o hábito depois dos 40 anos).

Tabela 4.3

Alcoólatras crônicos segundo a idade que tinham quando iniciaram o hábito de ingerir bebidas alcoólicas

Classe	Ponto médio	Frequência
5 — 10	7,5	2
10 — 15	12,5	9
15 — 20	17,5	34
20 — 25	22,5	28
25 — 30	27,5	12
30 — 35	32,5	9
35 — 40	37,5	2
40 — 45	42,5	4

FONTE: DANTAS (1979).

Finalmente, é preciso considerar que nem sempre existe interesse em apresentar todas as classes. Em alguns casos a primeira classe pode incluir todos os elementos menores do que determinado valor. Diz-se então que o extremo inferior da primeira classe não está definido. Assim, pode ser construída uma tabela de distribuição de frequências das pessoas conforme a estatura, com as seguintes classes:

menos de 150 cm
150 — 160 cm
160 — 170 cm etc.

Outras vezes são agrupados, na última classe, todos os elementos que têm valor igual ou maior do que determinado valor. Diz-se então que o extremo superior da última classe não está definido. Assim, pode ser construída uma tabela de distribuição de frequências com as seguintes classes de idade:

menos de 10 anos
de 10 a 19 anos
de 20 a 29 anos
de 30 a 39 anos
de 40 a 49 anos
de 50 a 59 anos
de 60 a 69 anos
de 70 anos e mais

4.3 FREQUÊNCIA RELATIVA

As tabelas de distribuição de frequências podem apresentar, além das frequências, as *frequências relativas*. Para calcular a frequência relativa de determinada classe basta dividir a frequência dessa classe pelo total de elementos da amostra. Esse resultado, multiplicado por 100, dá a frequência relativa em percentagem. Então:

$$\text{Frequência relativa} = \frac{\text{Frequência}}{\text{Total}}$$

Para entender como são calculadas as frequências relativas, observe os dados apresentados na tabela 4.4.

Tabela 4.4

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980

Classes (em salários mínimos)	Frequência (em 1 000 pessoas)	Frequência relativa (em %)
Até 1/2 salário mínimo	7 479	16,2
Mais de 1/2 a 1 salário mínimo	10 380	22,4
Mais de 1 a 2 salários mínimos	13 040	28,2
Mais de 2 a 3 salários mínimos	5 458	11,8
Mais de 3 a 5 salários mínimos	4 787	10,4
Mais de 5 a 10 salários mínimos	3 119	6,7
Mais de 10 a 20 salários mínimos	1 359	2,9
Mais de 20 salários mínimos	646	1,4
Total	46 268	100,0

FONTE: IBGE (1983).

A frequência relativa da primeira classe ("até 1/2 salário mínimo") é 16,2%. Esse valor foi obtido dividindo 7 479, que é a frequência da primeira classe, por 46 268, que é o total. O resultado, multiplicado por 100, deu a frequência relativa em percentagem, isto é:

$$\frac{7\,479}{46\,268} \cdot 100 = 16,2\%$$

É fácil interpretar frequência relativa. Assim, o resultado calculado mostra que 16,2% das pessoas com dez anos ou mais, que tinham algum rendimento mensal no Brasil em 1980, recebiam menos de 1/2 salário mínimo.

4.4 FREQUÊNCIA ACUMULADA

As tabelas de distribuição de frequências também podem apresentar *frequências acumuladas*. Para obter a frequência acumulada de determinada classe basta somar, à frequência dessa classe, as frequências das classes anteriores. Observe os dados apresentados na tabela 4.5.

Tabela 4.5

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980

Classes (em salários mínimos)	Frequência (em 1 000 pessoas)	Frequência acumulada (em 1 000 pessoas)
Até 1/2 salário mínimo	7 479	7 479
Mais de 1/2 a 1 salário mínimo	10 380	17 859
Mais de 1 a 2 salários mínimos	13 040	30 899
Mais de 2 a 3 salários mínimos	5 458	36 357
Mais de 3 a 5 salários mínimos	4 787	41 144
Mais de 5 a 10 salários mínimos	3 119	44 263
Mais de 10 a 20 salários mínimos	1 359	45 622
Mais de 20 salários mínimos	646	46 268

FONTE: IBGE (1983).

A frequência acumulada até a terceira classe é 30 899. Para obter esse valor foram somadas as frequências da primeira, segunda e terceira classes, isto é:

$$7\,479 + 10\,380 + 13\,040 = 30\,899$$

Da mesma forma, a frequência acumulada até a quinta classe é:

$$7\,479 + 10\,380 + 13\,040 + 5\,458 + 4\,787 = 41\,144$$

Convém observar, neste ponto, que a frequência acumulada da primeira classe é igual à frequência dessa classe. Não existem classes anteriores

à primeira. A frequência acumulada da última classe é igual ao total porque, para obter a frequência acumulada da última classe, são somadas as frequências de todas as outras.

4.5 FREQUÊNCIA RELATIVA ACUMULADA

Também podem ser apresentadas, nas tabelas de distribuição de frequências, as *frequências relativas acumuladas*. Para obter a frequência relativa acumulada até determinada classe basta somar, à frequência relativa dessa classe, as frequências relativas das classes anteriores. Observe os dados apresentados na tabela 4.6.

Tabela 4.6

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980

Classes (em salários mínimos)	Frequência (em 1 000 pessoas)	Frequência relativa	Frequência relativa acumulada
Até 1/2 salário mínimo	7 479	16,2	16,2
Mais de 1/2 a 1 salário mínimo	10 380	22,4	38,6
Mais de 1 a 2 salários mínimos	13 040	28,2	66,8
Mais de 2 a 3 salários mínimos	5 458	11,8	78,6
Mais de 3 a 5 salários mínimos	4 787	10,4	89,0
Mais de 5 a 10 salários mínimos	3 119	6,7	95,7
Mais de 10 a 20 salários mínimos	1 359	2,9	98,6
Mais de 20 salários mínimos	646	1,4	100,0

FONTE: IBGE (1983).

A frequência relativa acumulada até a terceira classe é 66,8%. Para obter este valor foram somadas as frequências relativas da primeira classe, da segunda e da terceira, isto é:

$$16,2 + 22,4 + 28,2 = 66,8$$

É fácil interpretar frequência relativa acumulada. Assim, por exemplo, para a classe "mais de 1 a 2 salários mínimos" a frequência relativa acumulada é 66,8%. Isto significa que 66,8% das pessoas com dez anos ou mais, que tinham algum rendimento no Brasil em 1980, tinham rendimento mensal igual ou inferior a dois salários mínimos.

4.6 APRESENTAÇÃO GRÁFICA

Os dados organizados em classes podem ser apresentados em gráficos. São bastante utilizados o histograma e o polígono de freqüências.

4.6.1 Histograma

É mais fácil entender como se faz um *histograma* através de um exemplo. Observe os dados apresentados na tabela 4.3. Para fazer um histograma — quando os intervalos de classe são iguais — siga os seguintes passos:

- trace o sistema de eixos cartesianos;
- marque os extremos de classes no eixo das abscissas;
- no eixo das ordenadas escreva as freqüências ou as freqüências relativas;
- para cada classe da distribuição de freqüências, trace um retângulo com base igual ao intervalo de classe e altura igual à freqüência, ou à freqüência relativa da classe;
- coloque título no gráfico.

Os dados da tabela 4.3 estão apresentados em histograma na figura 4.1. Os traços feitos no interior do histograma, embora sejam muito comuns, são desnecessários. A figura 4.2 mostra o mesmo histograma apresentado na figura 4.1, mas sem os traços em seu interior. O desenho foi hachurado e foram feitas linhas auxiliares, o que facilita a leitura.

Figura 4.1

Alcoólatras crônicas segundo a idade em que iniciaram o hábito de ingerir bebidas alcoólicas

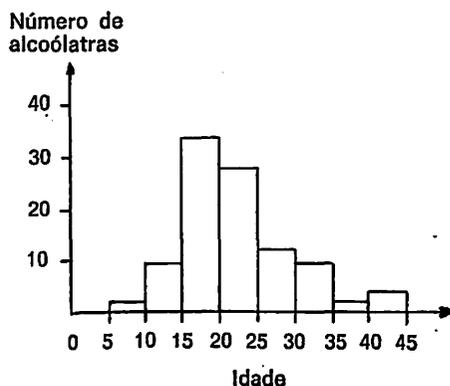
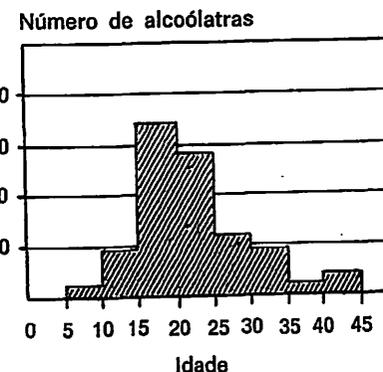


Figura 4.2

Alcoólatras crônicas segundo a idade em que iniciaram o hábito de ingerir bebidas alcoólicas



Para fazer um *histograma* — nos casos em que os intervalos de classe são diferentes — siga os seguintes passos:

- calcule a *densidade de freqüência relativa* para cada classe, que é definida como o quociente entre a freqüência relativa e o intervalo de classe;
- trace um sistema de eixos cartesianos;
- marque os extremos de classe no eixo das abscissas;
- no eixo das ordenadas marque as densidades de freqüência relativa de classe;
- para cada classe da distribuição de freqüências trace um retângulo com base igual ao intervalo de classe e altura igual à densidade de freqüência relativa de classe;
- coloque título no gráfico.

Para melhor entender como se faz um histograma — nos casos em que os intervalos de classe são diferentes —, observe os dados apresentados na tabela 4.7. A densidade de freqüência relativa da primeira classe, por exemplo, foi obtida dividindo a freqüência relativa dessa classe pelo intervalo de classe. Então, a densidade de freqüência relativa da primeira classe é:

$$\frac{16,2}{0,5} = 32,4$$

É importante observar que estão incluídas na última classe todas as pessoas que tinham rendimentos mensais de mais de 20 salários mínimos. Então, o extremo superior da última classe não está definido. Para fazer o cálculo fixou-se o extremo superior da última classe em 60 salários mínimos. O valor fixado não tem, na realidade, muita importância em termos dos resultados obtidos. Isto porque a porcentagem de pessoas, com rendimentos mensais iguais ou superiores a 20 salários mínimos, é muito pequena. Então, para a última classe, a densidade de frequência relativa foi calculada como segue:

$$\frac{1,4}{40} \approx 0,04$$

Tabela 4.7

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980

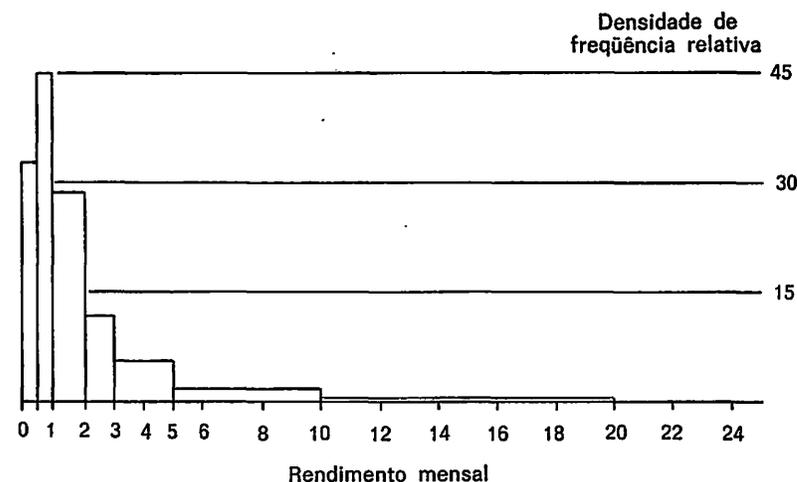
Classes (em salários mínimos)	Frequência (em 1 000 pessoas)	Frequência relativa	Densidade de frequência relativa
Até 1/2 salário mínimo	7 479	16,2	32,4
Mais de 1/2 a 1 salário mínimo	10 380	22,4	44,8
Mais de 1 a 2 salários mínimos	13 040	28,2	28,2
Mais de 2 a 3 salários mínimos	5 458	11,8	11,8
Mais de 3 a 5 salários mínimos	4 787	10,4	5,2
Mais de 5 a 10 salários mínimos	3 119	6,7	1,34
Mais de 10 a 20 salários mínimos	1 359	2,9	0,29
Mais de 20 salários mínimos	646	1,4	0,04

FONTE: IBGE (1983).

Os dados da tabela 4.7 estão apresentados graficamente na figura 4.3. Observe que a área de cada retângulo do histograma é proporcional à frequência relativa da classe correspondente.

Figura 4.3

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980



É importante lembrar, neste ponto, que para construir um histograma — nos casos em que os intervalos de classe são diferentes — também pode ser usada a *densidade de frequência* de classe. Para obter a densidade de frequência de uma classe, divide-se a frequência pelo intervalo de classe.

4.6.2 Polígono de frequências

É mais fácil mostrar como se faz *polígono de frequências* usando um exemplo. Observe os dados apresentados na tabela 4.3. Para fazer um polígono de frequências para esses dados, siga os seguintes passos:

- trace o sistema de eixos cartesianos;
- marque os pontos médios de classe no eixo das abscissas;
- no eixo das ordenadas, coloque as frequências;
- faça um ponto para representar cada classe. Esses pontos terão abscissa igual ao ponto médio de classe e ordenada igual à frequência de classe, como mostra o desenho à esquerda, no esquema dado em seguida.
- marque, no eixo das abscissas, um ponto que corresponda ao ponto médio de uma classe anterior à primeira; marque também um ponto que corresponda ao ponto médio de uma classe posterior à última, como mostra o desenho à direita do esquema dado em seguida;

- f) una todos esses pontos por segmentos de reta;
- g) coloque um título e está pronto o gráfico.

Os dados da tabela 4.3 estão mostrados em polígono de freqüências na figura 4.4.

Desenhos para a construção de um polígono de freqüências

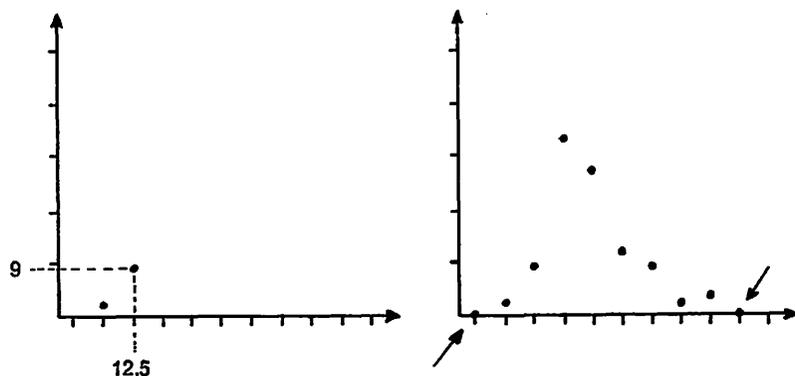
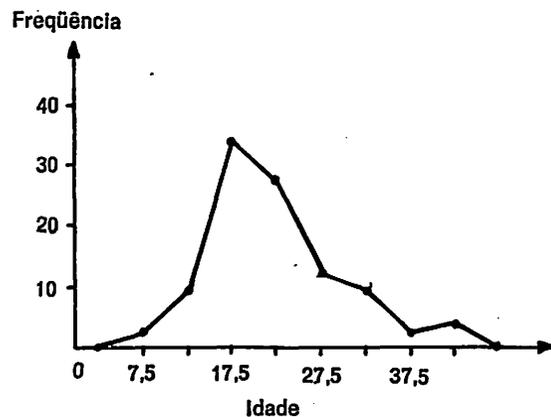


Figura 4.4

Alcoólatras crônicos segundo a idade em que iniciaram o hábito de ingerir bebidas alcoólicas



4.7 EXERCÍCIOS

- São dadas as notas de 40 alunos. Construa uma tabela de distribuição de freqüências considerando classes com os seguintes intervalos: 0 — 5, 5 — 7, 7 — 9, 9 — 10.

Notas de 40 alunos, em Matemática

7	3	4	9	5
8	5	3	8	7
9	3	1	9	9
5	7	7	10	3
4	8	8	8	7
8	8	9	3	8
9	9	7	8	1
6	10	7	7	9

- É dada uma tabela de distribuição de freqüências que apresenta pesos, em quilogramas, de recém-nascidos vivos. Calcule a freqüência relativa e a freqüência acumulada.

Distribuição de pesos, em quilogramas, de recém-nascidos vivos

Classe	Ponto médio	Frequência
0,5 — 1,0	0,75	1
1,0 — 1,5	1,25	3
1,5 — 2,0	1,75	22
2,0 — 2,5	2,25	115
2,5 — 3,0	2,75	263
3,0 — 3,5	3,25	287
3,5 — 4,0	3,75	99
4,0 — 4,5	4,25	32

FONTE: UNICAMP (1979).

- Imagine que você vai fazer uma tabela para apresentar o número de filhos tidos por mulher com idade entre 15 e 50 anos. Dê as classes que você usaria para construir uma tabela de distribuição de freqüências.
- Reorganize os dados da tabela apresentada em seguida, considerando as seguintes classes: nenhum dia com atraso, 1 ou 2, 3 ou 4, 5 ou mais.

Distribuição dos empregados de uma firma segundo o número de dias que chegaram com atraso ao trabalho, no mês de fevereiro de 1986

Número de dias	Número de empregados
0	32
1	12
2	7
3	5
4	2
5	1
6	2
7	0
8	1

5. Faça um histograma para apresentar os dados de idade da população brasileira. Para isso, reorganize a tabela, usando intervalos de classe iguais (todos com 10 anos). Considere que o extremo superior da última classe é 80 anos.

População residente no Brasil, segundo os grupos de idade, de acordo com o censo demográfico de 1980

Grupos de idade	População
0 a 4 anos	16 423 700
5 a 9 anos	14 773 741
10 a 14 anos	14 263 322
15 a 19 anos	13 575 971
20 a 24 anos	11 513 220
25 a 29 anos	9 442 217
30 a 39 anos	14 039 109
40 a 49 anos	10 377 274
50 a 59 anos	7 250 094
60 a 69 anos	4 474 511
70 anos e mais	2 741 506

FONTE: IBGE (1984).

8. Jogue um dado 30 vezes. Depois faça uma tabela de distribuição de freqüências para apresentar os resultados.
9. Pergunte a seus colegas de classe durante quanto tempo eles estudaram para fazer determinada prova. Organize os dados em uma tabela de distribuição de freqüências. Depois pergunte ao professor quanto tempo seria necessário estudar para fazer uma boa prova. Compare.
10. Pergunte para pessoas de ambos os sexos quanto tempo elas despendem, em média, por semana, assistindo a programas de televisão. Organize os dados em duas tabelas de distribuição de freqüências. Depois faça dois histogramas sobrepostos (um para cada sexo) e compare.

6. Faça um histograma para apresentar os dados de idade da população brasileira. Considere os intervalos dados.
7. Faça um polígono de freqüências para apresentar os dados de peso de recém-nascidos vivos do exercício 2.

5

Medidas de Tendência Central

Os dados apresentados em tabelas e em gráficos fornecem informação sobre o assunto em estudo. No entanto, pode existir interesse em apresentar essa informação de maneira condensada. Assim, imagine que um aluno fez cinco provas em Matemática e obteve as notas apresentadas na tabela 5.1.

Tabela 5.1

Notas do aluno José da Silva, em Matemática.

Prova	Nota
1.ª prova	7,0
2.ª prova	3,0
3.ª prova	5,5
4.ª prova	6,5
5.ª prova	8,0

Pode haver interesse em apresentar a informação contida na tabela 5.1. mais resumidamente. Afinal, para estabelecer se o aluno foi ou não foi aprovado, basta observar a média das notas obtidas — não é preciso “ver” todas as notas. Então, o conjunto de valores apresentados na tabela 5.1 pode ser resumido em um só valor — a média aritmética. Esse valor é apenas parte da informação dada pelas cinco notas, mas, no caso, constitui informação suficiente para estabelecer se o aluno foi ou não aprovado.

De maneira geral, existe interesse em apresentar um resumo das informações disponíveis. A maneira mais simples de resumir a informação

contida em um conjunto de dados é estabelecer o valor de um ponto central em torno do qual os dados se distribuem. Tais valores constituem as *medidas de tendência central*, definidas neste capítulo. Serão estudadas aqui média aritmética, média ponderada, média geométrica, média harmônica, moda e mediana.

5.1 MÉDIA ARITMÉTICA

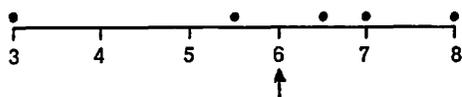
Para obter a *média aritmética* de um conjunto de dados basta dividir a soma dos valores de todos os dados pelo número deles. Então, a média aritmética das notas apresentadas na tabela 5.1 é:

$$\frac{7,0 + 3,0 + 5,5 + 6,5 + 8,0}{5} = 6,0$$

A média aritmética é uma medida de tendência central porque dá a abscissa de um ponto em torno do qual os dados se distribuem. Para entender esta idéia, imagine os valores da tabela 5.1 representados sobre um eixo, como mostra a figura 5.1. É fácil verificar, observando essa figura, que a média aritmética é a abscissa de um ponto em torno do qual os dados se distribuem.

Figura 5.1

Distribuição de dados em torno da média



A média aritmética é representada por \bar{x} (lê-se x - traço ou x - barra). A fórmula para o cálculo da média é:

$$\bar{x} = \frac{\sum x}{n}$$

A letra grega Σ (sigma) é usada como símbolo matemático para indicar que todos os valores observados de X devem ser somados. Essa letra é lida, em Matemática, como somatório. Então, quando se escreve Σx , lê-se "somatório de x ". No entanto, a maneira mais correta de indicar que todos os n valores assumidos pela variável X , isto é,

$$x_1, x_2, \dots, x_n$$

devem ser somados, é:

$$\sum_{i=1}^n x_i$$

que se lê "somatório de x índice i , i variando de 1 a n ". Com esta indicação, fica claro que existem n valores x_i . Então, a maneira mais correta de escrever a fórmula da média aritmética é:

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

5.1.1 Média aritmética de dados apresentados em tabelas de distribuição de freqüências

É mais fácil entender como se calcula a média aritmética de dados apresentados em tabelas de distribuição de freqüências através de um exemplo. Observe os dados fictícios apresentados na tabela 5.2.

Tabela 5.2

Distribuição dos alunos do curso colegial da Escola "João de Melo" segundo o número de advertências e/ou suspensões escolares

Número de advertências e/ou suspensões escolares	Número de alunos
0	211
1	40
2	30
3	12
4	4
5	2
6	1
Total	300

Para obter a média dos dados apresentados na tabela 5.2, siga os seguintes passos:

- some as frequências, isto é, calcule Σf_i como mostra a tabela 5.3. Esta soma dá o número de alunos do curso;
- multiplique cada valor observado pela respectiva frequência, isto é, calcule $x_i f_i$, como mostra a mesma tabela;
- some esses produtos, isto é, calcule $\Sigma x_i f_i$. Este valor é dado na tabela 5.3;
- aplique a fórmula:

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

Tabela 5.3

Cálculos intermediários para a obtenção da média dos dados da tabela 5.2

x_i	f_i	$x_i f_i$
0	211	0
1	40	40
2	30	60
3	12	36
4	4	16
5	2	10
6	1	6
Total	$\Sigma f_i = 300$	$\Sigma x_i f_i = 168$

A média aritmética dos dados apresentados na tabela 5.2 é:

$$\bar{x} = \frac{168}{300} = 0,56$$

ou seja, no curso colegial ocorreram, em média, 0,56 advertências e/ou suspensões escolares por aluno.

A média de dados agrupados em classes e apresentados em tabelas de distribuição de frequências é calculada pelo mesmo procedimento, com uma única diferença: os valores x_i são substituídos pelos pontos médios de classe. Observe os dados fictícios apresentados na tabela 5.4.

Tabela 5.4

Distribuição das estaturas, em centímetros, dos alunos do Colégio "João de Melo", em 1984

Classe	Ponto médio	Frequência
135 — 145	140	15
145 — 155	150	150
155 — 165	160	250
165 — 175	170	70
175 — 185	180	10
185 — 195	190	5

Para determinar a média de dados agrupados em classes, siga os seguintes passos:

- some as frequências, isto é, calcule Σf_i ;
- multiplique cada ponto médio de classe (x_i) pela frequência da classe, isto é, calcule $x_i f_i$;
- some esses produtos, isto é, calcule $\Sigma x_i f_i$;
- aplique a fórmula:

$$\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i}$$

Tabela 5.5

Cálculos intermediários para a obtenção da média dos dados da tabela 5.4

Classe	x_i	f_i	$x_i f_i$
135 — 145	140	15	2 100
145 — 155	150	150	22 500
155 — 165	160	250	40 000
165 — 175	170	70	11 900
175 — 185	180	10	1 800
185 — 195	190	5	950
Total		$\Sigma f_i = 500$	$\Sigma x_i f_i = 79 250$

A média aritmética dos dados apresentados na tabela 5.4 é:

$$\bar{x} = \frac{79 250}{500} = 158,5$$

ou seja, os alunos tinham, em média, 158,5 cm.

5.2 MÉDIA PONDERADA

Imagine que um aluno fez três provas em determinada disciplina em que a matéria é acumulativa, isto é:

- na primeira prova foram feitas perguntas sobre a matéria lecionada até a data dessa primeira prova;
- na segunda prova foram feitas perguntas sobre a matéria lecionada desde o início do curso até a data dessa segunda prova;
- na terceira prova foram feitas perguntas sobre toda a matéria lecionada.

É fácil argumentar que a primeira prova foi mais fácil do que a segunda e que esta foi mais fácil do que a terceira. Afinal, para a primeira prova havia menos matéria para estudar do que para a segunda e, para a segunda, havia menos matéria do que para a terceira. Então, a nota da primeira prova deve ter menos peso do que a da segunda e a nota da segunda prova deve ter menos peso do que a da terceira. Logo, é razoável estabelecer que:

- o peso da primeira prova é 1;
- o peso da segunda prova é 2;
- o peso da terceira prova é 3.

Para entender como se calcula a *média ponderada*, imagine que o aluno obteve as seguintes notas: 4, 7 e 6. A média ponderada dessas notas, usando os pesos 1, 2 e 3, respectivamente, é obtida seguindo os passos:

- some os pesos, isto é, calcule $\sum p_i$;
- multiplique cada nota pelo respectivo peso, isto é, calcule $x_i p_i$;
- some os produtos, isto é, calcule $\sum x_i p_i$;
- aplique a fórmula:

$$\bar{x} = \frac{\sum x_i p_i}{\sum p_i}$$

Tabela 5.6

Cálculos intermediários para obtenção da média ponderada

Nota (x_i)	Peso (p_i)	Produto ($x_i p_i$)
4	1	4
7	2	14
6	3	18
$\sum p_i = 6$		$\sum x_i p_i = 36$

A média ponderada do aluno é

$$\bar{x} = \frac{36}{6} = 6,0$$

5.3 MÉDIA GEOMÉTRICA

Considere dois números quaisquer, como 2 e 8. A *média geométrica* desses dois números é a raiz quadrada do produto deles, isto é:

$$G = \sqrt{2 \cdot 8} = \sqrt{16} = 4$$

Pode ser definida a média geométrica de mais de dois números. Assim, dados n valores da variável X ,

$$x_1, x_2, \dots, x_n$$

a média geométrica é, por definição:

$$G = \sqrt[n]{\prod_{i=1}^n x_i}$$

A letra grega Π (pi) é usada como símbolo matemático para indicar que todos os valores x_i , isto é,

$$x_1, x_2, \dots, x_n$$

devem ser multiplicados. Usa-se a letra maiúscula (não é o valor $\pi = 3,1416$). Em Matemática, a letra grega pi maiúscula é lida como produtório. Então quando se escreve:

$$\prod_{i=1}^n x_i$$

lê-se "produtório de x_i , i variando de 1 até n ". Isto significa que todos os valores x_i , i variando de 1 até n , devem ser multiplicados, isto é:

$$\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n$$

Para maior facilidade, pode-se escrever, simplesmente, Πx_i . Então, a média geométrica de n valores x_i é dada por:

$$G = \sqrt[n]{\prod x_i}$$

Para calcular G , pode-se aplicar logaritmo. Então:

$$\log G = \frac{\sum \log x_i}{n}$$

ou seja, o logaritmo da média geométrica é a média aritmética dos logaritmos dos dados. Segue-se que:

$$G = \text{antilog} \left(\frac{1}{n} \sum \log x_i \right)$$

Apenas como exemplo, considere os seguintes valores de X : 2, 3, 5 e 10. A média geométrica desses quatro valores é:

$$G = \sqrt[4]{2 \cdot 3 \cdot 5 \cdot 10}$$

Para obter o valor de G , pode-se aplicar logaritmo. Então:

$$\log G = \frac{1}{4} (\log 2 + \log 3 + \log 5 + \log 10)$$

Utilizando logaritmos decimais, obtém-se:

$$\log G = \frac{1}{4} (0,30103 + 0,47712 + 0,69897 + 1)$$

$$\log G = \frac{1}{4} \cdot 2,47712 = 0,61928$$

e, finalmente:

$$G \cong 4,16$$

5.4 MÉDIA HARMÔNICA

Imagine dois números, x_1 e x_2 . Para achar a média harmônica desses números:

a) inverta os números, isto é, calcule:

$$\frac{1}{x_1} \text{ e } \frac{1}{x_2}$$

b) determine a média aritmética desses inversos, isto é, calcule:

$$\frac{1}{n} \sum \frac{1}{x_i}$$

c) inverta essa média aritmética, isto é, calcule:

$$H = \frac{1}{\frac{1}{n} \sum \frac{1}{x_i}}$$

que é a média harmônica dos números dados.

Apenas como exemplo, considere os números 2 e 4. Para calcular a média harmônica:

a) inverta os números:

$$\frac{1}{2} \text{ e } \frac{1}{4}$$

b) determine a média aritmética:

$$\frac{\frac{1}{2} + \frac{1}{4}}{2} = \frac{\frac{2+1}{4}}{2} = \frac{3}{8}$$

c) inverta essa média:

$$H = \frac{8}{3} = 2,67,$$

obtendo, assim, a média harmônica de 2 e 4.

A média harmônica de n valores observados da variável X é o inverso da média aritmética dos inversos desses valores, isto é:

$$H = \frac{1}{\frac{1}{n} \sum \frac{1}{x_i}}$$

5.5 MODA

‘*Moda*’ é o valor que ocorre com maior frequência em um conjunto de dados. Para entender o conceito de moda, imagine que um aluno obteve as seguintes notas:

7; 8; 5; 7; 7; 9

A moda desse conjunto de notas é 7 porque, enquanto os outros números aparecem uma única vez, o número 7 ocorre três vezes. Observe agora os dados fictícios apresentados na tabela 5.7.

Tabela 5.7

Distribuição das classes da Escola "Santa Cecília" segundo o número de alunos reprovados por classe

Número de alunos reprovados por classe	Número de classes
0	1
1	5
2	8
3	13
4	7
5	0
6	2
Total	36

Conforme mostra a tabela 5.7, existiam 36 classes na escola. O número de alunos reprovados por classe variou entre zero (em uma classe não houve reprovações) e 6 (em duas classes houve seis reprovações). No entanto, ocorreram, com maior frequência, classes com três reprovados. Então, neste exemplo, a moda de alunos reprovados por classe é 3.

5.5.1 A moda de dados organizados em classes

Os dados fictícios apresentados na tabela 5.8 ajudam a entender como se determina a moda de dados organizados em classes, quando os intervalos de classe são iguais.

Tabela 5.8

Distribuição dos empregados da Empresa Santos & Santos, segundo a idade

Classe de idade	Frequência
10 — 15	2
15 — 20	7
20 — 25	15
25 — 30	27
30 — 35	20
35 — 40	17
40 — 45	12
45 e mais	9

A simples observação da tabela 5.8 mostra que, na firma em estudo, a classe de idade que ocorre com maior frequência é a classe 25 — 30. A frequência nessa classe é 27, maior que em todas as outras. Então essa é a classe modal (são mais comuns empregados com idade entre 25 — 30). Quando os dados estão organizados em classes, em geral basta informar a *classe modal*. Não é preciso informar o valor da moda, embora isso também possa ser feito.

A moda de dados agrupados em classes de intervalos iguais pode ser obtida através da fórmula:

$$m = L_h + \frac{a(f_h - f_{h-1})}{2f_h - (f_{h-1} + f_{h+1})}$$

Nessa fórmula, m representa a moda e:

- L_h = limite inferior da classe modal
- a = intervalo de classe
- f_h = frequência da classe modal
- f_{h-1} = frequência da classe anterior à classe modal
- f_{h+1} = frequência da classe posterior à classe modal

Para os dados apresentados na tabela 5.8, a classe modal é 25 — 30. Como:

$$\begin{aligned} L_h &= 25 \\ a &= 5 \\ f_h &= 27 \\ f_{h-1} &= 15 \\ f_{h+1} &= 20, \end{aligned}$$

segue-se que a moda é:

$$m = 25 + \frac{5(27 - 15)}{2 \cdot 27 - (15 + 20)} \approx 28,16$$

Quando os intervalos de classe são iguais, a classe modal é a classe que apresenta maior frequência. Se os intervalos de classe são diferentes, a classe modal será aquela que apresentar maior densidade de frequência e, conseqüentemente, maior densidade de frequência relativa. Para determinar a moda, nesses casos, é preciso calcular a densidade de frequência ou a densidade de frequência relativa de classe, como foi mostrado na seção 4.4. Como exemplo, observe a tabela 5.9. A classe "mais de 1 a 2 salários mínimos" apresenta a maior frequência (13 040) e a maior frequência relativa (28,2). No entanto, a classe "mais de 1/2 a 1 salário mínimo" apresenta maior densidade de frequência relativa (44,8). Então, esta é a classe modal.

Tabela 5.9

Pessoas residentes no Brasil, com dez anos ou mais, que tinham algum rendimento, segundo o valor do rendimento mensal, de acordo com o censo demográfico de 1980

Classe (em salários mínimos)	Frequência	Frequência relativa	Densidade de frequência relativa
Até 1/2 salário mínimo	7 479	16,2	32,4
Mais de 1/2 a 1 salário mínimo	10 380	22,4	44,8
Mais de 1 a 2 salários mínimos	13 040	28,2	28,2
Mais de 2 a 3 salários mínimos	5 458	11,8	11,8
Mais de 3 a 5 salários mínimos	4 787	10,4	5,2
Mais de 5 a 10 salários mínimos	3 119	6,7	1,34
Mais de 10 a 20 salários mínimos	1 359	2,9	0,29
Mais de 20 salários mínimos	646	1,4	0,04

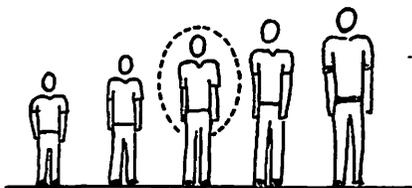
FONTE: IBGE (1983).

5.6 MEDIANA

Mediana é o valor que ocupa a posição central de um conjunto de dados ordenados. Para entender esse conceito observe a figura 5.2. Nessa figura estão representados cinco homens, organizados em fila por ordem de estatura. É fácil verificar que o terceiro homem ocupa a posição central, isto é, existem tantos homens antes dele (dois) quanto depois dele (dois). Como os homens estão organizados em fila por ordem crescente de estatura, diz-se que a estatura do terceiro homem é a mediana.

Figura 5.2

Distribuição de elementos em torno da mediana



Para determinar a mediana é preciso ordenar os dados. Se o número de dados é *ímpar*, a mediana é o valor que ocupa a posição central dos dados ordenados. Observe os dados fictícios apresentados na tabela 5.10. Para determinar a mediana é preciso organizar os dados em ordem crescente, como mostra a tabela 5.11. A mediana é o valor que ocupa o centro dos dados ordenados, isto é, no caso desse exemplo a mediana é 7,0.

Tabela 5.10

Notas em português dos alunos do 3.º ano colegial

Aluno	Nota
André	5,0
Carla	5,5
Eliana	8,5
Júlio	7,0
Pedro	8,0

Tabela 5.11

Dados organizados em ordem crescente, para a determinação da mediana

Ordem	Aluno	Nota
1.º	André	5,0
2.º	Carla	5,5
3.º	Júlio	7,0
4.º	Pedro	8,0
5.º	Eliana	8,5

Se o número de dados é *par*, a mediana é a média aritmética dos dois valores que ocupam a posição central dos dados ordenados. Observe os dados fictícios apresentados na tabela 5.12.

Tabela 5.12

Estaturas, em centímetros, dos alunos da 8.ª série A, da Escola "Santa Marcelina"

160	140	169	149	159
164	147	164	147	163
153	150	168	167	142
162	171	152	163	162

Para obter a mediana é preciso colocar os dados em ordem crescente. Na tabela 5.13, os dados estão ordenados em colunas. Como o número de dados é 20, a mediana é a média aritmética dos valores que estão em décimo e décimo primeiro lugar, porque existem nove números antes do décimo e nove números depois do décimo primeiro. Então, a mediana dos dados apresentados na tabela 5.13 é a média aritmética dos valores 160 e 162, isto é:

$$\frac{160 + 162}{2} = 161$$

Tabela 5.13

Dados ordenados para determinação da mediana

140	149	159	163	167
142	150	160	163	168
147	152	162	164	169
147	153	162	164	171

A mediana é uma *separatriz* porque separa o conjunto de dados em dois:

- o que antecede a mediana (dados iguais ou menores que a mediana);
- o que sucede a mediana (dados iguais ou maiores que a mediana).

Se um conjunto de dados tem n números, a mediana é o valor que ocupa a posição de ordem:

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

Então, se um conjunto de dados tem cinco elementos, a mediana é o elemento de ordem:

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

isto é, o terceiro elemento. Se um conjunto tem 20 elementos, como é o caso do exemplo apresentado na tabela 5.12, a mediana é o elemento de ordem,

$$\frac{20 + 1}{2} = 10,5$$

86 ou seja, é a média aritmética entre o décimo e o décimo primeiro valor.

5.6.1 A mediana de dados apresentados em tabelas de distribuição de freqüências

Para entender como se calcula a mediana de dados organizados em classes nas tabelas de distribuição de freqüências, observe a tabela 5.14.

Tabela 5.14

Distribuição das notas obtidas pelos alunos no exame vestibular de 1980, nas Faculdades Reunidas "Santa Mônica"

Classe	Freqüência	Freqüência acumulada
0 — 10	4	4
10 — 20	109	113
20 — 30	216	329
30 — 40	209	538
40 — 50	135	673
50 — 60	80	753
60 — 70	32	785
70 — 80	15	800
80 — 90	12	812
90 — 100	5	817

A tabela 5.14 mostra a distribuição de 817 notas de alunos que prestaram o vestibular. A mediana é o valor de ordem

$$\frac{817 + 1}{2} = 409$$

ou seja, no conjunto de notas, organizado em ordem crescente, a mediana será a nota que ocupar a posição 409.

Para determinar em que classe está a mediana é preciso observar as freqüências acumuladas. A primeira classe (0 |— 10) contém quatro notas, de ordem 1 a 4. A segunda classe (10 |— 20) contém as notas de ordem 5 a 113. A terceira classe (20 |— 30) contém as notas de ordem 114 a 329. Então, a mediana, que é de ordem 409, está na quarta classe (30 |— 40), já que essa classe, conforme mostra a freqüência acumulada, contém as notas de ordem 330 até 538.

O valor da mediana, representada por d , é obtido através da fórmula:

$$d = L_h + \frac{a}{f_h} \left(\frac{n}{2} - F_{h-1} \right)$$

onde:

- L_h = limite inferior da classe que contém a mediana
- a = intervalo de classe
- f_h = frequência da classe que contém a mediana
- n = número de dados
- F_{h-1} = frequência acumulada até a classe anterior à classe que contém a mediana

No exemplo apresentado:

- $L_h = 30$
- $a = 10$
- $f_h = 209$
- $n = 817$
- $F_{h-1} = 329$

Então:

$$d = 30 + \frac{10}{209} \left(\frac{817}{2} - 329 \right) = 33,8$$

5.7 QUARTIS E DECIS

Embora quartis e decis não sejam medidas de tendência central, serão definidos aqui devido à analogia com mediana. A mediana divide um conjunto de dados ordenados em duas partes iguais, isto é, a mediana é precedida e seguida por igual número de dados. Os *quartis* dividem o conjunto de dados em quatro partes iguais; os *decis* dividem o conjunto de dados em dez partes iguais. Observe os dados fictícios apresentados na tabela 5.15.

Tabela 5.15

Notas de Matemática dos alunos da 8.ª série, em maio de 1984

Aluno	Nota	Aluno	Nota
Júlio	7,5	André	2,5
Lúcia	5,5	Sílvia	9,0
Pedro	7,5	Fernando	9,0
Marina	9,0	Renata	7,5
Rogério	8,0	Luciana	5,5
Luís	6,5	Ana	6,0
Carlos	3,0	Renato	4,5
Paulo	7,5	João	7,0
Angélica	4,5	Alice	9,5
Leila	9,0	Alexandre	3,0

Para obter os *quartis* é preciso primeiro organizar as notas em ordem crescente, como mostra a tabela 5.16; depois, é preciso dividir o conjunto em quatro subconjuntos, com igual número de elementos. São 20 alunos. Então existem quatro subconjuntos de cinco alunos. O primeiro quartil é a média aritmética das notas de 5.ª e 6.ª ordem, isto é:

$$\frac{4,5 + 5,5}{2} = 5,0$$

O segundo quartil, que também é a mediana, é a média aritmética das notas de ordem 10 e 11, isto é:

$$\frac{7,0 + 7,5}{2} = 7,25$$

e o terceiro quartil é:

$$\frac{8,0 + 9,0}{2} = 8,5$$

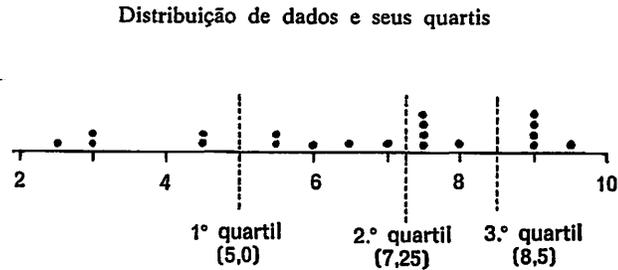
Tabela 5.16

Organização dos dados da tabela 5.15 para obtenção de quartis

Ordem	Aluno	Nota	Ordem	Aluno	Nota
1.º	André	2,5	11.º	Renata	7,5
2.º	Alexandre	3,0	12.º	Júlio	7,5
3.º	Carlos	3,0	13.º	Pedro	7,5
4.º	Angélica	4,5	14.º	Paulo	7,5
5.º	Renato	4,5	15.º	Rogério	8,0
6.º	Lúcia	5,5	16.º	Sílvia	9,0
7.º	Luciana	5,5	17.º	Marina	9,0
8.º	Ana	6,0	18.º	Leila	9,0
9.º	Luís	6,5	19.º	Fernando	9,0
10.º	João	7,0	20.º	Alice	9,5

Como mostra a figura 5.3, os quartis separam o conjunto de dados em quatro subconjuntos. Então os quartis são separatrizes.

Figura 5.3



Para obter os *decis* também é preciso organizar os dados em ordem crescente; depois divide-se o conjunto de dados em dez subconjuntos, com igual número de elementos. Os *decis* serão as separatrizes. Evidentemente, só tem sentido calcular *decis* quando o conjunto é formado por grande número de dados.

5.8 OBSERVAÇÕES

Quando se discutem medidas de tendência central, é importante lembrar que:

- a média aritmética é a medida de tendência central mais conhecida e mais utilizada, porque é de entendimento fácil e simples de calcular;
- a média ponderada é muito usada para notas escolares e de concursos, quando se dá maior peso (o que corresponde à maior importância) a determinada(s) prova(s);
- só tem sentido calcular a moda quando o conjunto de dados é bastante grande;
- a mediana, embora de fácil compreensão, é menos conhecida do que a média aritmética;
- as médias geométrica e harmônica são relativamente difíceis de calcular e, talvez devido a essa característica, são muito pouco usadas.

5.9 EXERCÍCIOS

- São dadas as idades dos membros da diretoria de um clube. Calcule a idade média.

Idade dos membros da diretoria do Clube "Amigos do Bairro"

54	61	55
60	55	58

- É dado o número de atendimentos, por dia, em um serviço de emergência, durante um mês. Calcule o número médio de atendimentos por dia.

Número de atendimentos, por dia, em um serviço de emergência, em novembro de 1985

Número de atendimentos	Frequência
0	2
1	3
2	3
3	5
4	10
5	6
6	1

- Calcule o tempo médio de duração de uma chamada telefônica interurbana com base nos dados fictícios apresentados na tabela dada em seguida.

Duração das chamadas telefônicas feitas em uma cidade, em setembro de 1975

Duração da chamada (minutos)	Frequência
0 — 2	100
2 — 6	50
6 — 10	30
10 — 15	20
15 — 20	5
20 — 30	5
30 — 40	1
40 — 60	1

4. São dadas as notas de cinco alunos em três provas. A primeira prova tem peso 2, a segunda tem peso 3 e a terceira tem peso 5. Calcule a média ponderada de cada aluno.

Notas de cinco alunos em Matemática

Aluno	1.ª prova	2.ª prova	3.ª prova
Ana	7	6	5
Cláudia	1	2	9
Marcos	5	5	5
Pedro	10	10	0
Sérgio	5	7	3

5. São dados os valores 4 e 9. Calcule a média geométrica.
 6. São dados os valores 12 e 16. Calcule a média harmônica.
 7. É dado o número de faltas não justificadas por empregado de uma firma no mês. Determine a moda.

Número de faltas não justificadas por empregado durante o mês de abril de 1985

Número de faltas	Frequência
0	120
1	13
2	5
3	1
4	0
5	1

8. Calcule a moda dos dados apresentados no exercício 2 do capítulo anterior.
 9. Pergunte a seus colegas de classe as notas que eles obtiveram em diferentes disciplinas. Organize os dados em uma tabela. Determine a nota mediana de cada disciplina. Discuta.
 10. Calcule a mediana dos dados apresentados no exercício 2 do capítulo anterior.

6 Medidas de Dispersão

As medidas de tendência central, vistas no capítulo anterior, dão o valor da abscissa de um ponto em torno do qual os dados se distribuem. Então, dada a informação de que um aluno foi aprovado com média 5, fica claro que as notas desse aluno se distribuíram em torno de 5. A média, no entanto, nada diz sobre o grau de dispersão das notas. Assim, por exemplo, se um aluno fez quatro provas durante o curso e foi aprovado com média 5, pode ter tirado qualquer das combinações de notas apresentadas a seguir:

5; 5; 5; 5
 4; 6; 4; 6
 0; 4; 6; 10
 0; 0; 10; 10,

além de várias outras combinações. É fácil verificar, observando as combinações de notas, que:

- a) as notas 5; 5; 5; 5 não têm dispersão;
 b) as notas 4; 6; 4; 6 têm dispersão pequena;
 c) as notas 0; 4; 6; 10 têm dispersão maior do que as anteriores;
 d) as notas 0; 0; 10; 10 têm a maior dispersão.

O exemplo dado mostra claramente que mesmo conjuntos com médias iguais podem apresentar dados com *dispersão* diferente. Muitas vezes existe interesse em medir o grau de dispersão dos dados de um conjunto. Neste capítulo serão descritas as seguintes *medidas de dispersão*: amplitude, variância e desvio-padrão. Também será descrito o coeficiente de variação, que é uma *medida de dispersão relativa*.

6.1 AMPLITUDE

Amplitude é a diferença entre o maior e o menor valor observado em um conjunto de dados. Observe que o maior valor apresentado na tabela 6.1 é 47 e o menor é 7. Então, a amplitude dos dados apresentados é:

$$47 - 7 = 40.$$

Tabela 6.1

Número de respostas corretas obtidas em uma prova simulada com 50 questões aplicada aos candidatos ao vestibular da Faculdade de Engenharia

30	9	19	31	29
22	43	25	12	31
45	39	22	47	42
27	25	29	25	29
15	13	36	23	30
10	7	15	12	25
8	16	14	18	10

É fácil interpretar a amplitude. No exemplo dado, a amplitude é 40. Então, o aluno que respondeu corretamente maior número de questões acertou 40 questões a mais do que o aluno que respondeu corretamente menor número de questões.

Quando se usam valores de amplitude é preciso não esquecer de que o cálculo dessa medida é feito com base apenas nos valores extremos. Os demais dados não são considerados para o cálculo da amplitude. Então, dois conjuntos de dados podem apresentar a mesma amplitude, mesmo que tenham dispersão diferente.

Para entender esta idéia, imagine dois grupos de pessoas. No primeiro grupo, as pessoas têm idades:

3; 4; 9; 11; 19; 20

e, no segundo grupo, as pessoas têm idades:

3; 10; 11; 11; 11; 20

A amplitude é a mesma nos dois casos, isto é:

$$20 - 3 = 17$$

mas a dispersão é, evidentemente, maior no primeiro grupo. Este exemplo mostra que a amplitude, embora fácil de calcular, não deve ser usada nor-

malmente como medida de dispersão. Isto porque a amplitude nem sempre distingue conjuntos com dispersão diferente.

6.2 VARIÂNCIA

Os dados distribuem-se em torno da média. Então, o grau de dispersão de um conjunto de dados pode ser medido pelos desvios dos valores observados em relação à média. Entende-se por *desvio em relação à média* a diferença entre o valor observado e a média do conjunto de dados. Assim, por exemplo, se a média de estatura dos jogadores de um time de futebol for $\bar{x} = 1,72$ m, o jogador que tiver estatura $x_i = 1,82$ m terá um desvio em relação à média de:

$$x_i - \bar{x} = 1,82 - 1,72 = 0,10 \text{ m}$$

Os desvios em relação à média medem a dispersão. É preciso considerar, no entanto, que cada dado tem um desvio em relação à média. Então, para julgar o grau de dispersão de todo o conjunto de dados, com base nos desvios, seria preciso observar todos os desvios. Não-se pode usar a soma dos desvios como medida de dispersão. A soma dos desvios é necessariamente igual a zero, como acontece no exemplo apresentado na tabela 6.2 e pode ser demonstrado matematicamente.

Tabela 6.2

Conjunto de dados e respectivos desvios em relação à média

Dados x_i	Desvios ($x_i - \bar{x}$)	Somas parciais
2	-5	-6
6	-1	
8	1	6
9	2	
10	3	
$\bar{x} = 7$	$\Sigma(x_i - \bar{x}) = 0$	

A soma dos desvios é sempre igual a zero porque os desvios, com sinais positivos e negativos, se compensam. É claro que o problema de achar um valor único para representar os desvios fica resolvido se, em lugar da soma dos desvios, for usada a *soma dos quadrados dos desvios*. Como os quadrados de números negativos são positivos, toda soma de quadrados é positiva ou, no mínimo, nula (só é nula quando todos os desvios são iguais a zero). O procedimento para obter a soma de quadrados de desvios é mostrado na tabela 6.3, através de um exemplo.

Tabela 6.3

Cálculo da soma de quadrados de desvios

Dados	Desvios	Quadrados dos desvios
x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
2	-5	25
6	-1	1
8	1	1
9	2	4
10	3	9
$\bar{x} = 7$	$\Sigma(x_i - \bar{x}) = 0$	$\Sigma(x_i - \bar{x})^2 = 40$

A soma de quadrados de desvios não é usada como medida de dispersão porque seu valor é diretamente proporcional ao número de dados. Para entender esta idéia, imagine dois grupos de pessoas. No primeiro grupo, as pessoas têm pesos:

60; 70; 80

e, no segundo grupo, as pessoas têm pesos:

60; 60; 70; 70; 80; 80.

Observe os valores desses dados e compare as somas de quadrados de desvios apresentadas na tabela 6.4. A soma dos quadrados dos desvios do segundo grupo é igual ao dobro da soma dos quadrados dos desvios do primeiro grupo, embora a dispersão nos dois grupos seja a mesma. Isto se explica: o segundo grupo é igual ao primeiro repetido duas vezes.

Tabela 6.4

Cálculo da soma de quadrados de desvios

Grupo I			Grupo II		
x	$(x - \bar{x})$	$(x - \bar{x})^2$	x	$(x - \bar{x})$	$(x - \bar{x})^2$
60	-10	100	60	-10	100
70	0	0	60	-10	100
80	10	100	70	0	0
			70	0	0
			80	10	100
			80	10	100
	0	200		0	400

Para medir a dispersão dos dados em torno da média usa-se a *variância*, que pode ser definida como a soma dos quadrados dos desvios, dividida pelo número de dados, isto é:

$$\frac{\Sigma(x_i - \bar{x})^2}{n}$$

Por razões que não serão discutidas aqui, também se define a variância como a soma dos quadrados dos desvios, dividida pelo número de graus de liberdade. O *número de graus de liberdade* é igual ao número de dados menos 1, isto é, $n - 1$.

Então, a variância, que é indicada por s^2 , pode ser definida pela fórmula:

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$$

Neste livro, a variância só será calculada usando como divisor o número de graus de liberdade, isto é, $n - 1$. Quando se usam máquinas de calcular que dão diretamente a variância, ou programas prontos para computadores, é preciso verificar se a variância é obtida dividindo a soma de quadrados por n , ou por $n - 1$. Os resultados são, evidentemente, diferentes.

Para mostrar como se calcula a variância são dados os valores 1, 2, 3, 4 e 5. Os cálculos intermediários para a obtenção da variância estão apresentados na tabela 6.5.

Tabela 6.5

Cálculos intermediários para a obtenção da variância

Dados	Desvios	Desvios ao quadrado
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	-2	4
2	-1	1
3	0	0
4	1	1
5	2	4
$\bar{x} = 3$	$\Sigma(x_i - \bar{x}) = 0$	$\Sigma(x_i - \bar{x})^2 = 10$

Aplicando a fórmula, vem:

$$s^2 = \frac{10}{4} = 2,5$$

Desenvolvendo algebricamente a fórmula de variância, obtém-se:

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1}$$

Embora esta fórmula pareça, à primeira vista, de aplicação difícil, ela é, na realidade, mais fácil do que a anterior. Apenas para conferir, foi feito o cálculo da variância dos valores 1, 2, 3, 4 e 5, já apresentada anteriormente, usando esta fórmula.

Tabela 6.6

Cálculos intermediários para a obtenção da variância

x_i	x_i^2
1	1
2	4
3	9
4	16
5	25
$\sum x_i = 15$	$\sum x_i^2 = 55$

Então:

$$s^2 = \frac{55 - \frac{(15)^2}{5}}{4} = 2,5$$

ou seja, o mesmo valor obtido anteriormente.

É importante deixar claro que variância mede dispersão. Para entender esta afirmativa, imagine que quatro alunos fizeram quatro provas e foram todos aprovados com média igual a 7. As notas desses alunos estão apresentadas na tabela 6.7.

Tabela 6.7

Notas de Matemática dos alunos do 3.º colegial

Aluno	Prova				Média
	1.ª	2.ª	3.ª	4.ª	
Antônio	7	7	7	7	7
Carlos	6	8	6	8	7
João	4	6	8	10	7
Pedro	0	8	10	10	7

Observe cuidadosamente as notas apresentadas na tabela 6.7. É fácil ver que:

- as notas de Antônio não têm dispersão;
- as notas de Carlos têm dispersão pequena;
- as notas de João têm dispersão maior do que as de Carlos;
- as notas de Pedro têm a maior dispersão.

Estas observações são confirmadas pelas variâncias dessas notas. As variâncias estão apresentadas na tabela 6.8.

Tabela 6.8

Médias e variâncias das notas apresentadas na tabela 6.7

Aluno	Média	Variância
Antônio	7	0
Carlos	7	1,33
João	7	6,67
Pedro	7	22,67

6.2.1 Variância de dados agrupados

Se os dados estão apresentados em uma tabela de distribuição de frequências, para obter a variância aplica-se a fórmula:

$$s^2 = \frac{\sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}}{n - 1}$$

onde x_i são os valores dos dados, y_i são as frequências e $n = \sum f_i$ é o número total de observações.

Os dados apresentados na tabela 6.8 serão utilizados para mostrar como se calcula a variância.

Tabela 6.8

Distribuição dos alunos da 1.ª série da Escola de 1.º Grau da Vila Santo Antônio segundo o número de faltas às aulas em maio de 1985

Número de faltas	Número de alunos
0	15
1	10
2	5
3	5
4	1
5	1
6	0
7	3

Antes de calcular a variância, será calculada a média. Os cálculos intermediários estão apresentados na tabela 6.9.

Tabela 6.9

Cálculos intermediários para a obtenção da média dos dados da tabela 6.8

x	f	xf
0	15	0
1	10	10
2	5	10
3	5	15
4	1	4
5	1	5
6	0	0
7	3	21
$\Sigma f = 40$		$\Sigma xf = 65$

A média dos dados é:

$$\bar{x} = \frac{65}{40} = 1,625$$

ou seja, o número médio de faltas por aluno foi 1,625. Para obter a variância foram feitos os cálculos apresentados na tabela 6.10.

Tabela 6.10

Cálculos intermediários para obtenção da variância

x	f	xf	x^2	x^2f
0	15	0	0	0
1	10	10	1	10
2	5	10	4	20
3	5	15	9	45
4	1	4	16	16
5	1	5	25	25
6	0	0	36	0
7	3	21	49	147
$\Sigma f = 40$		$\Sigma xf = 65$	$\Sigma x^2f = 263$	

Aplicando a fórmula, obtém-se:

$$s^2 = \frac{263 - \frac{(65)^2}{40}}{39} = \frac{157,375}{39} = 4,035$$

6.3 DESVIO-PADRÃO

Como medida de dispersão, a variância apresenta a desvantagem de ter unidade de medida igual ao quadrado da unidade de medida dos dados. Para entender essa idéia imagine que um aluno fez três provas, com 60 questões cada uma. Na primeira prova acertou 40 questões, na segunda acertou 45 e na terceira acertou 50. A média de acertos, nas três provas, é:

$$\bar{x} = \frac{40 + 45 + 50}{3} = 45$$

ou seja, o aluno acertou, em média, 45 questões. A variância é:

$$s^2 = \frac{(40 - 45)^2 + (45 - 45)^2 + (50 - 45)^2}{3 - 1} = \frac{50}{2} = 25$$

ou seja, 25 questões ao quadrado (porque os desvios são elevados ao quadrado). Evidentemente, "questões ao quadrado" não têm qualquer sentido prático. Então é preciso definir uma medida de dispersão que seja a raiz quadrada da variância.

Por definição, *desvio-padrão* é a raiz quadrada, com sinal positivo, da variância. O desvio-padrão é representado por s .

Para o exemplo apresentado tem-se que:

$$s = \sqrt{25} = 5$$

ou seja, o desvio-padrão é 5 questões. O desvio-padrão, como a variância, mede a dispersão dos dados, mas tem a vantagem de usar a mesma unidade de medida dos dados.

6.4 COEFICIENTE DE VARIAÇÃO

O coeficiente de variação é uma medida de *dispersão relativa* porque estabelece uma relação entre o desvio-padrão e a média. Para entender a idéia de dispersão relativa imagine dois grupos de pessoas. No primeiro grupo, as pessoas têm idades:

1; 3; 5

e no segundo grupo as pessoas têm idades:

53; 55; 57.

A média de idade no primeiro grupo é:

$$\frac{1 + 3 + 5}{3} = \frac{9}{3} = 3$$

e no segundo grupo é:

$$\frac{53 + 55 + 57}{3} = \frac{165}{3} = 55$$

A variância no primeiro grupo é:

$$\frac{(1 - 3)^2 + (3 - 3)^2 + (5 - 3)^2}{2} = 4$$

e, no segundo grupo, é:

$$\frac{(53 - 55)^2 + (55 - 55)^2 + (57 - 55)^2}{2} = 4$$

A variância do primeiro grupo é igual à variância do segundo grupo. Logo, a dispersão dos dados em torno da média é exatamente a mesma, nos dois grupos. No entanto, diferenças de dois anos são muito importantes no primeiro grupo, onde a idade média é 3 anos. Dois anos de diferença, nesse grupo, significam grandes mudanças físicas e de comportamento. Esta observação não é, porém, verdadeira para as pessoas do segundo grupo que têm, em média, 55 anos. Nesse grupo, dois anos a mais, ou a menos, não fazem grande diferença.

Essas observações refletem a idéia de dispersão relativa, ou seja, de dispersão em relação à média. Para medir dispersão relativa usa-se o coeficiente de variação. Por definição, coeficiente de variação (C.V.) é a razão entre o desvio-padrão e a média. Essa razão é multiplicada por 100. Então, o coeficiente de variação é dado em porcentagem. Logo:

$$C.V. = \frac{s}{\bar{x}} \cdot 100$$

Observe o valor dos coeficientes de variação para os dados de idade, apresentados como exemplo. Para o primeiro grupo calculou-se:

$$\bar{x} = 3 \text{ e } s^2 = 4$$

Para o segundo grupo, calculou-se:

$$\bar{x} = 55 \text{ e } s^2 = 4$$

Então, o coeficiente de variação, para o primeiro grupo, é:

$$C.V. = \frac{2}{3} \cdot 100 = 66,7\%$$

e para o segundo grupo é:

$$C.V. = \frac{2}{55} \cdot 100 = 3,6\%$$

Os valores dos coeficientes de variação confirmam a observação de que a dispersão dos dados, em relação à média, é muito importante no primeiro grupo (C.V. = 66,7%) e pouco importante no segundo grupo (C.V. = 3,6%).

6.5 EXERCÍCIOS

1. É dado o rendimento mensal de 20 pessoas. Calcule a amplitude.

Rendimento mensal, em salários mínimos, de 20 pessoas

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

2. Imagine os valores que podem ser obtidos quando se joga um dado. Calcule os desvios em relação à média e apresente os resultados numa tabela.
3. Dados os valores 8; 0; 5; 7, calcule a variância.
4. Dados os pesos de dez casais, calcule a variância do peso dos homens e a variância do peso das mulheres. Onde ocorre maior dispersão?

Peso do marido e peso da mulher, em quilogramas

Marido	Mulher
82	61
75	56
67	71
65	49
90	62
58	57
78	58
61	54
79	65
65	65

5. Dez pessoas apostam no número de caras que irão ocorrer quando se joga uma moeda quatro vezes. Calcule a média e a variância das apostas.

Distribuição das apostas no número de caras que irão ocorrer em quatro lançamentos de uma moeda

Aposta	Frequência
0	2
1	2
2	5
3	1
4	0

6. As notas dos 30 alunos de uma classe foram agrupadas em quatro grandes grupos. Calcule a variância.

Distribuição das notas finais dos alunos da 3.ª série

Classe	Ponto médio	Frequência
50 — 60	55	11
60 — 70	65	11
70 — 80	75	5
80 — 90	85	3

7. Calcule o desvio-padrão para os seguintes conjuntos de dados:

- a) 10, 10, 0, 0, 5;
 b) 4, 4, 4, 6, 8;
 c) 0, 2, 3, 4, 6.

8. São dados o peso e a estatura de quatro pessoas. Calcule os coeficientes de variação. Qual é a variável que tem maior dispersão relativa?

Peso e estatura de quatro pessoas

Peso (kg)	Estatura (cm)
60	1,60
75	1,70
70	1,75
75	1,65

9. Lance dois dados diferentes dez vezes consecutivas e anote os resultados. Calcule a média e a variância dos resultados obtidos para cada dado. Discuta.
10. De posse dos dados de idade das pessoas de sua família, calcule a média e a variância.

7

Correlação e Regressão Linear Simples

Muitas vezes o crescimento de uma variável é acompanhado pelo crescimento de outra variável. Isto acontece com idade e estatura de crianças. À medida que aumenta a idade, aumenta a estatura das crianças. Outras vezes o crescimento de uma variável é acompanhado pela diminuição de outra variável. Isto acontece com idade e expectativa de vida adicional. À medida que aumenta a idade, diminui o tempo de vida que se espera viver.

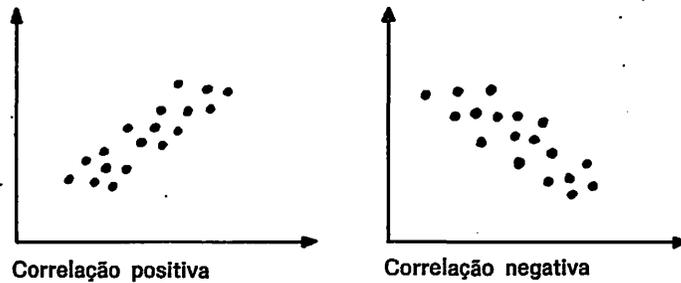
O comportamento conjunto de duas variáveis pode ser observado através de um gráfico, denominado *diagrama de dispersão*. Para entender como se desenha esse gráfico, imagine que são dados peso e estatura de pessoas. É fácil desenhar o diagrama de dispersão. Para isso:

- a) trace o sistema de eixos cartesianos;
- b) apresente uma das variáveis — por exemplo, a estatura — no eixo das abscissas e apresente a outra variável no eixo das ordenadas;
- c) para cada par de valores — no caso o peso e a estatura de uma pessoa — marque um ponto;
- d) coloque título no gráfico.

O diagrama de dispersão dá idéia do comportamento conjunto das duas variáveis. Se quando uma das variáveis cresce e a outra em média também cresce, o diagrama de dispersão tem o aspecto mostrado na figura 7.1, à esquerda. Nesses casos diz-se que entre as duas variáveis existe *correlação positiva*. Se quando uma das variáveis cresce e a outra em média diminui, o diagrama de dispersão tem o aspecto mostrado na figura 7.1, à direita. Nesses casos diz-se que entre as duas variáveis existe *correlação negativa*.

Figura 7.1

Diagramas de dispersão

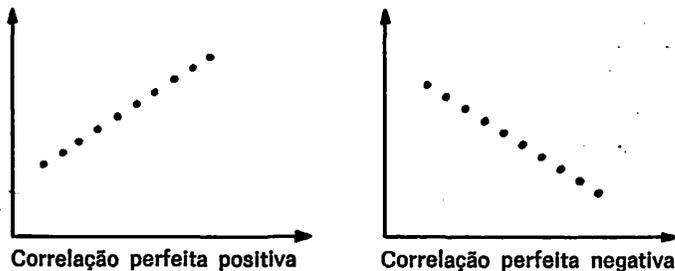


7.1 COEFICIENTE DE CORRELAÇÃO

A correlação entre duas variáveis pode ser perfeita. A correlação é *perfeita positiva* quando acréscimos constantes em uma das variáveis determinam acréscimos constantes na outra variável. Nesse caso, o diagrama de dispersão tem o aspecto apresentado na figura 7.2, à esquerda. A correlação é *perfeita negativa* quando acréscimos constantes em uma das variáveis determinam decréscimos constantes na outra variável. O diagrama de dispersão tem então o aspecto mostrado na figura 7.2, à direita. Note que, nos dois diagramas de dispersão, os pontos estão sobre retas. Essa configuração é típica da correlação perfeita.

Figura 7.2

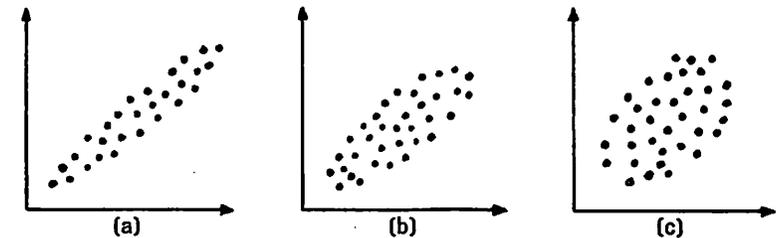
Aspecto gráfico de correlação perfeita



Mas a correlação raramente é perfeita. O grau de correlação entre duas variáveis varia bastante. Observe os diagramas de dispersão apresentados na figura 7.3. A correlação é positiva nos três gráficos, mas diminui de (a) para (b), e de (b) para (c).

Figura 7.3

Exemplos de correlação positiva



Para medir o grau de correlação entre duas variáveis usa-se o *coeficiente de correlação*. Esse coeficiente varia entre -1 e $+1$. Se a correlação for negativa, o sinal do coeficiente será negativo; se a correlação for positiva, o sinal do coeficiente será positivo. Quanto mais próximo de -1 estiver o valor do coeficiente de correlação, maior será a correlação negativa entre as variáveis. Da mesma forma, quanto mais próximo de $+1$ estiver o valor do coeficiente de correlação, maior será a correlação positiva entre as variáveis. O valor -1 indica correlação perfeita negativa e o valor $+1$ indica correlação perfeita positiva.

$r = 1$	Correlação perfeita positiva
$r = 0$	Correlação nula
$r = -1$	Correlação perfeita negativa

O coeficiente de correlação é indicado por r e calculado através da fórmula:

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n} \right] \left[\Sigma y^2 - \frac{(\Sigma y)^2}{n} \right]}}$$

Embora esta fórmula pareça complicada, à primeira vista, é fácil calcular o valor de r . Para entender como se calcula o valor de r , imagine o lançamento de um dado. Seja X a variável que indica os valores que podem ocorrer, quando se joga um dado. Então X pode assumir os valores:

1, 2, 3, 4, 5 e 6

Seja Y a variável que indica os valores que podem estar na face em contato com a mesa, quando se joga um dado. Então Y também pode assumir os valores:

1, 2, 3, 4, 5 e 6

Todo dado tem a mesma disposição de faces. Então, quando sai 1, o valor na face em contato com a mesa é 6; quando sai 2, o valor na face em contato com a mesa é 5, e assim por diante, como mostra a tabela 7.1. Estes valores também estão apresentados na figura 7.4.

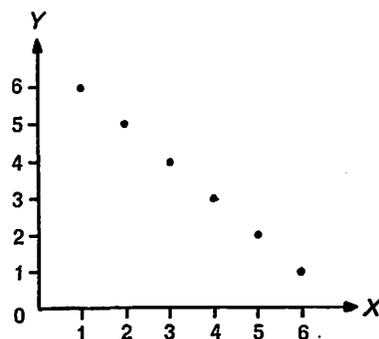
Tabela 7.1

Valores possíveis no jogo de um dado (X) e os respectivos valores da face em contato com a mesa (Y)

X	Y
1	6
2	5
3	4
4	3
5	2
6	1

Figura 7.4

Valores possíveis (X) de obter no jogo de um dado e valor (Y) da face em contato com a mesa



Os valores apresentados na tabela 7.1 e na figura 7.4 mostram que 110 acréscimos constantes em uma das variáveis determinam decréscimos cons-

tantes na outra variável. Então existe correlação perfeita negativa entre as duas variáveis. Se a correlação é perfeita negativa, o valor de r é igual a -1 . É fácil verificar este resultado. Os cálculos necessários para obter o valor de r estão apresentados na tabela 7.2.

Tabela 7.2

Cálculos intermediários para obtenção do valor de r

x	y	x^2	y^2	xy
1	6	1	36	6
2	5	4	25	10
3	4	9	16	12
4	3	16	9	12
5	2	25	4	10
6	1	36	1	6
$\Sigma x = 21$	$\Sigma y = 21$	$\Sigma x^2 = 91$	$\Sigma y^2 = 91$	$\Sigma xy = 56$

Substituindo os valores calculados na fórmula de r , obtém-se:

$$r = \frac{56 - \frac{21 \cdot 21}{6}}{\sqrt{\left[91 - \frac{(21)^2}{6}\right] \left[91 - \frac{(21)^2}{6}\right]}} = \frac{-17,5}{17,5} = -1$$

Para entender como se interpreta o valor de r , imagine que um professor quer saber se existe correlação entre notas em Português e notas em Matemática. Uma correlação positiva entre as notas dessas duas disciplinas implica a idéia de que os alunos que obtêm notas boas em Português geralmente obtêm notas boas em Matemática. Já os alunos que obtêm notas baixas em Português, em geral também obtêm notas baixas em Matemática. Uma correlação negativa entre as notas dessas duas disciplinas implica a idéia de que os alunos que obtêm notas boas em Português obtêm, em geral, notas baixas em Matemática e, vice-versa, os alunos que obtêm notas baixas em Português, em geral obtêm notas boas em Matemática.

Para discutir essas idéias com base em um exemplo numérico, observe os dados fictícios apresentados na tabela 7.3. Os cálculos intermediários para obtenção do coeficiente de correlação estão apresentados na tabela 7.4.

Tabela 7.3

Notas em Português e em Matemática obtidas pelos alunos do Curso Vestibular "A. de Abreu"

Nome do aluno	Português	Matemática
Antônio	8	5
Augusta	6	1
Beatriz	5	7
Carlos	3	2
Clóvis	8	6
Darci	4	5
Eduardo	5	3
Eunice	7	4
Fábio	10	9
Fernando	7	7
Francisco	3	3
Gláucia	6	8

Tabela 7.4

Cálculos intermediários para a obtenção do coeficiente de correlação

Português x	Matemática y	x ²	y ²	xy
8	5	64	25	40
6	1	36	1	6
5	7	25	49	35
3	2	9	4	6
8	6	64	36	48
4	5	16	25	20
5	3	25	9	15
7	4	49	16	28
10	9	100	81	90
7	7	49	49	49
3	3	9	9	9
6	8	36	64	48
72	60	482	368	394

Substituindo os valores na fórmula, tem-se que:

$$r = \frac{394 - \frac{72 \cdot 60}{12}}{\sqrt{\left[482 - \frac{(72)^2}{12}\right] \left[368 - \frac{(60)^2}{12}\right]}} = \frac{34}{\sqrt{50 \cdot 68}} = 0,583$$

Como o valor de r é 0,583, pode-se concluir que existe correlação positiva entre as notas obtidas pelos alunos, em Português e em Matemática.

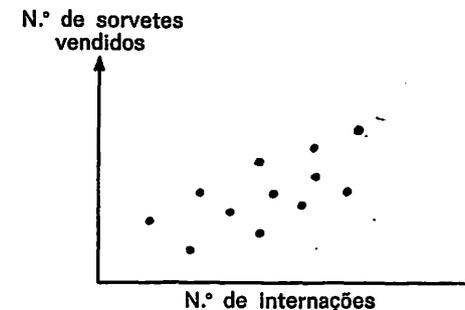
7.2 CUIDADOS NA INTERPRETAÇÃO DO COEFICIENTE DE CORRELAÇÃO

É importante deixar claro, neste ponto, que se duas variáveis têm correlação positiva, isto significa apenas que as variáveis crescem no mesmo sentido. No entanto, isso não implica a idéia de que o crescimento de uma das variáveis determina o crescimento da outra.

Para entender esta idéia, imagine que foram obtidos, em determinada cidade, dados sobre o número de sorvetes vendidos por mês e o número de internações hospitalares por desidratação, para cada um dos 12 meses de um ano. É bastante possível que os dados tenham, em gráfico, o aspecto mostrado na figura 7.5.

Figura 7.5

Diagrama de dispersão



A figura 7.5 mostra que existe correlação positiva entre número de sorvetes vendidos e número de internações por desidratação, mas não permite concluir que sorvete causa desidratação. No caso deste exemplo, o fato é conhecido: um aumento da temperatura determina tanto um aumento no consumo de sorvetes como maior incidência de casos de desidratação.

Da mesma forma, se duas variáveis têm correlação negativa, isto significa apenas que as variáveis crescem em sentidos opostos, isto é, à medida que X cresce, Y, em média, decresce. Isto não significa que os acréscimos em uma das variáveis sejam determinados pelos decréscimos na outra variável. Como exemplo, observe os dados apresentados na tabela 7.5.

Tabela 7.5

Consumo individual diário de proteínas de origem animal e coeficiente de natalidade, segundo o país

País	Consumo individual diário de proteínas de origem animal (g)	Coefficiente de natalidade (por mil)
Formosa	4,7	45,6
Malásia	7,5	39,7
Índia	8,7	33,0
Japão	9,7	27,0
Iugoslávia	11,2	25,9
Grécia	15,2	23,5
Itália	15,2	23,4
Bulgária	16,8	22,2
Alemanha	37,3	20,0
Irlanda	46,7	19,1
Dinamarca	56,1	18,3
Austrália	59,9	18,0
Estados Unidos	61,4	17,9
Suécia	62,6	15,0

FONTE: CASTRO (1961).

Os dados apresentados na tabela 7.5 sugerem que o aumento de consumo individual diário de proteínas de origem animal está associado à diminuição do coeficiente de natalidade. É fácil medir o grau de correlação

entre essas variáveis. Fazendo X indicar consumo individual diário de proteínas de origem animal e Y indicar o coeficiente de natalidade, obtém-se:

$$\begin{aligned} \Sigma x &= 413,0 & \Sigma y &= 348,6 \\ \Sigma x^2 &= 19.114,0 & \Sigma y^2 &= 9.706,02 \\ & & \Sigma xy &= 8.217,85 \end{aligned}$$

Então:

$$r = \frac{8.217,85 - \frac{413 \cdot 348,6}{14}}{\sqrt{\left[19.114 - \frac{(413)^2}{14}\right] \left[9.706,02 - \frac{(348,6)^2}{14}\right]}} = -0,775$$

Existe correlação negativa entre consumo individual diário de proteínas de origem animal e coeficiente de natalidade, mas isto não significa, necessariamente, que um aumento no consumo de proteínas de origem animal determine redução de fertilidade. Neste caso, a correlação negativa entre as duas variáveis pode ser explicada pelo fato de ambas serem afetadas por certas características da economia de cada país. O desenvolvimento econômico, associado à industrialização, urbanização e crescimento de renda *per capita*, induz à diminuição do coeficiente de natalidade e leva a maior consumo de proteínas de origem animal.

7.3 RETA DE REGRESSÃO

Podem haver interesse em expressar a variação de uma variável em função da variação de outra variável. Assim, por exemplo, pode haver interesse em expressar a variação do peso de indivíduos adultos em função da estatura. É bastante comum que a variação de uma variável, em função de outra, seja linear, isto é, que a variação de uma variável, em função de outra, possa ser descrita por uma reta. Observe os dados fictícios apresentados na tabela 7.6.

Tabela 7.6

Um exemplo de função linear com duas variáveis

X	Y
1	1
2	2
3	4
4	5
5	8

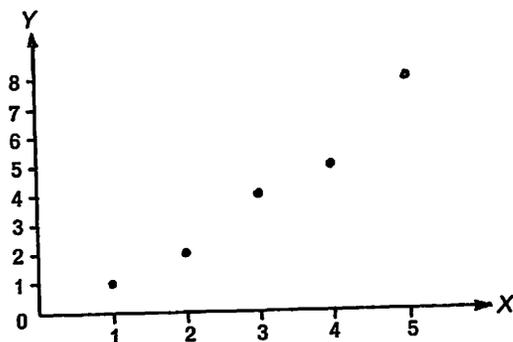
Os dados da tabela 7.6, apresentados em gráfico na figura 7.6, sugerem que Y pode ser considerado como uma função linear de X . Então existe uma reta de equação:

$$Y = a + bX$$

que expressa o valor de Y como função do valor de X . Então, Y é a variável dependente e X é a variável independente.

Figura 7.6

Um exemplo de função linear com duas variáveis



Os valores de a e b são calculados através das fórmulas:

$$b = \frac{\frac{\sum xy}{n} - \bar{x}\bar{y}}{\frac{\sum x^2}{n} - \bar{x}^2}$$

e

$$a = \bar{y} - b\bar{x}$$

onde \bar{y} e \bar{x} são as médias das variáveis Y e X , respectivamente. Para obter os valores de a e b , no caso do exemplo dado na tabela 7.6, foram feitos os cálculos apresentados na tabela 7.7.

Tabela 7.7

Cálculos intermediários para a obtenção da reta

x	y	x^2	y^2	xy
1	1	1	1	1
2	2	4	4	4
3	4	9	16	12
4	5	16	25	20
5	8	25	64	40
15	20	55	110	77

Com base nos dados apresentados na tabela 7.7, é fácil calcular:

$$b = \frac{77 - \frac{15 \cdot 20}{5}}{55 - \frac{(15)^2}{5}} = \frac{17}{10} = 1,7$$

e

$$a = \frac{20}{5} - 1,7 \cdot \frac{15}{5} = -1,1$$

A reta pode ser traçada no diagrama de dispersão. Basta dar dois valores quaisquer para X e calcular os valores de Y . Os valores calculados de Y não coincidem, necessariamente, com os valores observados de Y . Os valores calculados de Y são indicados por \hat{Y} . A equação da reta é então escrita como segue:

$$\hat{Y} = -1,1 + 1,7X$$

Para traçar a reta, faça $X = 0$ e $X = 5$. Para $X = 0$:

$$\hat{Y} = -1,1 + 1,7 \cdot 0 = -1,1$$

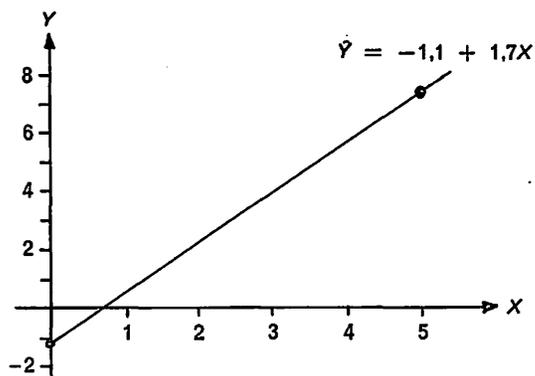
e, para $X = 5$:

$$\hat{Y} = -1,1 + 1,7 \cdot 5 = 7,4$$

Esses dois pontos (0; -1,1) e (5; 7,4) devem ser colocados no gráfico, como mostra a figura 7.7. A reta passa sobre esses pontos.

Figura 7.7

Apresentação gráfica da reta



A idéia de descrever a variação de uma variável em função de outra é antiga. No século passado, Galton, um matemático inglês, quis verificar se a estatura de um homem variava em função da estatura do pai. Para isso obteve dados de estatura de pais e filhos e os colocou em um gráfico. Galton verificou então que a estatura dos indivíduos pode ser analisada como função linear da estatura dos pais, mas que pais muito altos tendem a ter filhos menores do que eles. O matemático concluiu que estava havendo “regressão à mediocridade” e denominou a reta, que mostrava a estatura do indivíduo em função da estatura do pai, reta de regressão.

A expressão *reta de regressão* foi incorporada à estatística para indicar a reta que dá a variação de uma variável qualquer em função de outra. Também se fala em *regressão linear simples* para indicar o processo envolvido na obtenção de uma reta de regressão.

7.4 INTERPOLAÇÃO E EXTRAPOLAÇÃO

Imagine que foi obtida a equação da reta de regressão

$$\hat{Y} = a + bX$$

De posse dessa equação é possível calcular o valor provável de Y , isto é, uma estimativa de Y para dado valor de X que não foi observado. Como exemplo, observe os pares de valores de X e Y , apresentados na tabela 7.6. Não existe o valor $X = 3,5$, mas foi estabelecida a reta de regressão:

$$\hat{Y} = -1,1 + 1,7X$$

Então, embora o valor de $X = 3,5$ não tenha sido observado, de posse da equação da reta de regressão é possível estimar o valor de Y quando X assume o valor 3,5. Basta fazer, na reta de regressão, $X = 3,5$. Obtém-se:

$$\hat{Y} = -1,1 + 1,7 \cdot 3,5 = 4,85$$

O valor estimado de Y , quando $X = 3,5$ é $\hat{Y} = 4,85$. Nos dados apresentados na tabela 7.6 também não consta o valor $X = 10$. No entanto, é possível estimar o valor de Y para esse valor de X . Basta calcular:

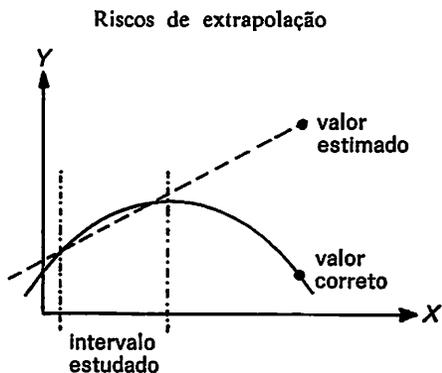
$$\hat{Y} = -1,1 + 1,7 \cdot 10 = 15,9$$

Quando se estimam valores de Y dentro do intervalo estudado, isto é, dentro do intervalo de valores observados de X , diz-se que foi feita uma *interpolação*. No exemplo, X varia entre 1 e 5, inclusive. Então, qualquer estimativa de Y , para valores de X dentro desse intervalo, é uma interpolação. Quando se estimam valores de Y fora do intervalo estudado, isto é, para valores de X menores, ou maiores do que os observados, diz-se que foi feita uma *extrapolação*. Uma vez que no exemplo apresentado X varia entre 1 e 5, inclusive, qualquer estimativa de Y para valores de X menores do que 1, ou maiores do que 5, é uma extrapolação.

Toda vez que se faz uma extrapolação é preciso ter muito cuidado. Primeiro porque, por razões teóricas que não serão discutidas aqui, os valores obtidos por extrapolação têm menor precisão (maior variância) do que os valores obtidos por interpolação. Depois porque o fato de Y ser função linear de X em determinado intervalo, não significa que Y continue sendo função linear de X fora desse intervalo.

A figura 7.8 explica este argumento. Imagine uma variável com o comportamento apresentado nessa figura, isto é, os valores crescem até atingir um ponto de máximo e depois decrescem. Uma variável com esse tipo de comportamento pode ser, eventualmente, observada apenas em determinado intervalo, como é mostrado na figura 7.8. Nesse caso pode-se ajustar uma reta aos dados, mas qualquer extrapolação levaria a um resultado desprovido de qualquer sentido.

Figura 7.8



Para entender esta idéia, imagine que alguém ajustasse uma reta de regressão aos dados apresentados na tabela 7.8.

Tabela 7.8

Temperatura média do ar em Curitiba, segundo o mês, em 1983

Mês	Temperatura média (°C)
Janeiro	21,2
Fevereiro	20,9
Março	18,7
Abril	17,6
Maio	15,9
Junho	12,4

FONTE: IBGE (1984).

É claro que qualquer “previsão” da temperatura média em Curitiba para meses posteriores aos observados, com base na reta de regressão ajustada aos dados da tabela 7.8, daria valores absurdos. Isto se explica: o fenômeno não é linear, mas cíclico.

Muitas previsões de grandezas econômicas e demográficas são feitas com base em regressão. Estas previsões, embora extremamente úteis, estão sujeitas a erro. Por isso devem ser julgadas com o devido cuidado. Os erros, nesses casos, são explicados pelos diversos fatores de difícil previsão que afetam essas grandezas. É interessante lembrar, neste ponto, que pou-

co antes de 1940 foi feita uma previsão da população que teriam os Estados Unidos em cada década, desde 1940 até 2100. Já apareceram desvios consideráveis das previsões feitas, por uma razão muito simples: o crescimento de uma população depende de condições sócio-econômicas bastante complicadas e dificilmente previsíveis a longo prazo, como guerras, restrições à imigração e crises econômicas.

7.5 EXERCÍCIOS

1. Calcule o coeficiente de correlação dos dados apresentados em seguida.

X	Y
1	1
2	3
3	1
4	3
5	1

2. Calcule o coeficiente de correlação dos dados apresentados em seguida.

X	Y
1	2
2	4
3	7
4	8
5	9

3. Calcule o coeficiente de correlação dos dados apresentados em seguida.

X	Y
1	9
2	8
3	7
4	4
5	2

4. São dados os valores médios de pesos de meninos com idades entre 7 e 12 anos, inclusive. Faça um diagrama de dispersão.

Pesos médios de crianças do sexo masculino, da cidade de São Paulo, segundo a idade, em 1975

Idade (anos)	Peso (kg)
7	22,2
8	24,4
9	27,1
10	29,5
11	32,4
12	35,4

FONTE: IBGE (1977).

5. Discuta o tipo de correlação (positiva, negativa ou nula) que, na sua opinião, devem ter os seguintes pares de variáveis:
- renda familiar e consumo de farinha de mandioca;
 - renda familiar e consumo de chocolate.
6. Ajuste uma reta de regressão aos dados apresentados no exercício 2. Trace a reta no gráfico.
7. Ajuste uma reta de regressão aos dados apresentados no exercício 3. Trace a reta no gráfico.
8. Se os filhos fossem, quando adultos, sistematicamente 5 cm mais altos do que os pais, qual seria a equação da reta de regressão que daria a estatura do filho em função da estatura do pai?
9. Com base nos dados apresentados no exercício 4, estime o peso de uma criança de 13 anos.
10. Com base nos dados apresentados na tabela 7.8 do texto, determine a reta de regressão da temperatura média mensal de Curitiba contra o tempo, em meses, ao longo do primeiro semestre de 1983. Para isso, faça X indicar o mês. Então, para indicar janeiro, use $X = 1$; para indicar fevereiro, use $X = 2$, e assim por diante. Depois estime, com base nessa reta, a temperatura média de Curitiba em dezembro de 1983. Discuta o resultado.

8 Probabilidade

Os jogos de azar, que se caracterizam por ações como girar uma roleta, lançar dados ou retirar cartas de baralho, têm duas características básicas: a incerteza e a regularidade. Assim, por exemplo, o resultado de um jogo de dados é incerto porque, toda vez que se joga um dado, pode ocorrer qualquer uma das faces. No entanto, o jogo, embora incerto, tem regularidade. Se forem feitos muitos lançamentos, espera-se que todas as faces ocorram igual número de vezes.

Essas características dos jogos de azar, percebidas há muito tempo, criaram a idéia de que seria possível achar uma "fórmula", ou um "método", que permitisse ao jogador ganhar sempre ou, pelo menos, ganhar na maioria das vezes. Isto não é possível, mas foi essa idéia que incentivou o estudo de tais jogos, o que levou à formulação da teoria de probabilidade, base da estatística moderna.

8.1 CONCEITO DE PROBABILIDADE

O estudo de *probabilidade* tem seus primórdios ligados aos jogos de azar. Essa origem permitiu conceituar probabilidade. Imagine que se deseja saber a probabilidade de ocorrer cara, quando se lança uma moeda. Se a moeda for honesta, espera-se que, numa série muito grande de lançamentos, ocorram caras e coroas igual número de vezes. Então, numa série muito grande de lançamentos ocorre cara metade das vezes. Logo, a probabilidade de ocorrer cara é $1/2$.

Então, se são possíveis n eventos mutuamente exclusivos e igualmente prováveis, se m desses eventos têm a característica A , a probabilidade de ocorrer um evento com a característica A é:

$$P(A) = \frac{m}{n}$$

É fácil aplicar esta fórmula. Para isso imagine o jogo de um dado. Quando se lança um dado, são possíveis seis eventos, isto é, podem ocorrer as faces 1, 2, 3, 4, 5 ou 6. Esses seis eventos são mutuamente exclusivos, porque duas ou mais faces não podem ocorrer ao mesmo tempo. Se o dado for honesto, esses seis eventos são igualmente prováveis.

Então, a probabilidade de ocorrer um número ímpar, quando se joga um dado, é:

$$\frac{3}{6} = \frac{1}{2}$$

porque existem seis eventos possíveis mutuamente exclusivos e igualmente prováveis, dos quais apenas três, isto é, as faces 1, 3 e 5, têm a característica "ser ímpar". Já a probabilidade de ocorrer a face 6, quando se lança um dado, é:

$$\frac{1}{6}$$

porque existem seis eventos mutuamente exclusivos e igualmente prováveis, dos quais apenas um (a face 6) tem a característica desejada.

A forma de calcular probabilidade, descrita aqui, é suficiente para as finalidades deste livro. A definição geral de probabilidade deve ser procurada em textos mais avançados.

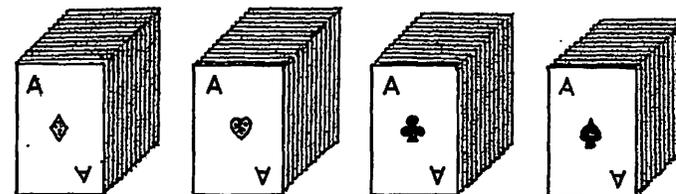
8.2 PROBABILIDADE CONDICIONAL

O conceito de probabilidade dado aqui permite responder a questões do tipo: "Qual é a probabilidade de ocorrer um ás de espadas, quando se retira ao acaso uma carta do baralho?" Para responder a esta pergunta é preciso saber que um baralho tem 52 cartas, conforme mostra a figura 8.1. Dessas 52 cartas, apenas uma tem a característica "ser ás de espadas". Então, a probabilidade de sair "ás de espadas", quando se retira uma carta ao acaso de um baralho, é:

$$\frac{1}{52}$$

Figura 8.1

Cartas de um baralho



Imagine agora que foi feita a mesma pergunta, isto é, perguntou-se: "Qual é a probabilidade de ter sido retirado um ás de espadas de um baralho?", mas se deu uma informação adicional: "Sabe-se que saiu carta de espadas". Ora, esta informação limita o número de eventos possíveis. Conforme mostra a figura 8.1, existem 13 cartas de espadas. Então, dada a informação de que saiu carta de espadas, a probabilidade de ter ocorrido um ás de espadas é:

$$\frac{1}{13}$$

e não mais 1/52, como anteriormente.

Este exemplo ilustra a idéia de *probabilidade condicional*, ou seja, a idéia de que a probabilidade de ocorrer um evento pode ser modificada quando se impõe determinada condição. No exemplo, a probabilidade de ocorrer às de espadas, que é 1/52, foi modificada para 1/13, quando se impôs a condição de ter ocorrido carta de espadas.

Então, a probabilidade condicional de ocorrer B, dado que ocorreu A, é a probabilidade de ocorrer B, sob a condição de ter ocorrido A, que é indicada como segue:

$$P(B|A)$$

8.2.1 Eventos Independentes

Imagine que uma moeda e um dado são lançados ao mesmo tempo. A probabilidade de ocorrer a face 6 no dado é 1/6, porque dos seis eventos possíveis, apenas um tem a característica "ser 6". No entanto, como o

dado e a moeda estão sendo lançados ao mesmo tempo, pode haver interesse em determinar a probabilidade de ocorrer face 6 no dado, sabendo-se que saiu cara na moeda. Esta é uma probabilidade condicional. Os eventos possíveis, quando se lança um dado e uma moeda, estão apresentados na tabela 8.1.

Tabela 8.1

Eventos possíveis quando se lança um dado e uma moeda ao mesmo tempo

Ocorrência no dado	Ocorrência na moeda	
	Cara	Coroa
1	1; cara	1; coroa
2	2; cara	2; coroa
3	3; cara	3; coroa
4	4; cara	4; coroa
5	5; cara	5; coroa
6	6; cara	6; coroa

É fácil verificar, observando a tabela 8.1, que a probabilidade de ocorrer a face 6 no dado é

$$\frac{1}{6}$$

quer se imponha, ou não, a condição de ter ocorrido cara na moeda. Este exemplo foi trazido aqui para mostrar que, em certos casos, a probabilidade de ocorrer um evento pode não se modificar, mesmo quando se impõe a condição de ter ocorrido outro evento. Nesses casos, os eventos são considerados *independentes*. Então, dois eventos, A e B, são independentes se:

$$P(B|A) = P(B)$$

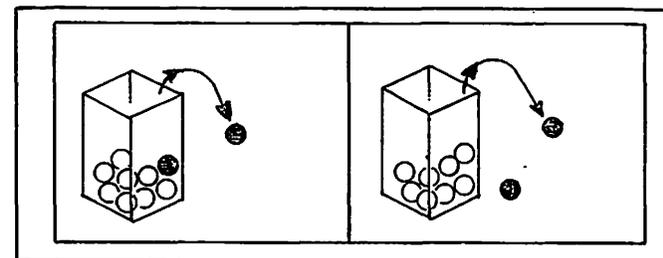
8.3 TEOREMA DO PRODUTO

Pode haver interesse em determinar a probabilidade de dois eventos ocorrerem ao mesmo tempo, ou um em seguida do outro. Assim, imagine que uma pessoa retira, ao acaso, uma bola de uma urna e, em seguida, sem que essa bola seja recolocada na urna, retira uma segunda bola. Imagine que essa urna contém duas bolas pretas e oito bolas brancas.

Qual é a probabilidade de terem sido retiradas as duas bolas pretas? Para responder a esta questão, observe a figura 8.2.

Figura 8.2

Duas bolas retiradas de uma urna



É fácil verificar que a probabilidade de uma pessoa retirar ao acaso uma bola preta de uma urna que contém duas bolas pretas e oito bolas brancas é:

$$\frac{2}{10}$$

Se sair a bola preta, e se essa bola não for recolocada na urna, a probabilidade de a pessoa retirar uma segunda bola preta é:

$$\frac{1}{9}$$

porque a urna passa a conter nove bolas (uma já foi retirada), das quais apenas uma é preta. Para determinar a probabilidade de ocorrer uma bola preta na primeira retirada e uma bola preta na segunda retirada, multipliquem-se as probabilidades, isto é, calcula-se:

$$\frac{2}{10} \cdot \frac{1}{9} = \frac{2}{90} = \frac{1}{45}$$

O *teorema do produto* pode ser enunciado como segue: A probabilidade de ocorrer um evento com a característica A e um evento com a característica B, isto é, a probabilidade de ocorrer um evento do conjunto $A \cap B$ é dada pela expressão:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

No exemplo, a probabilidade de ocorrer "bola preta na primeira retirada" e "bola preta na segunda retirada" é dada pelo produto de duas probabilidades: a probabilidade de sair bola preta na primeira retirada, multiplicada pela probabilidade de sair bola preta na segunda retirada, considerando-se que saiu bola preta na primeira.

É importante lembrar, neste ponto, que se A e B são eventos independentes:

$$P(B|A) = P(B)$$

Então, se A e B são eventos independentes, o teorema do produto fica como segue:

$$P(A \cap B) = P(A) \cdot P(B)$$

Como exemplo, imagine que são lançadas duas moedas. É claro que o fato de sair cara numa das moedas não influi sobre o fato de sair cara na outra moeda. Então, esses eventos são independentes. Conseqüentemente, a probabilidade de ocorrerem duas caras, quando se lançam duas moedas é:

$$\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

8.4 TEOREMA DA SOMA

Para entender o teorema da soma imagine que no jogo de dados o jogador ganha se sair 1 ou 6. Então, a probabilidade de o jogador ganhar é dada pela soma das probabilidades de ocorrer 1 ou 6, isto é:

$$\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

Imagine agora que no jogo de moedas o jogador ganha se sair pelo menos uma cara, em duas jogadas consecutivas. Então, a probabilidade de o jogador ganhar é dada pela probabilidade de sair cara na primeira jogada, somada à probabilidade de sair cara na segunda jogada, menos a probabilidade de ter saído cara nas duas jogadas, isto é:

$$\frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

Tabela 8.2

Eventos possíveis quando se lança uma moeda duas vezes consecutivas

Primeiro lançamento	Segundo lançamento	
	Cara	Coroa
Cara	Cara-Cara	Cara-Coroa
Coroa	Coroa-Cara	Coroa-Coroa

A tabela 8.2 mostra que os eventos com a característica desejada, isto é, "sair pelo menos uma cara", formam dois subconjuntos: a primeira linha — saiu cara na primeira jogada — e a primeira coluna — saiu cara na segunda jogada. No entanto, como o evento "cara-cara" foi contado duas vezes, precisa ser descontado uma vez.

O teorema da soma pode ser enunciado como segue: A probabilidade de ocorrer ou um evento com a característica A, ou um evento com a característica B, isto é, a probabilidade de ocorrer um evento do conjunto $A \cup B$ é:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Se $A \cap B$ é um conjunto vazio, o teorema da soma se reduz à expressão:

$$P(A \cup B) = P(A) + P(B)$$

Para tornar mais clara a aplicação do teorema da soma, imagine que se quer saber qual é a probabilidade de uma carta, retirada ao acaso de um baralho, ser um ás ou uma carta de espadas. De acordo com o teorema, a probabilidade de uma carta, retirada ao acaso de um baralho, ser um ás ou uma carta de espadas é:

$$\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

isto é, somam-se as probabilidades de se retirar um ás ($4/52$) e de se retirar uma carta de espadas ($13/52$), mas se subtrai a probabilidade de ocorrer um ás de espadas ($1/52$), porque essa probabilidade foi somada duas vezes, isto é, como ás e como carta de espadas.

8.5 FREQUÊNCIA RELATIVA

O conceito de probabilidade aplica-se facilmente nos casos de jogos de azar. Entretanto, a aplicação desse mesmo conceito fica difícil quando

se tenta responder a questões do tipo: Qual é a probabilidade de uma pessoa morrer antes de completar os 40 anos? Qual é a probabilidade de dois aviões se chocarem em pleno ar? Qual é a probabilidade de um bujão de gás explodir?

Todas essas questões são legítimas e estão associadas à teoria de probabilidades, mas não podem ser respondidas com base nos conceitos apresentados até aqui. É possível ampliar esses conceitos, mas primeiro observe os dados apresentados na tabela 8.3.

Tabela 8.3

Recém-nascidos segundo o sexo e a condição de vivo ou natimorto

Sexo	Condição		Frequência relativa (%)
	Vivo	Natimorto	
Masculino	1 513	37	2,4
Feminino	1 451	27	1,8
Total	2 964	64	2,1

FONTE: ARENA (1977).

Para obter a *freqüência relativa* de natimortos, no sexo masculino, dividiu-se o número de natimortos desse sexo (37) pelo total de nascidos do sexo masculino (1 513 vivos + 37 natimortos = 1 550). Obteve-se, dessa forma, o valor 0,024 ou 2,4%. A *freqüência relativa* é uma estimativa da probabilidade de natimorto do sexo masculino.

Analogamente, a *freqüência relativa* de natimortos de sexo feminino, que se obtém através do cálculo:

$$\frac{27}{1.478} = 0,018 \text{ ou } 1,8\%$$

pode ser uma estimativa da probabilidade de ocorrer um natimorto do sexo feminino.

Na área de saúde é comum usar o termo *risco*, como sinônimo de *probabilidade*. Assim, fala-se em *risco* de um nascituro apresentar defeito ou doença séria, em *risco* de uma pessoa contrair determinada doença, em *risco* de acidentes etc.

8.6 ALGUMAS TAXAS DEMOGRÁFICAS

As taxas demográficas são muito usadas nas áreas de saúde pública e de ciências sociais. Nesta seção serão apresentadas algumas dessas taxas, que podem ser interpretadas como *freqüência relativa*.

Entende-se por *taxa* ou *coeficiente*, em demografia, a razão entre o número de indivíduos que apresentam, ou apresentaram, determinada característica no decurso de certo período, e o total de indivíduos na população. Essa razão, multiplicada por 100, dá a taxa em porcentagem. No entanto, é mais comum multiplicar a razão obtida por 1 000, ou por 10 000. Dessa forma, as taxas ficam expressas por 1 000, ou por 10 000 indivíduos. Serão definidas aqui a taxa de mortalidade, a taxa de mortalidade infantil e a taxa de natalidade.

A *taxa de mortalidade* é a razão entre o número de óbitos, registrados em determinada região durante um ano, e a população da região, no início ou no meio desse ano. Essa razão é usualmente multiplicada por 1 000. Então, a taxa de mortalidade é, por definição:

$$\frac{N.^{\circ} \text{ de } \acute{o}bitos}{popula\c{c}\tilde{a}o} \cdot 1\,000$$

É importante considerar, quando se discute o valor da taxa de mortalidade, que os óbitos de não-residentes, ocorridos na região, também são incluídos no cálculo. Conseqüentemente, se a região tiver centros médicos especializados em doenças de alto risco, são registrados, na região, muitos óbitos de não-residentes. Isto ocorre, por exemplo, na cidade de São Paulo. As taxas de mortalidade para essa cidade são afetadas pelo fato de a cidade ser um grande centro de atendimento médico.

Também podem ser calculadas taxas específicas de mortalidade, isto é, taxas de mortalidade para determinada idade, devido a determinada causa ou para determinado grupo de pessoas. É interessante observar, por exemplo, que o número de pessoas que morreram com idade entre X e $X + 1$, dividido pelo número de pessoas que atingiram a idade X , dá a *probabilidade de morte na idade X*. O conhecimento dessa probabilidade é de grande interesse para as companhias de seguros de vida.

A *taxa de mortalidade infantil* — bastante utilizada como indicador da qualidade de vida de uma população — é dada pela razão entre o total de óbitos de menores de um ano de idade (excluídos os nascidos mortos) e o total de nascidos vivos, em determinado período de tempo (normalmente um ano). Essa razão é usualmente multiplicada por 1 000. Então, a taxa de mortalidade infantil é dada por:

$$\frac{n.^{\circ} \text{ de } \acute{o}bitos \text{ de menores de } 1 \text{ ano}}{n.^{\circ} \text{ de nascidos vivos}} \cdot 1\,000$$

Para verificar que a taxa de mortalidade infantil realmente dá idéia da qualidade de vida, observe os dados apresentados na tabela 8.4. É fácil verificar que as taxas de mortalidade infantil são bem menores nos países desenvolvidos.

Tabela 8.4

Taxa de mortalidade infantil, segundo o país, em 1982

País	Taxa de mortalidade infantil (por mil)
Bolívia	126
Índia	94
Brasil	73
México	53
Argentina	44
Reino Unido	11
EUA	11
França	10
Japão	7
Suécia	7

FONTE: Banco Mundial (1984).

Entende-se por *taxa de natalidade* a razão entre o número de nascidos vivos, em determinada região e em determinado período, e o total da população, nessa região e nesse período. Essa razão é multiplicada por 1 000. Então, a taxa de natalidade é dada por:

$$\frac{\text{n.º de nascidos vivos}}{\text{população}} \cdot 1\,000$$

A taxa de natalidade tende a diminuir com a urbanização, isto é, os casais residentes na área urbana têm, em geral, menos filhos do que os casais residentes nas áreas rurais. Ainda, as taxas de natalidade são menores em países e em regiões desenvolvidos.

8.7 EXERCÍCIOS

- Jogam-se dois dados. Qual é a probabilidade de a soma de pontos ser par?
- Suponha que a probabilidade de um casal ter um filho homem é 1/2. Nestas condições, qual é a probabilidade de um casal com cinco filhos ter os cinco filhos homens?
- Qual é a probabilidade de acertar, ao acaso, 13 pontos na Loteria Esportiva, sem fazer nenhuma aposta dupla?
- Uma urna branca contém duas bolas brancas e oito pretas. Uma urna preta contém duas bolas pretas e oito brancas. Se uma pessoa retirar ao acaso uma bola de cada urna, qual é a probabilidade de ter retirado pelo menos uma bola branca da urna branca ou da urna preta?
- Joga-se um dado duas vezes. Qual é a probabilidade de sair um número ímpar em pelo menos uma das jogadas?
- Calcule a frequência relativa de indivíduos dos diferentes grupos sanguíneos, com base nos dados da tabela apresentada em seguida.

Distribuição de indivíduos de acordo com o grupo sanguíneo

Grupo sanguíneo			
A	B	AB	O
292	94	17	417

FONTE: GARCIA (1977)

- Calcule a taxa de mortalidade infantil para as grandes regiões do Brasil, no ano de 1983. Os dados necessários para o cálculo estão na tabela apresentada em seguida. Discuta os resultados.

Nascidos vivos e número de óbitos registrados no Brasil, segundo a região, em 1983

Região	Nascidos vivos	Óbitos
Norte	97 592	9 204
Nordeste	663 472	59 875
Sudeste	1 306 371	62 093
Sul	473 217	16 593
Centro-Oeste ...	169 698	6 723

FONTE: IBGE (1984).

9. Sabe-se que uma moeda é honesta, isto é, a probabilidade de sair cara é igual a $1/2$. Suponha que a moeda foi jogada quatro vezes e ocorreram quatro caras. Numa próxima jogada é mais, ou menos, provável ocorrer cara?
10. Pergunte a um grupo de pessoas se elas gostam de poesia. Separe as respostas por sexo. Calcule as frequências relativas e discuta os resultados.

9

Números - Índices

Atualmente é bastante comum usar índices para caracterizar a situação econômica de um país, de um Estado ou de uma região. Os jornais e revistas publicam vários índices, como o índice geral de preços, o índice de custo de vida e o índice de cotação das ações na bolsa de valores. Embora esses índices sejam normalmente calculados e apresentados por instituições oficiais, é desejável que seus usuários tenham alguma idéia sobre a forma de calculá-los. Afinal, a discussão sobre os valores de tais índices exige algum conhecimento técnico como, por exemplo, a compreensão das fórmulas de cálculo.

No caso de índices de preços, é importante saber que para fazer o cálculo é preciso estabelecer:

- a amostra, ou seja, a relação de produtos cujos preços são utilizados no cálculo do índice;
- o período-base, ou seja, o período de tempo que será utilizado como base de comparação;
- o método de cálculo, ou seja, a fórmula usada para obtenção do índice.

9.1 ÍNDICE RELATIVO DE PREÇO

O *índice relativo de preço*, ou *preço relativo*, é dado pela relação entre o preço de um produto em determinado período e o preço do mesmo produto, no período definido como base. Esse resultado, multiplicado por 100, dá o índice relativo de preço em porcentagem. Como exemplo, imagine que um produto alcançou, em determinado período, preço igual a 56. No período-base, o preço desse produto era 35. Então, o índice relativo de preço é:

$$\frac{56}{35} \cdot 100 = 160$$

Pode ser dada uma fórmula para o cálculo do índice relativo de preço. Seja P_t o preço de um produto em determinado período de tempo e seja P_o o preço desse mesmo produto, no período definido como base. Então, o índice relativo de preço é, por definição:

$$\frac{P_t}{P_o} \cdot 100$$

Para entender como se interpreta o índice relativo de preço, observe os valores fictícios apresentados na tabela 9.1. Os índices foram calculados tomando 1983 como período-base. Então, o índice relativo de preço do produto A, em qualquer dos anos em análise, é dado pela relação entre o preço do produto naquele ano e o preço do produto em 1983, ano definido como base. O resultado, multiplicado por 100, dá o índice em porcentagem. É fácil verificar que os índices relativos de preço, apresentados na tabela 9.1 e na figura 9.1, mostram a evolução do preço do produto, entre 1982 e 1986.

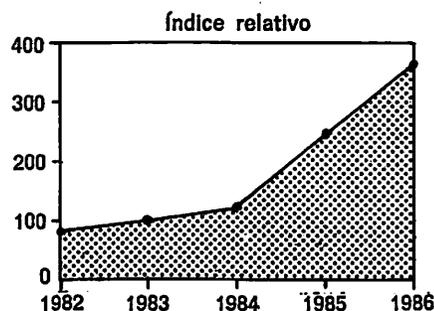
Tabela 9.1

Preços e índices relativos de preços do produto A no período 1982-85, tomando 1983 como base

Ano	Preço	Índice relativo
1982	28	80
1983	35	100
1984	42	120
1985	84	240
1986	126	360

Figura 9.1

Evolução do índice relativo de preço do produto A, no período 1982-86



O índice relativo de preço dá idéia da evolução do preço de um único produto. No entanto, na maioria das vezes, interessa estudar a evolução dos preços de um conjunto de produtos. Nesses casos é necessário usar os *índices gerais de preços*. Os índices gerais mais comumente utilizados são médias ponderadas dos preços relativos dos vários produtos. Quando se consideram os produtos consumidos pelas famílias e se pondera cada preço relativo de acordo com a importância do produto na despesa familiar, obtém-se um índice de custo de vida, que será analisado na próxima seção. Quando se consideram preços de todos os produtos, incluindo tanto bens de consumo como matérias-primas e bens de capital, obtém-se o índice geral de preços da economia. Na seção 9.3 será examinada a utilização do índice geral de preços na obtenção de valores reais ou deflacionados.

9.2 ÍNDICE DE CUSTO DE VIDA

Como a definição de índice de custo de vida é relativamente complicada, convém discutir alguns aspectos relacionados a esse índice, antes de defini-lo. O *índice de custo de vida* mostra o efeito das variações de preço sobre as despesas de uma "família típica". Para calcular esse índice é preciso determinar os itens que compõem a despesa de uma "família típica", bem como os preços que essas famílias pagam por esses itens. Os itens de despesa são os produtos que a "família típica" consome e os serviços que essa família utiliza. Os produtos de consumo referem-se a alimentação, vestuário, habitação, medicamentos, artigos para limpeza, mobiliário. Os serviços utilizados referem-se a transporte, água, luz, educação, saúde, lazer.

Para calcular o índice de custo de vida é preciso determinar também a proporção de cada item na despesa das famílias. Se, por exemplo, a "família típica" consome três produtos, A, B e C, e utiliza dois serviços, D e E, o total de despesa (100%) dessa família poderia ter a distribuição apresentada na tabela 9.2.

Os dados sobre as despesas das famílias são obtidos através de pesquisas. Para fazer esse tipo de pesquisa é preciso obter uma amostra da população de famílias cujo custo de vida se quer analisar e observar cada família da amostra, durante certo período de tempo, registrando todas as despesas. Obtém-se assim os gastos de cada família, com cada item e o total. A média dos gastos com cada item, dividida pela média dos totais, dá a proporção de gastos com cada item na amostra de famílias observadas.

Tabela 9.2

Distribuição da despesa de uma "família típica" segundo os itens de consumo e serviço

	Item	Proporção	Percentual
Consumo	A	0,3	30
	B	0,2	20
	C	0,1	10
Serviço	D	0,1	10
	E	0,3	30

É claro que o gasto de cada família com determinado item depende de fatores diversos, como número de pessoas que compõem a família, idade das pessoas, local da residência, costumes e modismos. No entanto, o fator que mais afeta a proporção de gastos de uma família com determinado item é o nível de renda. O gasto com alimentação, por exemplo, é proporcionalmente maior nas famílias mais pobres — não porque as famílias mais pobres comam mais —, mas simplesmente porque, para as famílias mais pobres, não sobra muito dinheiro para gastar com outros produtos de consumo. Já o gasto com automóveis próprios é maior nas famílias ricas, porque as famílias pobres simplesmente não têm automóvel. Feitas estas considerações, fica claro que é essencial informar, quando se constrói um índice de custo de vida, o nível de renda das famílias da amostra observada.

Para calcular o índice de custo de vida também é preciso obter os preços pagos pelas famílias da amostra, pelos diversos itens. Esses preços devem ser levantados onde as famílias fazem compras (supermercados, feiras, armazéns, quitandas, padarias, lojas etc.). Mais ainda, para que o índice mostre o verdadeiro crescimento do custo de vida é essencial obter os preços efetivamente pagos pelos consumidores. É um erro, por exemplo, utilizar preços tabelados pelo governo quando os consumidores são obrigados a pagar preços mais elevados, se quiserem comprar o produto com a qualidade usual.

Conhecidos os itens de despesa familiar, a proporção de gastos com cada item e os preços pagos pelos diversos itens, pode-se calcular o índice de custo de vida. Para entender como se faz esse cálculo, observe os valores fictícios apresentados na tabela 9.3.

Tabela 9.3

Preços de cinco itens, em janeiro de 1981 e janeiro de 1982

Mês	Produto			Serviço	
	A	B	C	D	E
Janeiro de 1981	15	10	25	10	20
Janeiro de 1982	24	30	45	20	44

É fácil obter o índice de custo de vida em janeiro de 1982, tomando janeiro de 1981 como base. Imagine que as proporções de A, B, C, D e E na despesa das famílias são os valores dados na tabela 9.2. Para obter o índice de custo de vida siga os passos:

- a) calcule o índice relativo de preço de cada produto e de cada serviço. No cálculo do índice de custo de vida, usam-se índices relativos de preço referidos à unidade — não em porcentagem. Então, para o produto A, o índice relativo de preço é dado pelo preço pago em janeiro de 1982, dividido pelo preço pago em janeiro de 1981, isto é:

$$\frac{24}{15} = 1,60$$

- b) multiplique o índice relativo de cada item pela proporção desse item na despesa familiar, e some:

$$\frac{24}{15} \cdot 0,3 + \frac{30}{10} \cdot 0,2 + \frac{45}{25} \cdot 0,1 + \frac{20}{10} \cdot 0,1 + \frac{44}{20} \cdot 0,3 =$$

$$= 1,6 \cdot 0,3 + 3,0 \cdot 0,2 + 1,8 \cdot 0,1 + 2,0 \cdot 0,1 + 2,2 \cdot 0,3 = 2,12$$

- c) multiplique o resultado por 100. Esse valor é o índice de custo de vida em porcentagem. Para o exemplo, tem-se:

$$2,12 \cdot 100 = 212,$$

ou seja, o índice de custo de vida em janeiro de 1982, tomando como base janeiro de 1981, é de 212%.

É interessante frisar, nesse ponto, que o valor de um índice no período-base é 1 ou 100%. No exemplo tomou-se janeiro de 1981 como base. Então, o índice de custo de vida nesse mês é 100%. O índice de custo de vida, calculado para janeiro de 1982, é 212%. Segue-se daí que o aumento no índice de custo de vida, de janeiro de 1981 para janeiro de 1982, foi:

$$212 - 100 = 112\%$$

Pode ser dada uma fórmula para o cálculo do índice de custo de vida. Considere que uma "família típica" tem despesa com n itens. Sejam x_i as proporções de cada um desses itens na despesa total e sejam P_{0i} e P_{1i}

os preços de cada item, no período-base e no período t , respectivamente. Então, o índice de custo de vida é, por definição:

$$\sum_{i=1}^n \frac{P_{ti}}{P_{oi}} \cdot x_i \cdot 100$$

ou seja, o índice de custo de vida definido aqui é a média ponderada de índices relativos de preço, usando como fatores de ponderação a proporção de cada item na despesa familiar.

9.3 CÁLCULO DE VALORES REAIS OU DEFLACIONADOS

Imagine que são dadas as estaturas de duas pessoas, mas um dos valores é dado em metros e o outro é dado em pés. Para que essas estaturas possam ser comparadas é preciso uniformizar a unidade de medida: ou as duas estaturas passam a ser dadas em metros, ou as duas estaturas passam a ser dadas em pés, ou ainda as duas estaturas passam a ser dadas em outra unidade, como centímetros, por exemplo.

Imagine agora que são dados dois valores monetários, pagos ou recebidos em diferentes épocas. À primeira vista parece que, se esses valores são dados em cruzeiros, ambos têm a mesma unidade de medida.¹ Entretanto, devido à inflação, um cruzeiro pago ou recebido em determinada época representa um valor de troca diferente de um cruzeiro pago ou recebido em outra época. Então, antes de comparar valores monetários de épocas diferentes, é preciso uniformizar a unidade de medida.

Para entender o procedimento, considere um aumento de preços da ordem de 100%, entre o período-base e o período t . Isto significa que o índice de preços passou de 100, no período-base, para 200 no período t , e o poder de compra da unidade monetária foi reduzido à metade.

Seja V_t um valor medido em cruzeiros do período t . Então V_t é o valor em cruzeiros correntes ou, simplesmente, V_t é o *valor corrente*. O mesmo valor medido em cruzeiros do período-base é o *valor real*.

O problema que se apresenta é estabelecer o valor real, isto é, o valor em cruzeiros do período-base, quando se conhece o valor corrente, isto é, o valor em cruzeiros no período t . No exemplo em discussão, os preços duplicaram-se. Então, o poder de compra do cruzeiro reduziu-se à metade. Conseqüentemente, o valor real (V_r) é igual à metade do valor corrente (V_t), ou seja:

$$V_r = \frac{V_t}{2}$$

Pode ser dada uma fórmula para o cálculo do valor real. Como o índice de preços passou de $I_0 = 100$ no período-base, para $I_t = 200$ no período t , segue-se que:

$$V_r = \frac{V_t}{I_t} \cdot 100$$

Esta fórmula dá o mesmo resultado obtido anteriormente. O índice geral de preços, usado para calcular valores reais, é denominado *deflator*. Por isso, o valor real também é conhecido como *valor deflacionado*.

Para melhor entender o que significam as expressões *valor corrente* e *valor real* ou *valor deflacionado*, observe os dados fictícios apresentados na tabela 9.4.

Tabela 9.4

Valor corrente da produção de arroz no município de Santa Cruz, em milhões de cruzeiros, e índice geral de preços, no período de 1980-83

Ano	Valor da produção de arroz	Índice geral de preços
1980	1 275	100
1981	2 430	150
1982	3 921	300
1983	5 735	500

É fácil verificar, observando a tabela 9.4, que o valor corrente da produção de arroz cresceu no período de 1980-83. No entanto, grande parte do crescimento do valor corrente é explicada pela *inflação*, isto é, pela perda de valor do cruzeiro. Logo, a contribuição da lavoura de arroz para o progresso econômico do município só pode ser analisada através da variação do valor real dessa produção.

Para obter o valor real, tomando 1980 como base, basta dividir o valor corrente pelo índice geral de preços e multiplicar por 100. Assim, por exemplo, o valor real da produção de arroz em 1982 é calculado como segue:

$$V_r = \frac{3\,921}{300} \cdot 100 = 1\,307$$

1. Devido à inflação acelerada, a moeda brasileira tem passado por várias mudanças de nome. Mas o método de obter valores deflacionados, exposto nesta seção, também se aplica a cruzados e cruzados novos.

É preciso lembrar, neste ponto, que o período-base do índice geral de preços, usado como deflator, é 1980. (Nesse ano, o valor do índice geral de preços é 100.) Então, no caso dos dados fictícios apresentados na tabela 9.4, o valor real está medido em cruzeiros de 1980. Isto significa que o valor corrente de 3 921 milhões de cruzeiros, em 1982, corresponde a 1 307 milhões de cruzeiros em 1980. Os valores reais para os outros anos são obtidos de maneira análoga e estão apresentados na tabela 9.5.

Tabela 9.5

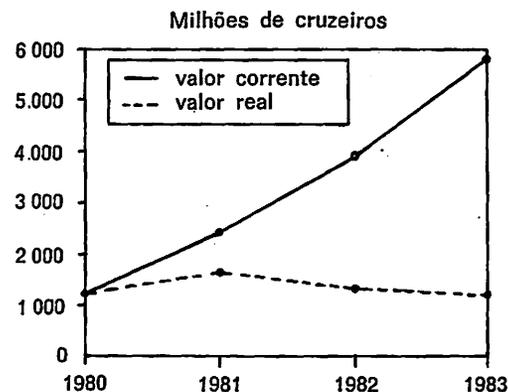
Valor real da produção de arroz no município de Santa Cruz, em milhões de cruzeiros, no período de 1980-83

Ano	Valor da produção
1980	1 275
1981	1 620
1982	1 307
1983	1 147

A figura 9.2 apresenta tanto os valores correntes como os valores reais da produção de arroz do exemplo em discussão. É fácil verificar, observando essa figura, que numa economia inflacionária é preciso analisar valores reais, nunca valores correntes.

Figura 9.2

Valor corrente e valor real da produção de arroz no município de Santa Cruz, em milhões de cruzeiros, no período de 1980-83



É preciso deixar claro, neste ponto, que no exemplo apresentado foram calculados valores reais em cruzeiros do período-base do deflator utilizado. (O valor real da produção foi dado em cruzeiros de 1980; nesse ano, o índice de preços era 100.) No entanto, também se pode obter valores reais relativos a um período k qualquer, desde que se disponha do valor do índice de preços usado como deflator para o período k .

Nesse caso, dado o valor corrente (V_t), o valor do deflator no mesmo período (I_t) e o valor do deflator em um período k (I_k), o valor real (V_r), em cruzeiros do período k , é dado por:

$$V_r = \frac{V_t}{I_t} \cdot I_k$$

A fórmula apresentada anteriormente é um caso particular desta última, com $I_k = 100$. Isto porque, naquele caso, o valor real obtido era medido em cruzeiros do período-base do índice.

Para ilustrar a aplicação dessa última fórmula serão utilizados os dados da tabela 9.4. Será calculado o valor real da produção de arroz do município de Santa Cruz, em cruzeiros de 1983. Nesse caso, $I_k = 500$. Então, para 1980 o valor real é:

$$\frac{1 275}{100} \cdot 500 = 6 375$$

Para 1981 o valor real é:

$$\frac{2 430}{150} \cdot 500 = 8 100$$

De maneira análoga podem ser obtidos os demais valores reais apresentados na tabela 9.6.

Tabela 9.6

Valor real da produção de arroz do município de Santa Cruz no período 1980-83, em milhões de cruzeiros de 1983

Ano	Valor da produção	Índice
1980	6 375	100,0
1981	8 100	127,1
1982	6 535	102,5
1983	5 735	90,0

Os principais índices de preços utilizados no cálculo de valores reais são o Índice Geral de Preços (IGP) da Fundação Getúlio Vargas e o Índice Nacional de Preços ao Consumidor Amplo (IPCA, que passou a ser denominado IPC a partir de março de 1986), do IBGE.

9.4 EXERCÍCIOS

1. O preço de um produto era 225 em janeiro de 1980 e passou para 333 em janeiro de 1981. Determine o índice relativo de preço desse produto em janeiro de 1981, com base em janeiro de 1980.
2. Se o preço de certo produto dobrou de valor, entre janeiro e dezembro de 1981, qual é o índice relativo de preço desse produto, em dezembro de 1981, com base em janeiro do mesmo ano?
3. A tabela a seguir mostra preços fictícios de quatro produtos em agosto e dezembro de 1986.

Preços dos produtos A, B, C e D
em agosto e dezembro de 1986

Mês	Produto			
	A	B	C	D
Agosto	50	10	87	20
Dezembro	60	9	87	26

Determine o índice relativo de preço de cada produto em dezembro, com base em agosto.

4. Para simplificar os cálculos, suponha que a despesa total das famílias é feita com apenas quatro produtos: A, B, C e D. Suponha, também, que a participação média desses produtos, na despesa das famílias, seja, respectivamente, 20%, 40%, 30% e 10%. Os preços dos quatro produtos, em agosto e dezembro de 1986, são os valores apresentados na tabela do exercício anterior. Determine o índice de custo de vida em dezembro, com base em agosto.
5. Suponha que a participação de um produto na despesa média das famílias seja de 5%. Suponha ainda que, em certo período, o preço desse produto cresceu 20%, enquanto os preços dos demais produtos permaneciam constantes. Qual foi o crescimento do custo de vida nesse período?

6. Entre dezembro de 1980 e dezembro de 1981 o preço de certo produto aumentou de 80 para 189 mil cruzeiros. Se o índice geral de preços, nesse mesmo período, cresceu de 100 para 210, qual é o preço real do produto em dezembro de 1981, medido em cruzeiros de dezembro de 1980? Qual foi o crescimento percentual do preço real do produto no período?
7. O preço corrente de um produto era 55 em agosto de 1976 e passou para 70 em agosto de 1977. Se o índice geral de preços aumentou 40% nesse período (isto é, passou de 100 para 140), determine a variação percentual do preço real de agosto de 1976 até agosto de 1977.
8. A série cronológica dos preços médios mensais do arroz em casca, em Cr\$/kg, recebidos pelos produtores do Sul do Brasil, de janeiro a dezembro de 1984, é a seguinte: 163, 195, 207, 217, 221, 239, 252, 271, 291, 345, 394 e 508. Calcule o valor do índice relativo de preço para cada mês, com base em janeiro de 1984.
9. A série cronológica dos valores do índice geral de preços da Fundação Getúlio Vargas, de janeiro a dezembro de 1984, com base em 1977, é a seguinte: 7 921,1, 8 892,1, 9 777,0, 10 651,1, 11 594,7, 12 667,2, 13 974,3, 15 458,7, 17 083,3, 19 232,2, 21 131,6 e 23 357,1. Utilizando os dados do exercício anterior, determine os preços reais do arroz em casca, recebidos pelos produtores do Sul, de janeiro a dezembro de 1984, em cruzeiros de 1977.
10. Refaça o exercício anterior, obtendo os preços reais em cruzeiros de dezembro de 1984.

Respostas aos Exercícios

CAPÍTULO 1

1. Idade, renda familiar, tempo disponível para estudo são variáveis cardinais. Sexo, religião, cor e atividades recreativas de interesse são variáveis nominais.
2. Variáveis nominais: nacionalidade, partido político que apóia.
Variáveis ordinais: status social, posição hierárquica no exército.
Variáveis cardinais: estatura, valor dos bens que possui.
3. Amostras de um elemento: A, B e C.
Amostras de dois elementos: A e B, A e C, B e C.
4. Podem ser colocadas, em uma urna, fichas com as letras A, B, C, D, E e F. Depois de misturadas, retira-se ao acaso uma amostra de três fichas.
5. Pode ser tomado um de cada dez elementos. Por exemplo, pode ser tomado o 7.º, o 17.º, o 27.º, e assim por diante, até o 97.º elemento.
6. Identificam-se por números, separadamente, os empregados de escritório, os empregados de oficina e os representantes. Depois se faz o sorteio de uma amostra aleatória de cada categoria. Poder-se-ia, por exemplo, sortear 10% dos elementos de cada categoria, obtendo, então, uma amostra estratificada com 10% dos empregados da firma.
7. a) Você acha que o Estado deve fornecer transporte coletivo gratuito para velhinhos e velhinhas que precisam usar ônibus?
b) Você acha que o dinheiro dos impostos deve ser gasto em subsídios para que as pessoas idosas tenham o privilégio de andar de ônibus de graça?

CAPÍTULO 2

1. Número de alunos matriculados no curso do Colégio "Dr. J. Alves", segundo a série, nos anos de 1984 e 1985

SÉRIE	1984	1985
1.º	40	42
2.º	35	36
3.º	29	32
Total	104	110

2. Alunos matriculados e alunos reprovados na Escola de 1.º Grau "D. Alcina", segundo a série, em 1985

Série	Alunos matriculados		Alunos reprovados	
	N.º	Freq. relat. (%)	N.º	Freq. relat. (%)
1.º	40	25	8	38
2.º	40	25	5	24
3.º	40	25	4	19
4.º	40	25	4	19
Total	160	100	21	100

3. Número de docentes no Brasil, segundo o ano

Ano	N.º de docentes
1981	942 955
1982	960 905
1983	967 975

FONTE: IBGE (1984).

4. Número de estabelecimentos de ensino de 1.º grau no Brasil, públicos e privados, segundo o ano

Ano	Estabelecimentos		Total
	Públicos	Privados	
1981	182 836	9 793	192 629
1982	187 567	10 463	198 030
1983	180 983	9 934	190 917

FONTE: IBGE (1984).

- 5.

Altura média e peso médio de crianças da região Sul do Brasil, em 1975, segundo a idade

Idade (anos completos)	Altura média (cm)	Peso médio (kg)
3	94,1	14,4
4	100,8	16,0
5	107,4	18,0
6	113,1	19,9
7	118,0	21,6
8	124,1	24,1
9	129,3	26,5
10	133,7	29,0

FONTE: IBGE (1977).

- 6.

Distribuição de 1 000 pessoas segundo a opinião a respeito da liberação do filme para exibição em televisão

Opinião	N.º de pessoas	Freq. relat. (%)
Favorável à liberação	432	43,2
Contrária à liberação	322	32,2
Não tem opinião	124	12,4
Não quiseram declarar	122	12,2
Total	1 000	100,0

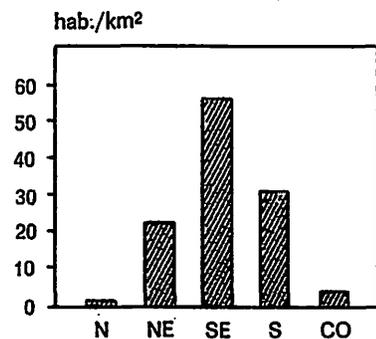
- 7.

Distribuição de 1 000 pessoas segundo a opinião da liberação do filme para exibição em televisão, por sexo

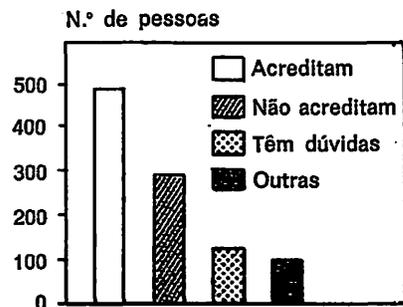
Opinião	Homens		Mulheres	
	N.º	Freq. relat.	N.º	Freq. relat.
Favorável à liberação	289	57,8	143	28,6
Contrária à liberação	120	24,0	202	40,4
Não tem opinião	13	2,6	111	22,2
Não quiseram declarar	78	15,6	44	8,8
Total	500	100,0	500	100,0

CAPÍTULO 3

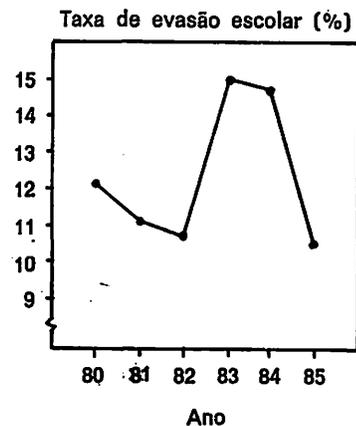
1. Densidade demográfica no Brasil, segundo a região, de acordo com o censo demográfico de 1980



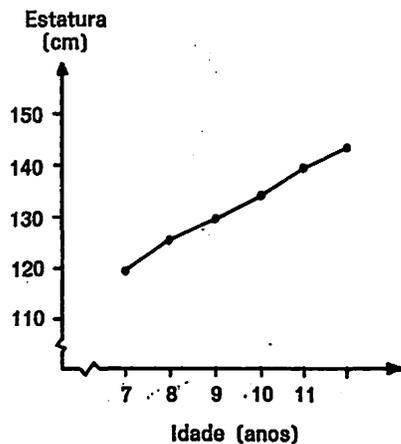
2. Distribuição de 1000 pessoas segundo a opinião sobre horóscopos



3. Taxa de evasão escolar na Escola de 1.º Grau "D. Pedro II", no período 1980-85

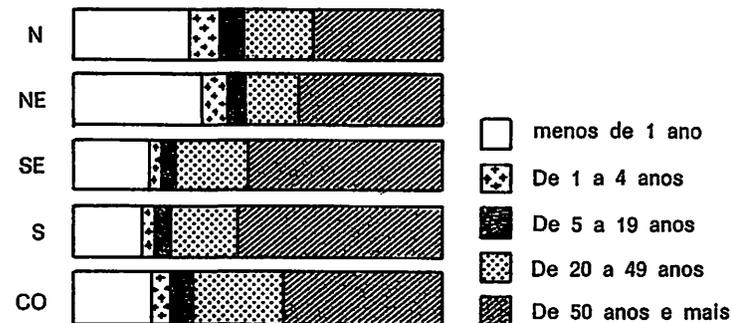


4. Altura média de crianças do sexo masculino residentes no Rio de Janeiro em 1975, segundo a idade



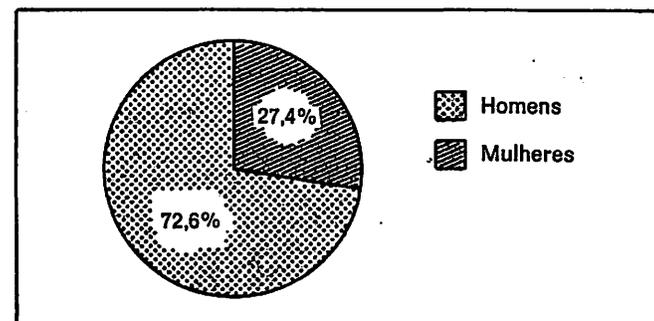
5.

Mortalidade proporcional por região do Brasil, segundo os grupos de idade, de acordo com o censo demográfico de 1980

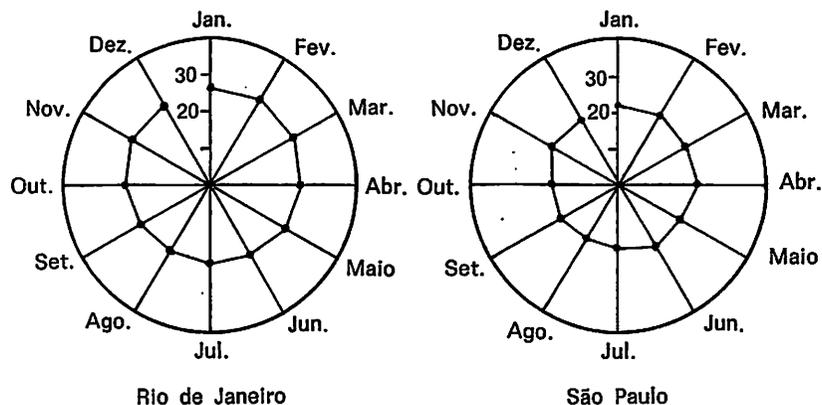


6.

Pessoas economicamente ativas no Brasil, em 1980, segundo o sexo



7. Temperatura do ar, segundo o mês, nas cidades do Rio de Janeiro e São Paulo, em 1983



CAPÍTULO 4

1. Distribuição de 40 alunos segundo a nota

Nota	Freqüência	Freq. relat. (%)
0 — 5	9	22,5
5 — 7	4	10,0
7 — 9	17	42,5
9 — 10	10	25,0
Total	40	100,0

2. Distribuição de recém-nascidos vivos segundo o peso

Classe	Ponto Médio	Freqüência	Freqüência relativa (%)	Freqüência acumulada
0,5 — 1,0	0,75	1	0,1	1
1,0 — 1,5	1,25	3	0,4	4
1,5 — 2,0	1,75	22	2,7	26
2,0 — 2,5	2,25	115	14,0	141
2,5 — 3,0	2,75	263	32,0	404
3,0 — 3,5	3,25	287	34,9	691
3,5 — 4,0	3,75	99	12,0	790
4,0 — 4,5	4,25	32	3,9	822
Total	—	822	100,0	—

3. As classes poderiam ser as seguintes:

0, 1, 2, 3, 4, 5, 6, 7 ou 8, 9 ou 10, mais de 10 filhos.

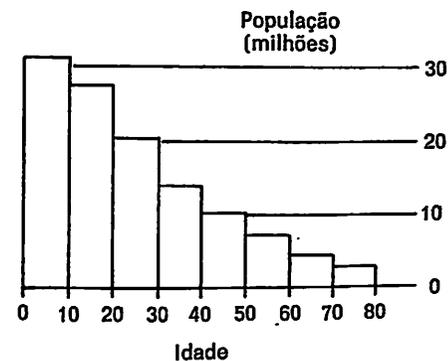
- 4.

Distribuição dos empregados de uma firma segundo o número de dias em que chegaram com atraso ao trabalho, no mês de fevereiro de 1986

N.º de dias	N.º de empregados
0	32
1 ou 2	19
3 ou 4	7
5 ou mais	4
Total	62

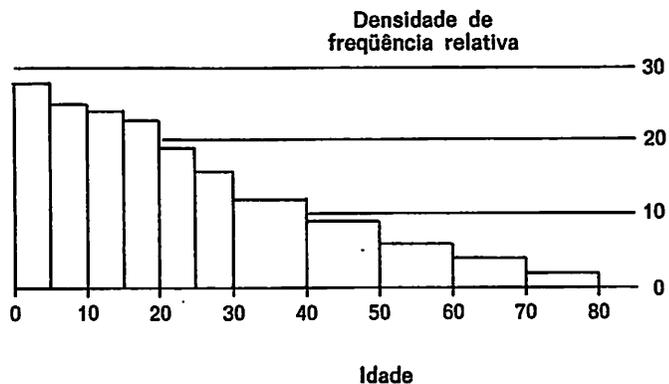
- 5.

População residente no Brasil, segundo os grupos de idade, de acordo com o censo demográfico de 1980



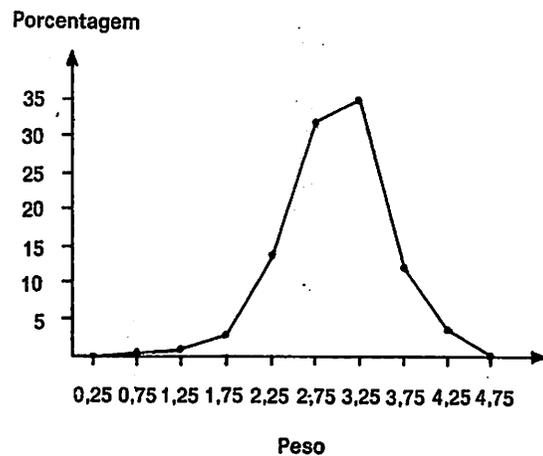
6.

População residente no Brasil, segundo os grupos de idade, de acordo com o censo demográfico de 1980



7.

Distribuição de pesos de recém-nascidos vivos



CAPÍTULO 5

- 57,17
- 3,3
- 5,13 minutos

4.

Médias ponderadas de cinco alunos

Aluno	Média
Ana	5,7
Cláudia	5,3
Marcos	5,0
Pedro	5,0
Sérgio	4,6

- A média geométrica é igual a 6.
- A média harmônica é 13,71.
- A moda é 0 (zero).
- A moda é 3,06.
- A mediana é 3,01.

CAPÍTULO 6

- A amplitude é 7,9
-

Resultado	Desvio
1	-2,5
2	-1,5
3	-0,5
4	0,5
5	1,5
6	2,5

- A variância é 12,67
- Marido: a variância é 106,4

Mulher: a variância é 40,2

Os pesos dos maridos apresentam maior dispersão.
- A média é 1,5 e a variância é 0,944.
- A variância é 96,55.
- a) O desvio-padrão é 5,00

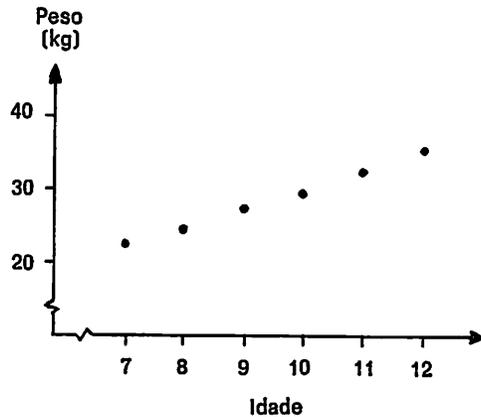
b) O desvio-padrão é 1,79

c) O desvio-padrão é 2,24

8. O coeficiente de variação do peso é 10,10% e o coeficiente de variação da estatura é 3,85%. Portanto, peso tem maior dispersão relativa.

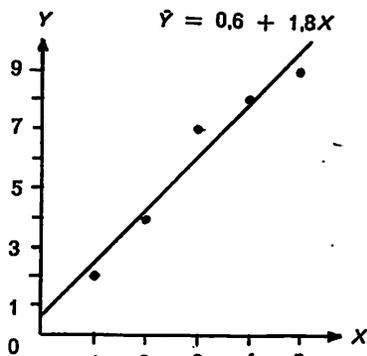
CAPÍTULO 7

- O coeficiente de correlação é zero
- O coeficiente de correlação é 0,976
- O coeficiente de correlação é -0,976
- Pesos médios de crianças do sexo masculino, da cidade de São Paulo, segundo a idade, em 1975

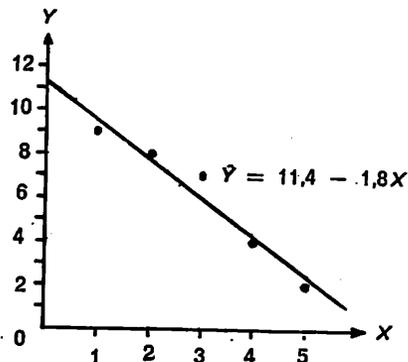


- No Brasil, o consumo de farinha de mandioca é maior entre famílias relativamente pobres. Correlação negativa.
- Chocolate é um alimento relativamente caro, que certamente é consumido em maior quantidade por famílias mais ricas. Correlação positiva.

6.



7.



- $\hat{Y} = 5 + X$, onde X e Y são, respectivamente, a altura do pai e a altura do filho, em centímetros.
- $\hat{Y} = 37,74$ kg
- $\hat{Y} = 23,79 - 1,717 X$

Para $X = 12$ obtemos $\hat{Y} = 3,2^\circ\text{C}$. Essa estimativa é uma extrapolação que, nesse caso, não tem validade. Embora a temperatura média mensal em Curitiba seja decrescente ao longo do primeiro semestre, essa tendência decrescente certamente não é válida para o segundo semestre.

CAPÍTULO 8

- 1/2
- 1/2
- 1/32
- $\left(\frac{1}{3}\right)^{13} = 6,27 \cdot 10^{-7}$ ou 627 vezes por bilhão.
- 0,84
- 0,75

7. Freqüência relativa dos grupos sanguíneos, em porcentagem

A	B	AB	O
35,6	11,5	2,1	50,8

8.

Taxa de mortalidade infantil, por 1 000 nascidos vivos, segundo a região do Brasil, em 1983

Região	Taxa de mortalidade infantil
Norte	94,3
Nordeste	90,2
Sudeste	47,5
Sul	35,1
Centro-Oeste	39,6

- A probabilidade de ocorrer cara continua igual a 1/2.

CAPÍTULO 9

1. 148
2. 200
3. 120, 90, 100 e 130
4. 103
5. 1%
6. 90 mil cruzeiros
Cresceu 12,5%.
7. Diminuiu 9,1%.

8.

Índices relativos de preços do arroz em casca, no período de janeiro a dezembro de 1984, tomando como base janeiro de 1984

Mês	Preço	Índice relativo
Jan.	163	100,0
Fev.	195	119,6
Mar.	207	127,0
Abr.	217	133,1
Mai	221	135,6
Jun.	239	146,6
Jul.	252	154,6
Ago.	271	166,3
Set.	291	178,5
Out.	345	211,7
Nov.	394	241,7
Dez.	508	311,7

9. e 10.

Índice geral de preços e preço do arroz, segundo o mês

Mês	Índice geral de preços, com base em 1977	Preço corrente do arroz	Preço real em Cr\$ de 1977	Preço real em Cr\$ de dez./84
Jan.	7.921,1	163	2,06	481
Fev.	8.892,1	195	2,19	512
Mar.	9.777,0	207	2,12	495
Abr.	10.651,1	217	2,04	476
Mai	11.594,7	221	1,91	445
Jun.	12.667,2	239	1,89	441
Jul.	13.974,3	252	1,80	421
Ago.	15.458,7	271	1,75	409
Set.	17.083,3	291	1,70	398
Out.	19.232,2	345	1,79	419
Nov.	21.131,6	394	1,86	435
Dez.	23.357,1	508	2,17	508

Referências Bibliográficas

01. ARENA, J. F. P. Estudo clínico-epidemiológico prospectivo das anomalias congênitas na população de Campinas, S.P. Campinas, 1974. 229 p. [Tese (Doutoramento) — FCM — Unicamp].
02. BANCO MUNDIAL. Relatório sobre o desenvolvimento mundial 1984. Rio de Janeiro, Fundação Getúlio Vargas, 1984. 305 p.
03. CASTRO, J. Geopolítica da fome. 6.ª ed. São Paulo, Brasiliense, 1961. v. 1.
04. DANTAS, R.O. Alcoolismo entre trabalhadores rurais e urbanos. *Ciênc. Cult.*, 31(7):774-6, 1979.
05. FUNDAÇÃO GETÚLIO VARGAS. Conjuntura estatística. *Conjuntura econômica*, 40(1), 1986.
06. GARCIA, M.V. Blood groups and dermatoglyphics in a Brazilian population of arabian origin. *Ciênc. Cult.*, 29(7):826-9, 1977.
07. IBGE. Anuário estatístico do Brasil. 1977. Rio de Janeiro, v. 38, 848 p.
08. IBGE. Censo demográfico: dados gerais — migração — instrução — fecundidade — mortalidade. Rio de Janeiro, 1983. v. 1, tomo 4, n.º 1.
09. IBGE. Censo demográfico: mão-de-obra. Rio de Janeiro, 1983. v. 1, tomo 5, n.º 1.
10. IBGE. Censo demográfico: famílias e domicílios. Rio de Janeiro, 1983. v. 1, tomo 6, n.º 1.
11. IBGE. Anuário Estatístico do Brasil. 1984. Rio de Janeiro, v. 45. 1104 p.
12. UNICAMP. Serviço de Arquivo Médico e Estatística do Hospital das Clínicas da Faculdade de Ciências Médicas, 1979.