

# Estatística básica

## ▶ Introdução

Neste capítulo, daremos continuidade ao estudo da Estatística por meio da associação de medidas quantitativas que caracterizam e resumem um conjunto de dados. Convém, no início, revisarmos rapidamente alguns conceitos aprendidos.

- **População:** conjunto de elementos que têm em comum a característica que está sendo investigada em uma pesquisa.
- **Amostra:** subconjunto da população cujos elementos fornecerão as informações que estão sendo investigadas por meio de uma pesquisa.
- **Variável:** é o objeto de estudo (ou item investigado) de uma pesquisa. As variáveis classificam-se em: **quantitativas** (são aquelas que apresentam números como resposta: quantidade de filhos, altura, renda etc.) ou **qualitativas** (que apresentam como resposta uma característica ou preferência do entrevistado: cor, nome do candidato em que o entrevistado vai votar nas próximas eleições etc.)
- **Tabelas de frequência:** são tabelas que organizam e resumem o conjunto de dados coletados em uma pesquisa. Em uma tabela de frequências geralmente constam a **frequência absoluta ( $F_a$ )**, que corresponde ao número de vezes que cada valor da variável aparece nos dados obtidos; a **frequência relativa ( $F_r$ )**, que é a razão entre a frequência absoluta e o número total de dados disponíveis; a frequência relativa pode ser expressa também na forma de porcentagem.
- **Classes (ou intervalos) de valores:** são intervalos reais usados para agrupar os valores de uma variável quantitativa, quando estes são demasiadamente diversificados, não havendo praticamente repetição de valores. Por exemplo, nas pesquisas sobre renda mensal, é comum apresentar os resultados em classes de valores. Em geral, usamos a notação  $a - b$  para representar o intervalo real  $[a, b[$ , cuja **amplitude** é  $b - a$ .
- **Representações gráficas:** o uso de gráficos é um importante recurso usado em diversas mídias (jornais, revistas, internet etc.) para representar um conjunto de dados. Entre as vantagens do uso de gráficos estão a rapidez da absorção de informações por parte do leitor, além de seu forte apelo visual e estético.



Jovens sendo entrevistados em uma pesquisa.

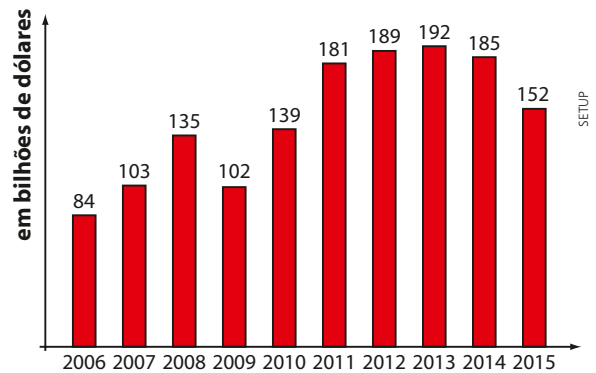
Estudamos as seguintes representações gráficas:

### Gráfico de barras

Podemos ter o gráfico de barras verticais e também o gráfico de barras horizontais.

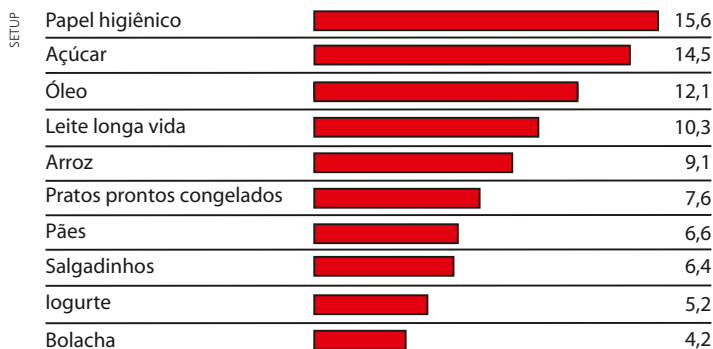
O gráfico de barras verticais ao lado mostra a evolução da corrente do comércio no Brasil, de 2006 a 2015, considerando os 5 primeiros meses de cada ano. A corrente do comércio é a soma das exportações e importações brasileiras, considerando um dado período.

**Corrente de comércio do Brasil — De janeiro a maio**



Fonte: O Estado de S. Paulo. Disponível em: <economia.estadao.com.br/noticias/geral,comercio-exterior-registra-queda-de-18-no-ano-,1700993>. Acesso em: 22 fev. 2016.

**Categorias com maior fatia de marca própria nos supermercados nas vendas do 1º semestre de 2015 (em porcentagem)**



Fonte: O Estado de S. Paulo. Disponível em: <economia.estadao.com.br/blogs/marcia-de-chiara/consumidor-resiste-a-marca-propria-mesmo-com-o-orcamento-apertado/>. Acesso em: 15 mar. 2016.

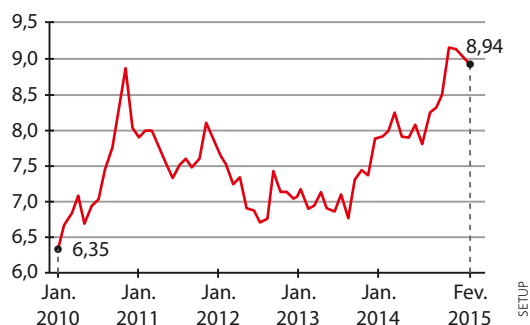
O gráfico de barras horizontais ao lado mostra, para alguns itens, o percentual de vendas (em relação ao total de vendas do item) representado por marcas próprias de grandes supermercados no Brasil.

### Gráfico de linhas

O gráfico de linhas é muito usado para representar valores de uma variável no decorrer de um intervalo de tempo.

No gráfico ao lado está representada a variação do preço do quilograma da carne no período de janeiro de 2010 a fevereiro de 2015.

**Preço da carne bovina no atacado da Grande São Paulo (em reais por quilograma)**



Fonte: O Estado de S. Paulo. Disponível em: <economia.estadao.com.br/noticias/geral,alta-no-preco-do-boi-derruba-frigorificos,1668492>. Acesso em: 15 mar. 2016.

## Gráfico de setores

É um tipo de gráfico que apresenta um círculo dividido em setores. No exemplo ao lado, temos o perfil brasileiro (brancos, pretos, amarelos, indígenas e pardos) segundo o último Censo, realizado em 2010.

## Pictograma

Nos pictogramas são usadas imagens que guardam relação com o assunto exposto. É uma forma de comunicar informações que desperta a atenção e a curiosidade do leitor.

No gráfico seguinte foram usadas imagens de carros para representar o aumento de vendas em quatro meses consecutivos de um automóvel recém-lançado no mercado.

**Número de vendas de automóveis**



Cada carro representado no gráfico corresponde a 2000 unidades vendidas. Observe também os fracionamentos do carro nos meses de abril, junho e julho.

## Histograma

É uma representação gráfica semelhante ao gráfico de barras, mas é usada quando se quer representar valores de uma variável agrupados em intervalos.

Você se lembra do significado do IDH?

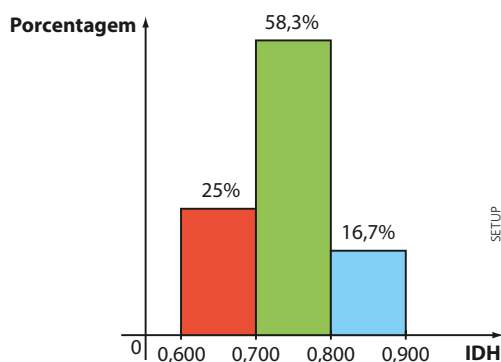
Ele é o Índice de Desenvolvimento Humano, que mede o bem-estar da população com base em expectativa de vida, escolaridade e PIB *per capita*. O IDH varia entre 0 e 1; quanto maior o seu valor, melhor a qualidade de vida.

Acompanhe, na tabela ao lado, o IDH dos países sul-americanos (dados de 2014).

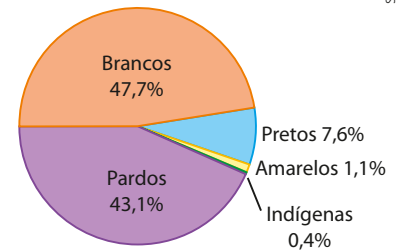
Podemos representar esse conjunto de dados agrupando-os em três intervalos de mesma amplitude.

Obtemos o histograma seguinte:

**Distribuição do IDH dos países sul-americanos – 2014**



**Perfil do brasileiro – 2010**



Fonte: IBGE. Disponível em: <biblioteca.ibge.gov.br/visualizacao/periodicos/94/cd\_2010\_religiao\_deficiencia.pdf>. Acesso em: 15 mar. 2016.

**IDH dos países sul-americanos – 2014**

País	IDH
Argentina	0,836
Bolívia	0,662
Brasil	0,755
Chile	0,832
Colômbia	0,720
Equador	0,732
Guiana	0,636
Paraguai	0,679
Peru	0,734
Suriname	0,714
Uruguai	0,793
Venezuela	0,762

Fonte: United Nations Development Programme – Human Development Index (HDI). Disponível em: <hdr.undp.org/en/data>. Acesso em: 15 mar. 2016.



## EXERCÍCIOS



- 1** Em uma pesquisa realizada com 150 trabalhadores, foram levantadas várias informações, como o tempo (em anos) que o trabalhador está em seu emprego atual. Os resultados estão apresentados na seguinte tabela de frequências:

Tempo (em anos)	Frequência absoluta	Frequência relativa	Porcentagem (%)
1	45	d	h
2	48	e	i
3	a	f	20
4	b	g	j
5 ou mais	c	0,16	k
<b>Total</b>	<b>150</b>	<b>1,00</b>	<b>100</b>

Dados elaborados pelo autor.

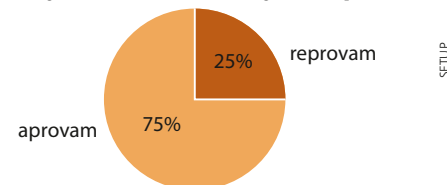


THINKSTOCK/GETTY IMAGES

- Qual é o valor de  $e + f + g$ ?
- Quantos funcionários estão há pelo menos 3 anos no atual emprego?
- Qual é o valor de  $h + k$ ?
- Se esse conjunto de dados fosse representado em um gráfico de setores, qual seria a medida aproximada do ângulo central do setor correspondente aos trabalhadores com 2 anos no emprego atual?

- 2** O gráfico ao lado ilustra o resultado de uma pesquisa sobre a aprovação da administração do prefeito de uma cidade um ano após sua posse. Sabe-se que foram ouvidas 480 pessoas.

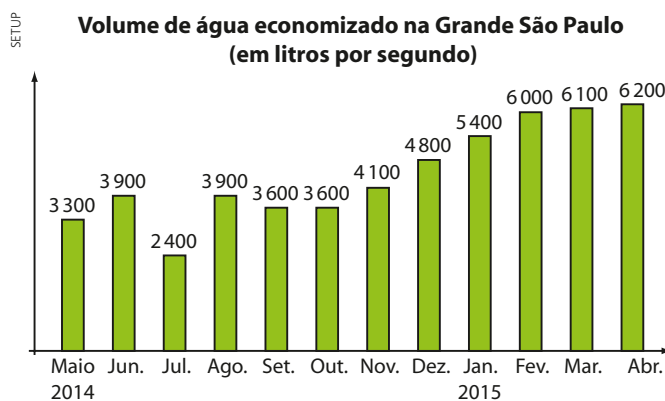
### Aprovação da administração do prefeito



Dados elaborados pelo autor.

- Quantas pessoas aprovam a administração do prefeito?
- Quais as medidas dos ângulos dos setores desse gráfico?
- Supondo que as mulheres representam 60% entre os que aprovam e 45% entre os que reprovam, determine a diferença entre o número de homens que aprovam e o número de homens que reprovam a administração desse prefeito.

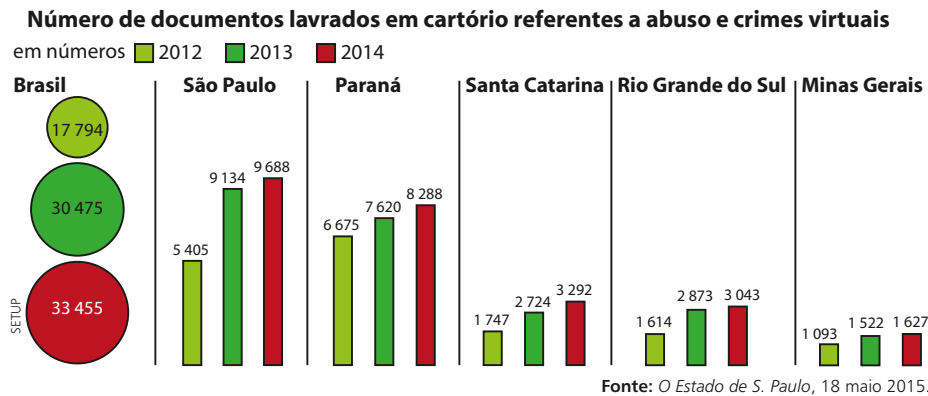
- 3** O gráfico abaixo apresenta o volume de água economizado (em L/s) pela população da Grande São Paulo, durante a crise hídrica iniciada em 2014.



Fonte: O Estado de S. Paulo. Disponível em: <saopaulo.estadao.com.br/noticias/geral,1-ano-de-bonus-da-agua-rende-o-equivalente-a-uma-guarapiranga,1684965>. Acesso em: 15 mar. 2016.

- Qual é a diferença entre o maior e o menor volume de água economizado por mês, em L/s, no período considerado?
- Considerando que o mês de abril tem 30 dias, determine o volume total de água economizado em abril de 2015.
- Se um estudante tivesse que apresentar esse conjunto de dados usando outro gráfico, escolhido entre os gráficos de setores ou de linhas, qual deles melhor o representaria? Por quê?

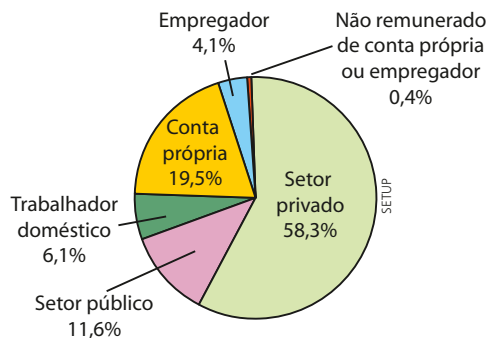
- 4 O gráfico comparativo abaixo mostra o crescimento do número de documentos lavrados em cartório que comprovam abusos e crimes virtuais, tais como vazamento de fotos e vídeos íntimos, perfis falsos, difamação, entre outros.



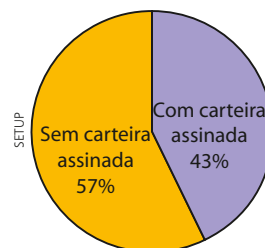
Com base nos gráficos, analise as afirmações seguintes classificando-as em verdadeira (V) ou falsa (F) e justificando as falsas:

- Em 2014, São Paulo registrava mais de  $\frac{1}{3}$  do total de ocorrências do país.
  - Entre os Estados listados, Santa Catarina registrou o maior aumento percentual do número de documentos lavrados em 2014, na comparação com 2013.
  - O aumento percentual ao qual se refere o item b, para Santa Catarina, é maior que 25%.
  - Considerando o número total de ocorrências na Região Sul, em 2014, é correto afirmar que o Paraná concentrava mais de 50% desse total.
  - Para o Estado de Minas Gerais, é possível dizer que o aumento no número de documentos nos três anos considerados é aproximadamente linear.
- 5 De acordo com os resultados da Pesquisa Mensal de Emprego (PME), em setembro de 2015, o número de pessoas potencialmente ativas para exercer alguma atividade de trabalho nas seis regiões metropolitanas onde a pesquisa é realizada — Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre — foi estimado em 43,9 milhões. Destes, 22,7 milhões de pessoas estão ocupadas.

**Distribuição da população ocupada segundo as categorias de posição na ocupação – Set. 2015**



**Trabalhadores domésticos – Set. 2015**



Fonte: IBGE. Disponível em: <ftp://ftp.ibge.gov.br/Trabalho\_e\_Rendimento/Pesquisa\_Mensal\_de\_Emprego/fasciculo\_indicadores\_ibge/2015/pme\_201509pubCompleta.pdf>. Acesso em: 15 mar. 2016.

Considerando os gráficos anteriores, determine:

- o número inteiro mais próximo da medida, em graus, do ângulo do setor correspondente a “conta própria”.
- o número de trabalhadores do setor privado nas seis regiões metropolitanas onde a pesquisa foi realizada.
- em relação ao total de pessoas ocupadas, o percentual aproximado dos trabalhadores domésticos que trabalham sem carteira assinada.
- o número de trabalhadores domésticos nas seis regiões metropolitanas.
- a diferença entre as medidas dos ângulos dos setores correspondentes aos trabalhadores domésticos sem e com carteira assinada.

- 6 Na tabela abaixo estão registrados os números (arredondados) de operações (pousos e decolagens) realizadas por aviões comerciais, nos cinco principais aeroportos de certo país, em determinado período.

**Número de pousos e decolagens**

Aeroporto	I	II	III	IV	V
<b>Número de operações</b>	7 500	10 500	13 500	4 500	6 750

Dados elaborados pelo autor.


- a) Construa um pictograma para representar os dados, utilizando desenhos da vista superior de aviões (lembre que o avião tem formato simétrico, sendo facilmente possível representar sua metade).
- b) Se esse conjunto de dados fosse representado em um gráfico de setores, qual seria a medida aproximada do ângulo correspondente ao aeroporto II?
- 7 Na tabela seguinte constam os valores da mortalidade infantil nos estados brasileiros em 2013. Tais valores referem-se ao número de crianças entre mil nascidas vivas que morrem durante o primeiro ano de vida.

**Taxa de mortalidade infantil em 2013 nos estados brasileiros**

Estado	%	Estado	%
Acre	19,2	Paraíba	19,0
Alagoas	24,0	Paraná	10,6
Amapá	23,9	Pernambuco	14,9
Amazonas	20,0	Piauí	21,1
Bahia	19,9	Rio de Janeiro	12,7
Ceará	16,6	Rio Grande do Norte	17,0
Distrito Federal	11,2	Rio Grande do Sul	10,5
Espírito Santo	10,1	Rondônia	21,3
Goiás	16,2	Roraima	17,8
Maranhão	24,7	Santa Catarina	10,1
Mato Grosso	18,1	São Paulo	10,8
Mato Grosso do Sul	15,4	Sergipe	18,9
Minas Gerais	12,6	Tocantins	17,4
Pará	18,3		

Fonte: Exame. Disponível em: <exame.abril.com.br/brasil/noticias/os-melhores-e-os-piores-estados-em-indicadores-de-saude#1>. Acesso em: 15 mar. 2016.

- a) Construa uma tabela de frequências, agrupando os dados em cinco classes de amplitude 3, começando pelo valor 10.
- b) Faça um histograma para representar esses valores, usando o item anterior.

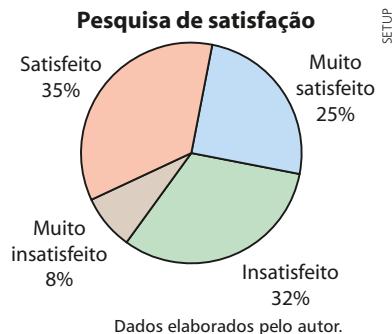
- 8 No pictograma ao lado estão representadas as populações das duas maiores regiões metropolitanas de um certo país. Cada  representa 1,5 milhão de habitantes.



- a) Determine as populações das regiões P e Q.
- b) Sabendo que a área de P é de 135 000 km<sup>2</sup>, obtenha sua densidade demográfica.

- 9 Foi realizada uma pesquisa com 1800 consumidores para conhecer o grau de satisfação do cliente em relação aos serviços prestados pela sua companhia de telefonia celular.

Os resultados são mostrados no gráfico de setores seguinte:



- a) Determine, para cada setor representado, a medida do ângulo central.
- b) Qual é o número de consumidores entrevistados que estão satisfeitos com os serviços prestados?
- c) Se  $\frac{5}{12}$  do número de consumidores insatisfeitos tivessem respondido "satisfeitos", qual seria o acréscimo, em graus, da medida do ângulo do setor "satisfeitos"? Admita que não haja outras alterações nas respostas.

- 10 (Enem-MEC) Uma empresa de alimentos oferece três valores diferentes de remuneração a seus funcionários, de acordo com o grau de instrução necessário para cada cargo. No ano de 2013, a empresa teve uma receita de 10 milhões de reais por mês e um gasto mensal com a folha salarial de R\$ 400 000,00, distribuídos de acordo com o gráfico 1. No ano seguinte, a empresa ampliará o número de funcionários, mantendo o mesmo valor salarial para cada categoria. Os demais custos da empresa permanecerão constantes de 2013 para 2014. O número de funcionários em 2013 e 2014, por grau de instrução, está no gráfico 2.

**Distribuição da folha salarial**

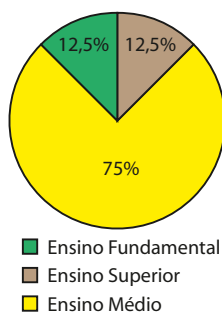


gráfico 1

**Número de funcionários por grau de instrução**

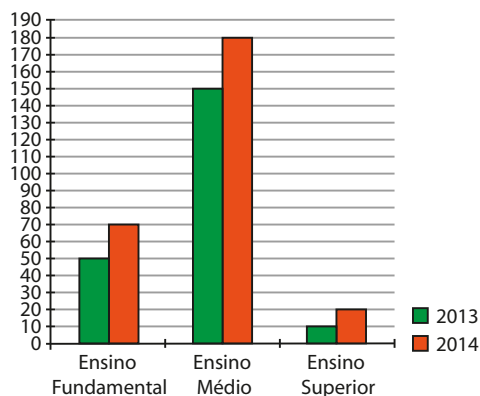
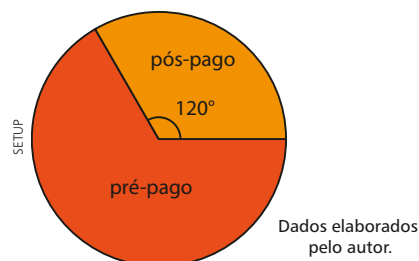


gráfico 2

Qual deve ser o aumento na receita da empresa para que o lucro mensal em 2014 seja o mesmo de 2013?

- a) R\$ 114 285,00                      d) R\$ 210 000,00
- b) R\$ 130 000,00                    e) R\$ 213 333,00
- c) R\$ 160 000,00
- 11 Os resultados de uma pesquisa realizada com um certo número de consumidores são mostrados a seguir:

Seu plano de celular é pré-pago ou pós-pago?



Dados elaborados pelo autor.

Para os que responderam "pré-pago" foi feita a seguinte pergunta:

Você acessa a internet pelo seu aparelho?



Dados elaborados pelo autor.

Com base nos gráficos, determine o percentual de consumidores em relação ao total de entrevistados na pesquisa que:

- a) possuem plano de celular pós-pago.
- b) possuem plano de celular pré-pago e não acessam a internet por meio do aparelho.

## As pesquisas eleitorais

No Brasil ocorrem eleições a cada dois anos, e, em decorrência dessa regularidade, o brasileiro se habituou a acompanhar pela mídia as pesquisas eleitorais, encomendadas aos institutos especializados, no período de eleições.

Como já sabemos, para colher as informações necessárias, usam-se amostras e, a partir dos resultados obtidos na amostra, características de todos os eleitores são estimadas, fazendo uso dos conceitos de **margem de erro** e do **nível de confiança**.

O conceito de margem de erro é mais conhecido do brasileiro. Imagine que uma pesquisa eleitoral aponte que a porcentagem de eleitores que irão votar no candidato **X** é de 30%, por exemplo. Considerar uma margem de erro de 2,5 pontos percentuais para mais ou para menos significa dizer que o intervalo  $[30 - 2,5; 30 + 2,5] = [27,5; 32,5]$  *provavelmente* contém a verdadeira proporção de eleitores do candidato **X** em toda a população.

O termo “provavelmente” está relacionado a um conceito bem menos divulgado na mídia e praticamente desconhecido da população brasileira: é o nível de confiança da pesquisa, que está associado aos intervalos de confiança. Em pesquisas eleitorais, costuma-se usar um nível de confiança de 95%. O que isso significa?

Imagine que com outra amostra de mesmo tamanho dessa população fosse realizada outra pesquisa nas mesmas condições que a anterior. Suponha que essa outra pesquisa tenha apontado um percentual de 28,5% para o candidato **X**; com a margem de erro da pesquisa, construímos o intervalo  $I_2 = [28,5 - 2,5; 28,5 + 2,5] = [26; 31]$  que provavelmente contém a verdadeira proporção de eleitores de **X** na população. Com uma terceira amostra, obteríamos, analogamente, um intervalo  $I_3$  e assim por diante.

Considerando 100 diferentes amostras de mesmo tamanho dessa população, teríamos 100 intervalos:  $I_1, I_2, I_3, \dots, I_{100}$ .

O nível de confiança de 95% significa que 95 desses 100 intervalos contém a verdadeira proporção de eleitores que votam no candidato **X**, em toda a população.

Assim, quando dispomos de uma informação como “o candidato **X** atingiu 30% das intenções de voto, em uma pesquisa cuja margem de erro é de 2,5 pontos percentuais, com nível de 95% de confiança”, é importante ter em mente que não é 100% certeza que o intervalo  $[27,5; 32,5]$  contém o verdadeiro percentual de eleitores do candidato **X** na população. Há 5% de chance de o intervalo não conter o verdadeiro percentual. Abre-se, desse modo, a possibilidade de um erro estatístico, inerente à natureza da pesquisa eleitoral.



Cabina de votação com urna eletrônica, dispositivo que armazena as escolhas dos eleitores e agiliza o processo de apuração dos votos.

ALEXANDRE TOKITAKA/PULSAR IMAGENS



## ▶ Medidas de centralidade e dispersão

Considere um levantamento feito, no mês de janeiro, sobre as temperaturas máximas diárias na cidade de Brasília. Já vimos que os dados referentes aos 31 valores medidos da temperatura podem ser representados em uma tabela de frequência ou em alguma representação gráfica.

Vamos, agora, estabelecer para esses dados algumas medidas (números reais) que sejam representativas, isto é, que resumam e caracterizem o conjunto de todas as temperaturas máximas registradas.

Há dois tipos de medidas:

- **medidas de centralidade** (ou de **tendência central** ou de **posição**): por meio de determinado valor, que revela uma tendência central para as temperaturas máximas, representaremos os 31 valores medidos. Vamos estudar três medidas de centralidade: a **média**, a **mediana** e a **moda**.
- **medidas de dispersão** (ou de **variabilidade**): de modo geral, essas medidas indicam se, em um conjunto de dados, os valores estão concentrados (menos dispersos, menos espalhados) ou não (mais dispersos, mais espalhados) em torno de um valor central (geralmente a média). As medidas de dispersão permitirão comparar dois conjuntos de dados quanto ao grau de homogeneidade, indicando em qual deles há maior (ou menor) variabilidade, isto é, em qual deles os dados variam mais (ou menos).

Vamos estudar quatro medidas de dispersão: a **amplitude**, a **variância**, o **desvio padrão** e o **desvio médio**.

Para simplificar a notação que será usada neste capítulo, vamos introduzir um importante símbolo da linguagem matemática: o somatório, indicado pela letra grega  $\Sigma$  (lê-se: "sigma"). Ele representa aqui a soma de um número finito de parcelas que têm alguma característica comum. Acompanhe os exemplos seguintes:

- O símbolo  $\sum_{i=1}^4 i^2$  (lê-se: "somatório (ou soma) de  $i^2$ , para  $i$  variando de 1 até 4") significa que devemos atribuir para  $i$ , sucessivamente, os valores 1, 2, 3 e 4, calcular os respectivos valores numéricos da expressão  $i^2$  e adicionar os valores encontrados, isto é:

$$\sum_{i=1}^4 i^2 = \underbrace{1^2}_{i=1} + \underbrace{2^2}_{i=2} + \underbrace{3^2}_{i=3} + \underbrace{4^2}_{i=4} = 1 + 4 + 9 + 16 = 30$$

- Para calcularmos o valor de  $\sum_{n=0}^2 \left(\frac{1}{n+1}\right)$  é preciso atribuir para  $n$  os valores 0, 1 e 2:

$$\begin{array}{l} n = 0 \Rightarrow \frac{1}{0+1} = 1 \\ n = 1 \Rightarrow \frac{1}{1+1} = \frac{1}{2} \\ n = 2 \Rightarrow \frac{1}{2+1} = \frac{1}{3} \end{array} \quad \left. \begin{array}{l} \diagup \\ \diagdown \end{array} \right\} \sum_{n=0}^2 \left(\frac{1}{n+1}\right) = 1 + \frac{1}{2} + \frac{1}{3} = \frac{11}{6}$$

ALAMY/FOTOARENA



Vista do Eixo Monumental, uma avenida localizada em Brasília (DF), 2015.

## ▶ Medidas de centralidade

### ▶ Média aritmética

Observe a imagem a seguir: nela estão representados seis suportes para copos descartáveis, distribuídos em um centro de convenções, onde será realizado um evento. A quantidade de copos em cada suporte está indicada abaixo.



Se todos os copos fossem retirados dos suportes e distribuídos igualmente entre o número de suportes, qual seria a quantidade de copos descartáveis em cada um?

- Devemos, inicialmente, adicionar a quantidade de copos:

$$73 + 41 + 36 + 19 + 59 + 84 = 312$$

- Dividimos a soma encontrada pela quantidade de suportes (seis):

$$312 : 6 = 52$$

O valor encontrado (52 copos) representa a **média aritmética** entre os números 73, 41, 36, 19, 59 e 84.

### Definição

Sejam  $x_1, x_2, \dots, x_n$  a relação dos valores assumidos por uma determinada variável quantitativa  $x$ . Definimos **média aritmética** (indica-se por  $\bar{x}$ ) como a razão entre a soma de todos esses valores e o número total de valores:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Usando o símbolo de somatório para representar o numerador dessa expressão, escrevemos:

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

Ao calcularmos a média de uma relação de valores assumidos por uma variável, obtemos um número real que pode ou não coincidir com algum dos valores assumidos pela variável.

**EXEMPLO 1**

Os valores seguintes referem-se às notas obtidas por um aluno em oito disciplinas do Ensino Médio em um certo bimestre do ano letivo:

$$7,5 — 6,0 — 4,2 — 3,9 — 4,8 — 6,2 — 8,0 — 5,4$$

Vamos calcular a média aritmética desses valores:

$$\bar{x} = \frac{\sum_{i=1}^8 X_i}{8} = \frac{7,5 + 6,0 + 4,2 + 3,9 + 4,8 + 6,2 + 8,0 + 5,4}{8} = \frac{46}{8} = 5,75$$

Qual é o significado desse valor?

Caso o aluno apresentasse a mesma nota (desempenho) em todas as disciplinas, ela deveria ser 5,75 a fim de que fosse obtida a pontuação total de 46 pontos, que é a soma dos pontos efetivamente obtidos nessas oito disciplinas.

Observe que em nenhuma disciplina o aluno obteve a nota média, que é 5,75. Nesse caso, a média aritmética não coincide com qualquer uma das notas obtidas pelo aluno.

Suponhamos agora que, no bimestre seguinte, o aluno tenha obtido as seguintes notas nas mesmas oito disciplinas:

$$6,6 — 7,2 — 7,8 — 6,4 — 5,9 — 6,0 — 6,5 — 4,8$$

A média aritmética  $\bar{x}$  desses valores é:

$$\bar{x} = \frac{6,6 + 7,2 + 7,8 + 6,4 + 5,9 + 6,0 + 6,5 + 4,8}{8} = \frac{51,2}{8} = 6,4$$

Nessa situação, a média obtida coincide com uma das notas obtidas pelo aluno.

**EXERCÍCIO RESOLVIDO**

- 1** A média dos salários de quinze funcionários de uma loja de autopeças é R\$ 1323,00. Se forem contratados mais dois funcionários, com salários de R\$ 1315,00 e R\$ 1450,00, qual será a nova média salarial da loja?

**Solução:**

A média inicial ( $\bar{x}$ ) de salários é 1323. Temos:

$1323 = \frac{\Sigma \text{salários}}{15} \Rightarrow \Sigma \text{salários} = 1323 \cdot 15 = 19845$ . Assim, antes das contratações, a soma de todos os salários dessa loja era de R\$ 19 845,00.

A soma dos salários, em reais, após a admissão dos dois funcionários será:

$$\Sigma^1 = 19845 + 1315 + 1450 = 22\ 610$$

e a nova média ( $\bar{x}^1$ ) de salários, também em reais, será:

$$\bar{x}^1 = \frac{\Sigma^1 \text{salários}}{17} = \frac{22\ 610}{17} = 1330$$

A nova média salarial passará a ser de R\$ 1330,00.

## Média aritmética ponderada

Considere o seguinte problema:

Em um espetáculo musical, foram vendidos 1200 ingressos cujos valores dependiam do setor escolhido no teatro, como mostra o quadro abaixo:

Sector	Números de ingressos vendidos	Preço unitário do ingresso
Pista	720	R\$ 50,00
Andar superior	400	R\$ 150,00
Camarote	80	R\$ 300,00

Qual foi o valor médio do ingresso pago nesse espetáculo?

Consideremos que a variável em estudo é o preço do ingresso. Fazendo a leitura do quadro, notamos que foram vendidos 720 ingressos a R\$ 50,00 cada; 400 ingressos a R\$ 150,00 cada e 80 ingressos a R\$ 300,00 cada.

Assim, o preço médio ( $\bar{p}$ ) do ingresso, em reais, é:

$$\begin{aligned}\bar{p} &= \frac{\overbrace{50 + 50 + \dots + 50}^{720 \text{ parcelas}} + \overbrace{150 + 150 + \dots + 150}^{400 \text{ parcelas}} + \overbrace{300 + \dots + 300}^{80 \text{ parcelas}}}{720 + 400 + 80} \Rightarrow \\ \Rightarrow \bar{p} &= \frac{720 \cdot 50 + 400 \cdot 150 + 80 \cdot 300}{1200} \Rightarrow \\ \Rightarrow \bar{p} &= \frac{36\,000 + 60\,000 + 24\,000}{1200} = \frac{120\,000}{1200} = 100\end{aligned}$$

A média obtida para o valor do ingresso, nesse problema, é chamada média aritmética ponderada dos valores R\$ 50,00, R\$ 150,00 e R\$ 300,00, em que o fator de ponderação (também chamado de peso) corresponde à quantidade de ingressos vendidos em cada setor.

Observe, nesse exemplo, que a média obtida não coincide com qualquer um dos preços do ingresso disponíveis para compra. O valor obtido para a média R\$ 100,00 é um valor teórico cujo significado é: se todos os 1200 ingressos tivessem sido vendidos pelo mesmo valor, este deveria ser de R\$ 100,00, a fim de que fosse obtida a arrecadação de R\$ 120 000,00.

De modo geral, consideremos uma relação de valores formada pelos elementos  $x_1, x_2, \dots, x_k$ , com frequências absolutas respectivamente iguais a  $n_1, n_2, \dots, n_k$ .

A média aritmética ponderada desses valores é:

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k (x_i \cdot n_i)}{n_1 + n_2 + \dots + n_k}$$

Podemos também expressar  $\bar{x}$  em termos da frequência relativa de cada  $x_i$  (com  $i = 1, 2, \dots, k$ ), a saber,  $f_i = \frac{n_i}{n_1 + n_2 + \dots + n_k}$ .



### PENSE NISTO:

Em que condições o preço médio do ingresso, em reais, seria

$$\begin{aligned}\frac{50 + 150 + 300}{3} &= \\ &= \frac{500}{3} \approx 166,67?\end{aligned}$$

Vamos escrever, convenientemente, a expressão obtida para  $\bar{x}$ :

$$\bar{x} = x_1 \cdot \underbrace{\frac{n_1}{n_1 + n_2 + \dots + n_k}}_{= f_1} + x_2 \cdot \underbrace{\frac{n_2}{n_1 + n_2 + \dots + n_k}}_{= f_2} + \dots + x_k \cdot \underbrace{\frac{n_k}{n_1 + n_2 + \dots + n_k}}_{= f_k}$$

Assim, obtemos:

$$\bar{x} = x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k = \sum_{i=1}^k (x_i \cdot f_i)$$

### EXEMPLO 2

Foi realizada uma pesquisa socioeconômica com duzentas pessoas de uma cidade para investigar alguns itens de conforto presentes na residência de cada entrevistado. Na tabela de frequências seguinte, constam os resultados referentes ao número de aparelhos de TV por domicílio:

Número de aparelhos de TV	Frequência absoluta	Frequência relativa
0	9	$\frac{9}{200} = 0,045$ ou 4,5%
1	93	$\frac{93}{200} = 0,465$ ou 46,5%
2	60	$\frac{60}{200} = 0,30$ ou 30%
3	22	$\frac{22}{200} = 0,11$ ou 11%
4	16	$\frac{16}{200} = 0,08$ ou 8%
<b>Total</b>	<b>200</b>	<b>1,00</b> ou <b>100%</b>

Qual é o número médio de TVs por domicílio?

A variável em estudo é o “número de TVs por domicílio” e ela assume os seguintes valores: 0, 1, 2, 3 e 4. Devemos, então, calcular a média aritmética ponderada desses valores usando como pesos a frequência absoluta, a frequência relativa na forma decimal ou a frequência relativa na forma percentual.

• frequência absoluta:

$$\text{Nesse caso, } \bar{x} = \frac{0 \cdot 9 + 1 \cdot 93 + 2 \cdot 60 + 3 \cdot 22 + 4 \cdot 16}{9 + 93 + 60 + 22 + 16} \Rightarrow \bar{x} = \frac{93 + 120 + 66 + 64}{200} = 1,715$$

• frequência relativa:

$$\text{Nesse caso, } \bar{x} = \frac{0 \cdot 0,045 + 1 \cdot 0,465 + 2 \cdot 0,30 + 3 \cdot 0,11 + 4 \cdot 0,08}{\underbrace{0,045 + 0,465 + 0,30 + 0,11 + 0,08}_{= 1}} \Rightarrow$$

$$\Rightarrow \bar{x} = 0 + 0,465 + 0,60 + 0,33 + 0,32 = 1,715$$

• frequência relativa (na forma percentual):

$$\text{Nesse caso, } \bar{x} = \frac{0 \cdot 4,5\% + 1 \cdot 46,5\% + 2 \cdot 30\% + 3 \cdot 11\% + 4 \cdot 8\%}{\underbrace{4,5\% + 46,5\% + 30\% + 11\% + 8\%}_{= 100\%}} \Rightarrow$$

$$\Rightarrow \bar{x} = \frac{46,5\% + 60\% + 33\% + 32\%}{100\%} = 1,715$$

Observe, novamente, que a média aritmética é um valor teórico, não coincidindo, nesse exemplo, com um valor assumido pela variável em estudo.



## EXERCÍCIOS



**12** Em cada caso, calcule a média aritmética dos valores:

- a) 23 — 20 — 22 — 21 — 28 — 20
- b) 7 — 9 — 9 — 9 — 7 — 8 — 8 — 9 — 9 — 9
- c) 0,1 — 0,1 — 0,1 — 0,1 — 0,2 — 0,2
- d) 4 — 4,5 — 4,5 — 5,0 — 5,0 — 5,5 — 6,5 — 5,0
- e) 3 — 3 — 3 — 3 — 3 — 3 — 3

**13** Em um edifício residencial com 54 apartamentos, 36 condôminos pagam taxa de condomínio de R\$ 270,00; para os demais, essa taxa é de R\$ 360,00. Qual é o valor da taxa média de condomínio nesse edifício?

**14** Seja  $a \in \mathbb{R}$ . A média aritmética entre  $a$ , 8,  $2a$ , 9 e  $(a + 1)$  é 6,8. Qual é o valor de  $a$ ?

**15** Um grupo **A** de 20 recém-nascidos tem massa média de 2,8 kg; um grupo **B** de 30 recém-nascidos tem massa média de 2,6 kg. Juntando os recém-nascidos dos grupos **A** e **B**, qual é o valor esperado para a média de massas?



**16** A média aritmética de uma lista formada por vinte números é 12. Qual será a nova média se:

- a) acrescentarmos o número 33 a essa lista?
- b) retirarmos o número 50 dessa lista?
- c) acrescentarmos o número 63 a essa lista e retirarmos o 51?

**17** Em uma fábrica, a média salarial das mulheres é R\$ 1408,00; para os homens a média salarial é R\$ 1632,00. Sabe-se, também, que a média geral de salários nessa fábrica é R\$ 1475,20.

- a) Sem fazer cálculos, responda: Há mais homens ou mulheres trabalhando na fábrica?
- b) Determine as quantidades de homens e de mulheres, sabendo que elas diferem de 32.

**18** A tabela seguinte informa a quantidade de cartões amarelos distribuídos, por um árbitro, em uma partida de futebol nos jogos por ele apitados durante uma temporada:

## Distribuição de cartões amarelos

Número de cartões	0	1	2	3	4
Frequência absoluta	30	18	7	3	2

Dados elaborados pelo autor.

- a) Quantos jogos o árbitro apitou na temporada?
- b) Qual é o número médio de cartões amarelos distribuídos por partida?
- c) Se esse conjunto de dados fosse representado em um gráfico de setores, qual seria a medida do ângulo do setor correspondente aos jogos nos quais foi distribuído um único cartão?

**19** Em um concurso público foram aprovados 80 candidatos, cuja média de notas foi 74,5 (em 100 pontos possíveis). A média das 40 menores notas dos candidatos aprovados foi 67,0. Qual foi a média das 40 maiores notas dos aprovados?

**20** Um professor calculou a média aritmética das notas dos quarenta alunos que submeteu a uma prova e obteve como resultado o valor de 5,5. Na hora de devolver as provas, verificou que havia cometido erro em duas delas. Na primeira, a nota correta era 9,5 em vez de 6,5 e, na segunda, a nota correta era de 5,5 em vez de 3,5. Feita a correção, de quanto foi acrescida a média das notas?

**21** Em um concurso, para o cálculo da nota final do candidato, adotam-se pesos nas provas, como mostra o quadro abaixo:

Prova	Peso
Redação	4
Matemática	3
Conhecimentos gerais	2

A nota final mínima para aprovação nesse concurso é 8,5.

- a) Um candidato **X** obteve 7,5, 9,0 e 9,5 nas provas de Redação, Matemática e Conhecimentos gerais, respectivamente. Ele foi aprovado?
- b) O candidato **Y** obteve 8,3 na Redação e 7,5 em Conhecimentos gerais. Qual é a nota mínima que ele deveria obter em Matemática, a fim de que fosse aprovado no concurso? Sabe-se que as notas do concurso são atribuídas de 0,1 em 0,1.

- 22** Em uma padaria trabalham 12 funcionários cujos cargos e salários estão abaixo descritos:

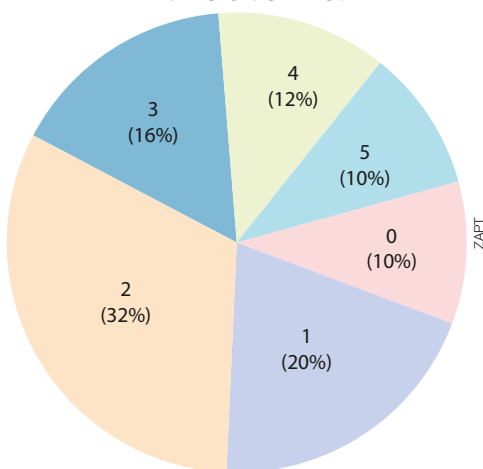
Cargos e salários

Cargo	Salário mensal	Número de funcionário
Gerente	R\$ 2 800,00	1
Atendente	R\$ 1 050,00	5
Padeiro	R\$ 1 300,00	2
Confeiteiro	R\$ 1 000,00	1
Caixa	R\$ 1 200,00	3

Dados elaborados pelo autor.

- a) Qual é o valor da folha de pagamento dessa padaria?
- b) Qual é a média salarial nessa padaria?
- c) O proprietário da padaria quer contratar dois segurancas especializados, mas sabe que a média salarial da padaria não pode ultrapassar R\$ 1 300,00. Qual é o maior salário que pode ser oferecido a cada um dos candidatos ao cargo de segurancas?
- 23** Realizou-se uma pesquisa entre as mulheres de uma cidade para levantar informações sobre o número de filhos. Foram entrevistadas 400 mulheres, e os dados obtidos estão representados no gráfico de setores abaixo.

Número de filhos



Dados elaborados pelo autor.

- a) Quantas mulheres da amostra têm três ou mais filhos?
- b) Qual é a média de filhos das mulheres dessa amostra?
- 24** Um professor de economia dá aula em 4 turmas: **A, B, C, D**. Na tabela seguinte, encontram-se os resultados obtidos pelas turmas no exame final, bem como a quantidade de alunos por turma:

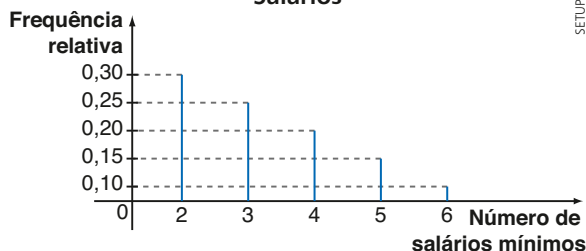
Notas

Turma	Números de alunos	Média das notas
A	30	6,2
B	35	7,2
C	55	5,4
D	?	5,0

Dados elaborados pelo autor.

- a) Considerando as turmas **A, B e C**, determine a média das notas obtidas nesse exame pelos alunos dessas 3 turmas reunidas.
- b) O professor não se lembra, ao certo, o número de alunos da turma **D**, mas ele sabe que a média geral das notas das 4 turmas reunidas não ultrapassou 5,8. Quantos alunos, no mínimo, podem pertencer à turma **D**?
- 25** O gráfico abaixo mostra a frequência relativa dos salários dos 40 funcionários (em salários mínimos) de uma pequena empresa:

Salários



Dados elaborados pelo autor.

- Sabendo que, no segundo semestre de 2015, o salário mínimo no Brasil era de R\$ 788,00, verifique se são verdadeiras ou falsas as afirmações seguintes, justificando.
- a) O salário médio nessa empresa era de 3,5 salários mínimos.
- b) A folha de pagamento dos salários dessa empresa era maior que R\$ 105 000,00.
- c) Se cada funcionário recebesse um aumento de R\$ 100,00 no salário, a média de salários dos funcionários dessa empresa ultrapassaria R\$ 2 800,00.
- d) Se todos os funcionários que recebem dois salários mínimos passassem a receber três salários mínimos, o salário médio da empresa passaria a ser maior do que R\$ 3 000,00.
- 26** Para um torneio de basquete foram convocados 12 jogadores para compor a seleção nacional. O time titular, com 5 jogadores, tinha altura média de 2,04 metros e o time reserva, com os demais, tinha altura média de 2,01 metros. Às vésperas do jogo de estreia um titular se contundiu e foi substituído por um reserva.

Com isso, a altura média do time titular aumentou 2 cm e a do time reserva diminuiu 1,5 cm. A seleção disputou o torneio com 11 jogadores.

- Qual a altura do jogador titular que se contundiu?
- Qual a altura do jogador reserva que substituiu o jogador contundido?

**27** Ao deixar um hotel *resort* na praia, os hóspedes são convidados a responder um questionário de avaliação do hotel. Ao final do questionário, pede-se uma nota de 0 a 5 (5 corresponde ao máximo de satisfação) para o hotel. Durante o 1º mês de uma temporada, os 2 000 questionários recebidos apontavam uma nota média de 3,9 para o hotel.



ALAMY/FOTOFARENA

- Qual é o número mínimo de questionários que devem ser respondidos, além dos que já foram, para que a nota média de avaliação do hotel passe a ser 4,6?
- É possível que a média do hotel passe a ser de 5,0 no bimestre, considerando os formulários que serão preenchidos no 2º mês? Em caso afirmativo, determine o número mínimo de questionários a serem preenchidos.

**28** A média aritmética de cinquenta números reais,  $x_1, x_2, x_3, \dots, x_{49}, x_{50}$ , é igual a 120. Qual é a média aritmética dos números reais  $x_1 + 1, x_2 + 2, x_3 + 3, \dots, x_{49} + 49$  e  $x_{50} + 50$ ?

## ► Mediana

Considere o consumo mensal de água, em metros cúbicos, de uma residência nos nove primeiros meses de um ano:

33 — 31 — 34 — 32 — 34 — 32 — 102 — 34 — 30

Calculando a média mensal de consumo, obtemos:

$$\bar{x} = \frac{33 + 31 + 34 + 32 + 34 + 32 + 102 + 34 + 30}{9} = \frac{362}{9} \approx 40,2$$

O valor encontrado para a média,  $40,2 \text{ m}^3$ , não representa, com fidelidade, uma medida de tendência central: o consumo mensal dessa residência aponta para um valor entre 30 e 35 metros cúbicos; além disso, dos 9 valores registrados, 8 são menores que a média e “distantes”, ao menos, 6 unidades dela e apenas 1 valor é maior que a média, estando muito distante dela.

Nessa situação, a média foi afetada por um valor muito discrepante do consumo, que destoa dos demais: o valor de  $102 \text{ m}^3$  registrado em um mês pode ser explicado por um vazamento de água em alguma válvula ou torneira ou esse consumo atípico pode ser devido a uma hospedagem temporária na casa (parentes de outra cidade, por exemplo).



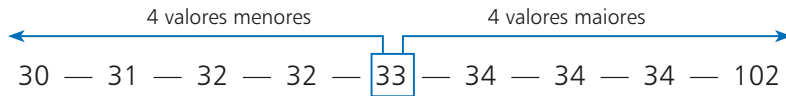
ULRICH NIEHOFF/IMAGETOCKER RM/DIOMEDIA

Hidrômetro, aparelho que mede o consumo de água nos imóveis.



Deste modo, é importante conhecermos outra medida de centralidade, além da média, a fim de que façamos uma análise mais completa para interpretar e caracterizar o conjunto de dados. Essa outra medida de centralidade é a **mediana**.

Colocando em ordem crescente os valores de consumo da relação anterior, obtemos:



O valor destacado, 33, separa o conjunto de dados em duas partes: na primeira  $\{30, 31, 32, 32\}$  todos os valores são menores que 33 e na outra  $\{34, 34, 34, 102\}$  todos os valores são maiores que 33.

O valor  $33 \text{ m}^3$  é chamado mediana e representa, nesse exemplo, uma medida de centralidade mais fiel ao conjunto de dados.

### Definição

Sejam  $x_1 \leq x_2 \leq \dots \leq x_n$  os  $n$  valores ordenados assumidos por uma variável quantitativa  $X$ , em um conjunto de observações.

Define-se a **mediana** (indicaremos por  $Me$ ) por meio da relação:

$$Me = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ for ímpar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ for par} \end{cases}$$

A definição garante que a mediana seja um valor central que divide o conjunto de dados em dois subconjuntos com o mesmo número de elementos. Em um subconjunto, todos os elementos são menores que a mediana ou iguais a ela; no outro subconjunto, todos os elementos são maiores que a mediana ou iguais a ela.

#### EXEMPLO 3

O controle de qualidade de uma indústria forneceu o seguinte número de peças defeituosas (por lote de 100 unidades):

6 — 4 — 9 — 6 — 3 — 8 — 1 — 4 — 5 — 6

Vamos determinar a mediana do número de peças defeituosas. Para isso, ordenamos esses valores:

1 — 3 — 4 — 4 — 5 — 6 — 6 — 6 — 8 — 9 \*

Como  $n = 10$  é par, pela definição a mediana será dada pela média aritmética entre o 5º e o 6º valores de \*, isto é:

$$Me = \frac{x_5 + x_6}{2} = \frac{5 + 6}{2} = 5,5$$

Note que, em \*, temos cinco valores menores que 5,5 (1 — 3 — 4 — 4 — 5) e cinco valores maiores que 5,5 (6 — 6 — 6 — 8 — 9).

## ► Moda

Após fazer um levantamento das idades dos alunos de uma de suas turmas do 3º ano do Ensino Médio, uma professora construiu a tabela de frequência ao lado:

Entre os valores assumidos pela variável “idade”: 16, 17, 18 ou 19, constatamos que o valor 17 ocorreu mais vezes: 28 dos 44 alunos têm 17 anos.

Dizemos, então, que a **moda** das idades nesse conjunto de dados é igual a 17.

A moda de uma relação de valores (indica-se  $M_o$ ) é o valor que ocorre mais vezes na relação, isto é, aquele que possui maior frequência absoluta.

Idades	Número de alunos
16	4
17	28
18	7
19	5

### EXEMPLO 4

Vamos encontrar a moda dos seguintes conjuntos de valores:

a) 5 — 8 — 11 — 8 — 3 — 4 — 8

A moda é  $M_o = 8$ , pois há três valores iguais a 8.

b) 2 — 3 — 9 — 3 — 4 — 2 — 6

Há duas modas: 2 e 3. Dizemos, então, que se trata de uma **distribuição de frequências bimodal**.

c) 1 — 3 — 4 — 6 — 9 — 11 — 2

Nesse caso, todos os valores aparecem com a mesma frequência unitária. Assim, não há moda nessa distribuição.

### OBSERVAÇÃO

Média, mediana e moda são as três medidas de tendência central mais usuais que podem ser associadas a um conjunto de dados. Cada uma delas possui, como vimos, interpretação e significado próprios. Dependendo da natureza dos dados, uma ou outra dessas medidas pode ser mais adequada para representá-los quantitativamente. Entretanto, a análise dos dados se torna mais completa quando conhecemos os valores das três medidas.



## EXERCÍCIO RESOLVIDO

- 2 Na tabela seguinte está representada a evolução do agronegócio no Brasil, expresso como porcentagem do Produto Interno Bruto (PIB), no período de 2003 a 2013.

**Agronegócio no Brasil**

Ano	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Porcentagem de participação em relação ao PIB brasileiro	26,3%	25,5%	23,6%	22,8%	23,2%	23,8%	22,5%	22,5%	23,1%	22,2%	22,5%

Fonte: Centro de Estudos Avançados em Economia Aplicada – Esalq/USP. Disponível em: <cepea.esalq.usp.br/pib/>. Acesso em: 15 mar. 2016.

Calcule as três medidas de centralidade associadas a esse conjunto de dados, interpretando-as.

### Solução:

- Média aritmética:

$$\frac{26,3 + 25,5 + 23,6 + 22,8 + 23,2 + 23,8 + 22,5 + 22,5 + 23,1 + 22,2 + 22,5}{11} = \frac{258,0}{11} \approx 23,45\%$$

O valor encontrado para a média indica o percentual teórico de participação do agronegócio no PIB caso todos os anos apresentassem o mesmo percentual.

- Moda:  $M_o = 22,5\%$

O valor encontrado para a moda significa que qualquer outro percentual da relação de valores se repete menos vezes que o valor 22,5%. Note que 22,5% aparece 3 vezes entre os dados.

- Mediana:

É preciso colocar os percentuais em ordem crescente:

$$22,2 — 22,5 — 22,5 — 22,5 — 22,8 — \boxed{23,1} — 23,2 — 23,6 — 23,8 — 25,5 — 26,3$$

Como o número de valores é ímpar ( $n = 11$ ), a mediana é igual ao valor da sexta posição da relação acima, isto é, a mediana é 23,1%.

Isso significa que, dos 11 percentuais listados, 5 são inferiores a 23,1% e 5 são superiores a 23,1%.



## EXERCÍCIOS

FAÇA NO  
CADERNO

- 29** Calcule a média ( $\bar{M}$ ), a mediana ( $Me$ ) e a moda ( $Mo$ ) para cada conjunto de valores:

- a) 2 — 2 — 3 — 3 — 3 — 4 — 4 — 4 — 4  
 b) 16 — 18 — 18 — 17 — 19 — 18  
 c) 1 — 5 — 3 — 2 — 4  
 d) 11 — 8 — 15 — 19 — 6 — 15 — 13 — 21  
 e) 44 — 43 — 42 — 43 — 45 — 44 — 40 — 41 — 49 — 46

- 30** Os dados ordenados abaixo referem-se ao tempo de espera (em minutos) de 10 pessoas que foram atendidas em um posto de saúde durante uma manhã:

$$1 — 5 — 8 — 9 — x — 16 \\ 18 — y — 23 — 26$$

Sabendo que o tempo médio de espera foi de 14 minutos e o tempo mediano foi de 15 minutos, determine os valores de  $x$  e de  $y$ .

- 31** Uma pesquisa realizada com 3 000 pessoas de certa região pretende levantar alguns aspectos socioeconômicos. Um dos itens do questionário era: "Qual é o número de banheiros em sua residência?" Os resultados encontram-se a seguir:

Número de banheiros	Porcentagem
1	42%
2	37%
3	16%
4	5%

- a) Qual é o número de entrevistados cuja residência possui até dois banheiros?  
 b) Calcule a média, a moda e a mediana para os dados coletados.

- 32** Na tabela seguinte constam os valores dos dez maiores PIBs das Américas em 2014.

**PIB nas Américas em 2014**

País	PIB (em bilhões de dólares)
Estados Unidos	17 420
Brasil	2 346
Canadá	1 785
México	1 295
Argentina	537,7
Venezuela	381,3
Colômbia	377,7
Chile	258,1
Peru	202,6
Equador	100,9

Fonte: Banco Mundial. Disponível em:

<[www.worldbank.org/en/country](http://www.worldbank.org/en/country)>. Acesso em: 15 mar. 2016.

- a) Calcule a média e a mediana dos dados apresentados. Por que a média é bem maior que a mediana?  
 b) Em que condição a média ficaria mais próxima da mediana? Faça os cálculos necessários.

**33** Uma empresa paga, todo ano, um bônus de fim de ano para seus funcionários. Neste ano, a empresa já pagou bônus a 40 dos seus 50 funcionários, como mostra a tabela seguinte:

**Bônus de fim de ano**

Número de funcionários	Valor do bônus (em reais)
8	300
14	600
18	1 000

Dados elaborados pelo autor.

- a) Qual é a mediana dos valores já pagos de bônus pela empresa?
- b) Sabe-se que os dez funcionários restantes receberão bônus de 600 ou 1 000 reais. Qual é o número de funcionários que devem receber bônus de R\$ 600,00 para que a mediana dos 50 valores seja R\$ 800,00?

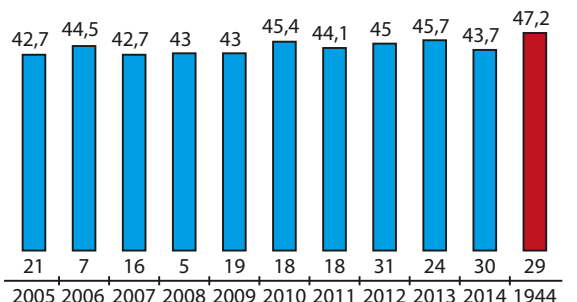
**34** Um corretor de imóveis relacionou, ao longo de dois anos de trabalho, a quantidade de imóveis comercializados (venda ou locação) mensalmente. Os resultados encontram-se a seguir:

Imóveis comercializados por mês	Número de meses
0	8
1	4
2	11
3	1

- a) Quais são os valores da média, da mediana e da moda da variável em questão?
- b) Se nos cinco meses subsequentes não for comercializado imóvel algum, qual será a mediana dos 29 valores?

**35** Em maio de 2015, uma onda de calor matou mais de mil pessoas na Índia. No gráfico seguinte, constam as temperaturas máximas registradas no mês de maio na capital Nova Délhi, além do recorde registrado em 29 de maio de 1944:

**Temperatura máxima registrada no mês de maio em diferentes anos em Nova Délhi (em °C)**

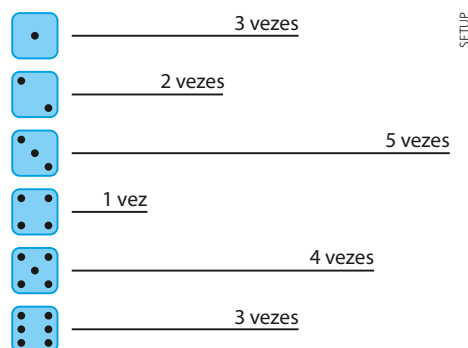


■ Recorde

Fonte: O Estado de S. Paulo, 28 maio 2015.

- a) Considerando a relação dos valores das temperaturas máximas de 2005 a 2014, determine a média, a moda e a mediana.
- b) Qual deveria ter sido a temperatura máxima registrada em maio de 2015 a fim de que a média das temperaturas máximas registradas nesses onze anos subisse 0,12 °C em relação à média do item a)?

**36** Um dado foi lançado 18 vezes consecutivas e os resultados obtidos encontram-se no esquema abaixo:



- a) Calcule a média, a mediana e a moda da distribuição de frequências das faces obtidas nos lançamentos.
- b) Decidiu-se lançar o dado mais 7 vezes. Sabendo que o 19º, 20º e o 21º lançamentos do dado resultaram, cada um, na face 4, é possível que a mediana da nova distribuição de frequências, considerando 25 lançamentos, seja 3,5? E 5? Em caso afirmativo, determine as condições para que isso seja possível.
- c) Nas condições do item b), existe a possibilidade de que a mediana da distribuição de frequências dos 25 lançamentos seja igual a 4? Em caso afirmativo, dê um exemplo de uma distribuição de frequências que satisfaz, explicitando as possibilidades para os quatro últimos lançamentos.

## ▶ Medidas de dispersão (ou variabilidade)

Procurando uma companhia aérea para as viagens de negócios dos funcionários de sua pequena empresa, um empresário obteve, na internet, os percentuais mensais de pontualidade dos voos de duas companhias aéreas, **A** e **B**, no período de sete meses anteriores à data de sua pesquisa. Os resultados encontram-se a seguir:

Companhia \ Mês	1	2	3	4	5	6	7
A	86%	92%	91%	95%	90%	89%	94%
B	93%	92%	90%	91%	90%	93%	88%

Inicialmente, ele calculou a média dos percentuais de pontualidade das duas companhias:

- companhia **A**:

$$\frac{86\% + 92\% + 91\% + 95\% + 90\% + 89\% + 94\%}{7} = \frac{637\%}{7} = 91\%$$

- companhia **B**:

$$\frac{93\% + 92\% + 90\% + 91\% + 90\% + 93\% + 88\%}{7} = \frac{637\%}{7} = 91\%$$

Conhecendo apenas o valor das médias das duas companhias, ele sabia que seria difícil optar por alguma. Surgiu, então, a ideia de saber qual companhia aérea apresentava desempenho mais regular, isto é, aquela cujos índices de pontualidade variassem menos.

Observando os valores da tabela, o empresário concluiu, sem muita convicção, que a companhia **B** era a mais regular.

A situação acima descrita mostra a necessidade de conhecermos as **medidas de dispersão** (ou variabilidade), que permitem quantificar a variabilidade de um conjunto de dados.

Vamos estabelecer quatro medidas de dispersão: a **amplitude**, a **variância**, o **desvio padrão** e o **desvio absoluto médio**.

### ▶ Amplitude

A **amplitude** de uma relação de valores assumidos por uma variável quantitativa é o número real dado pela diferença entre o maior e o menor valores registrados (nesta ordem).

No exemplo das companhias aéreas, temos:

- companhia **A**: A amplitude é igual a  $95\% - 86\% = 9\%$
- companhia **B**: A amplitude é igual a  $93\% - 88\% = 5\%$

Como a amplitude dos dados mensais da companhia **B** é menor que a amplitude dos dados mensais da companhia **A**, conclui-se que a companhia **B** é a mais regular, isto é, seus índices oscilaram menos do que os da outra companhia.

Definiremos agora três medidas de variabilidade que quantificam a dispersão (espalhamento) dos dados em relação à média.

ANTONIO CUNHA/CB.D.A PRESS



#### OBSERVAÇÃO

A amplitude de um conjunto de dados é uma medida de variabilidade fácil de ser calculada e, em geral, eficaz na análise da dispersão dos dados e na comparação da variabilidade de dois conjuntos de valores.

## ► Variância

Sejam  $x_1, x_2, \dots, x_n$  a relação de valores assumidos por uma variável quantitativa  $X$  e  $\bar{x}$  a média aritmética desses valores.

Para cada  $x_i$  ( $i = 1, 2, \dots, n$ ), calculamos o quadrado da diferença entre esse valor e a média, ou seja,  $(x_i - \bar{x})^2$ , que é chamado desvio quadrático.

A **variância** – indica-se por  $\text{Var } X$  ou  $\sigma^2$  (lê-se: sigma ao quadrado) – é a média aritmética dos desvios quadráticos, isto é:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

Usando a notação de somatório, escrevemos:

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

Observe que, em qualquer situação,  $\sigma^2$  é um número real não negativo, pois o numerador da expressão é uma soma de quadrados em  $\mathbb{R}$ .

Vamos calcular a variância dos dados referentes aos percentuais mensais de pontualidade das duas companhias (para facilitar, omitiremos nos cálculos os símbolos de %):

### Companhia A

Mês	$x_i$	$(x_i - \bar{x})^2$
1	86	$(86 - 91)^2$
2	92	$(92 - 91)^2$
3	91	$(91 - 91)^2$
4	95	$(95 - 91)^2$
5	90	$(90 - 91)^2$
6	89	$(89 - 91)^2$
7	94	$(94 - 91)^2$

$$\begin{aligned} \sigma_A^2 &= \frac{(86 - 91)^2 + (92 - 91)^2 + \dots + (94 - 91)^2}{7} \\ \sigma_A^2 &= \frac{(-5)^2 + 1^2 + 0^2 + 4^2 + (-1)^2 + (-2)^2 + 3^2}{7} = \\ &= \frac{25 + 1 + 0 + 16 + 1 + 4 + 9}{7} = \frac{56}{7} = 8 \end{aligned}$$

### Companhia B

Mês	$x_i$	$(x_i - \bar{x})^2$
1	93	$(93 - 91)^2$
2	92	$(92 - 91)^2$
3	90	$(90 - 91)^2$
4	91	$(91 - 91)^2$
5	90	$(90 - 91)^2$
6	93	$(93 - 91)^2$
7	88	$(88 - 91)^2$

$$\begin{aligned} \sigma_B^2 &= \frac{(93 - 91)^2 + (92 - 91)^2 + \dots + (88 - 91)^2}{7} \\ \sigma_B^2 &= \frac{2^2 + 1^2 + (-1)^2 + 0^2 + (-1)^2 + 2^2 + (-3)^2}{7} = \\ &= \frac{4 + 1 + 1 + 0 + 1 + 4 + 9}{7} = \frac{20}{7} \approx 2,86 \end{aligned}$$

Como a variância em **B** é menor que a variância em **A** ( $\sigma_B^2 = 2,86 < \sigma_A^2 = 8$ ), concluímos que a companhia **B** é mais regular, isto é, os percentuais mensais de pontualidade de **B** estão menos dispersos em relação à média do que os percentuais mensais de **A**.



### PENSE NISTO:

O que aconteceria com o valor da variância se na definição, no lugar dos desvios quadráticos, usássemos apenas os desvios  $(x_i - \bar{x})$ , isto é, se tivéssemos  $\frac{(x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x})}{n}$ ?

## OBSERVAÇÃO

A variância é definida como uma soma de quadrados (média dos desvios quadráticos), sendo, portanto, uma medida cuja unidade é quadrática. Por exemplo, se estivéssemos estudando a altura dos alunos de uma turma, a altura média seria expressa em metros (m), porém a variância seria expressa em metros ao quadrado ( $m^2$ ), o que geraria uma incompatibilidade em relação às unidades, pois m é a unidade de medida de comprimento e  $m^2$  é unidade de área. Para uniformizá-las, definiremos o desvio padrão.

## ► Desvio padrão

Seja  $x_1, x_2, \dots, x_n$  a relação dos valores assumidos por uma variável  $X$ . Chamamos **desvio padrão** de  $X$ , indica-se por  $DP(X)$  ou  $\sigma$ , a raiz quadrada da variância de  $X$ :

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Vamos calcular o desvio padrão dos percentuais de pontualidade das companhias aéreas:

- companhia **A**:  $\sigma_A^2 = 8 \Rightarrow \sigma_A = \sqrt{8} \approx 2,82$
- companhia **B**:  $\sigma_B^2 = 2,86 \Rightarrow \sigma_B = \sqrt{2,86} \approx 1,69$



## EXERCÍCIO RESOLVIDO

- 3 Num levantamento realizado em 100 jogos de futebol de um torneio foram colhidos os seguintes dados:

Gols por partida	0	1	2	3	4	5
Frequência de jogos	28	26	31	9	4	2

Calcule o desvio padrão do número de gols marcados por partida.

### Solução:

É importante lembrar que a variável em estudo é "número de gols marcados por partida", e ela assume os valores 0, 1, 2, 3, 4 e 5 com frequências absolutas respectivamente iguais a 28, 26, 31, 9, 4 e 2.

Calculemos a média ( $\bar{x}$ ) desses valores. Temos a seguinte média de gols/partida:

$$\bar{x} = \frac{0 \cdot 28 + 1 \cdot 26 + 2 \cdot 31 + 3 \cdot 9 + 4 \cdot 4 + 5 \cdot 2}{28 + 26 + 31 + 9 + 4 + 2} = \frac{141}{100} = 1,41$$

Calculemos, agora, os desvios quadráticos de cada valor que a variável assume com relação à média:

$$\underbrace{(0 - 1,41)^2}_{1,9881}; \underbrace{(1 - 1,41)^2}_{0,1681}; \underbrace{(2 - 1,41)^2}_{0,3481}; \underbrace{(3 - 1,41)^2}_{2,5281}; \underbrace{(4 - 1,41)^2}_{6,7081} \text{ e } \underbrace{(5 - 1,41)^2}_{12,8881}$$

Levando em conta as frequências absolutas, temos que a variância ( $\sigma^2$ ) é dada por:

$$\sigma^2 = \frac{28 \cdot 1,9881 + 26 \cdot 0,1681 + 31 \cdot 0,3481 + 9 \cdot 2,5281 + 4 \cdot 6,7081 + 2 \cdot 12,8881}{100}$$

$$\sigma^2 \approx \frac{55,67 + 4,37 + 10,79 + 22,75 + 26,83 + 25,77}{100} = \frac{146,18}{100} = 1,4618$$

Daí, o desvio padrão é  $\sigma = \sqrt{1,4618} \approx 1,21$  gol/partida.



## EXERCÍCIOS

**37** Para cada conjunto de valores, calcule a amplitude (a), a variância ( $\sigma^2$ ) e o desvio padrão ( $\sigma$ ):

- a) 3 — 3 — 4 — 4 — 4 — 6  
 b) 1 — 2 — 3 — 4 — 5  
 c) 15 — 22 — 18 — 20 — 21 — 23 — 14  
 d) 31 — 31 — 31 — 31 — 31 — 31 — 31 — 31  
 e) 5 — 6 — 6 — 7 — 7 — 7 — 8 — 8 — 8 — 8

**38** Um grupo de 12 estudantes passou um dia de verão em um parque aquático. Seus gastos com alimentação são dados a seguir (valores em reais):

24 — 16 — 30 — 20 — 28 — 30 — 20 —  
 40 — 18 — 16 — 30 — 16

A partir dos valores dados, obtenha:

- a) a variância;                      c) a amplitude.  
 b) o desvio padrão;
- 39** A quantidade de erros de digitação por página de uma pesquisa escolar com quarenta páginas é dada a seguir:

Erro por página	0	1	2
Número de páginas	28	8	4

- a) Determine as medidas de centralidade (média, mediana e moda) correspondentes à quantidade de erros.  
 b) Determine as medidas de dispersão (variância e desvio padrão) correspondentes.
- 40** Um professor de inglês está interessado em comparar o desempenho de suas quatro turmas de um mesmo curso. Para isso, considerou a média final dos cinco alunos de cada turma:

Turma A	3	5	7	5	5
Turma B	6	6	4	4	5
Turma C	9	1	6	5	4
Turma D	7	8	5	2	3

- a) Calcule a amplitude das notas de cada turma e use esses valores para ordená-las, da mais regular à menos regular.  
 b) Compare os desvios padrões das turmas C e D, indicando aquela com aproveitamento mais regular dos alunos.  
 c) Use o desvio padrão para comparar as turmas A e B quanto à regularidade das notas finais dos alunos.

**41** Nas tabelas a seguir estão representadas as taxas de analfabetismo, expressas em porcentagem, dos estados das regiões Centro-Oeste e Sudeste em 2013:

## Taxas de analfabetismo

Região Centro-Oeste	
Distrito Federal	3,2%
Goiás	7,1%
Mato Grosso	7,8%
Mato Grosso do Sul	7,2%

Região Sudeste	
São Paulo	3,7%
Rio de Janeiro	3,7%
Minas Gerais	7,6%
Espírito Santo	6,6%

Fonte: IBGE, 2013.

Qual região apresenta o conjunto de valores mais homogêneo? Calcule o desvio padrão das taxas de analfabetismo nas duas regiões para comprovar sua resposta.

**42** No quadro seguinte constam as notas que Pedro e Paulo tiraram nas cinco avaliações que fizeram em um curso de informática.

Pedro	7,0	4,5	5,5	5,0	3,0
Paulo	5,0	5,5	3,0	4,0	7,5

Calcule a variância das notas de cada aluno, indicando qual deles obteve desempenho mais homogêneo.

**43** Os salários dos 20 funcionários que trabalham em um certo setor de uma indústria estão apresentados na tabela:

## Salários da indústria

Salários (em reais)	Número de funcionários
1 200,00	10
1 440,00	6
2 400,00	4

Dados elaborados pelo autor.



- a) Calcule a média ( $\bar{x}$ ) e o desvio padrão ( $\sigma$ ) dos salários.
- b) Suponha que sejam contratados cinco funcionários, cada um com salário de R\$ 1 500,00. A média salarial dos 25 funcionários que trabalharão neste setor irá aumentar, diminuir ou permanecer constante? Por quê?

- 44 Os valores seguintes representam os resultados obtidos por 12 estudantes em um experimento cujo objetivo era calcular, em milímetros, uma determinada distância:

$$8,7 - 8,5 - 9,2 - 8,8 - 8,9 - 8,6 - 8,7 - 8,6 - 8,4 - 8,7 - 8,6 - 8,7$$

O professor considerou aceitáveis os resultados pertencentes ao intervalo  $[\bar{x} - \sigma, \bar{x} + \sigma]$ , em que  $\bar{x}$  é a média e  $\sigma$  o desvio padrão dos dados acima.

Quantos alunos não tiveram seu resultado considerado aceitável?

- 45 Para preencher uma vaga de emprego, os 5 candidatos finalistas foram submetidos a 3 provas (Redação, Raciocínio quantitativo e Inglês) e uma dinâmica de grupo. As notas obtidas pelos candidatos encontram-se no quadro seguinte:

	Redação	Dinâmica	Raciocínio quantitativo	Inglês
<b>Candidato A</b>	6,5	6,0	8,0	7,0
<b>Candidato B</b>	7,0	7,5	5,0	7,0
<b>Candidato C</b>	7,0	7,0	4,5	5,0
<b>Candidato D</b>	6,0	7,0	6,0	6,0
<b>Candidato E</b>	7,5	7,5	4,0	8,5

A escolha do candidato obedecerá ao seguinte critério:

- 1ª) O candidato deverá apresentar nota média maior ou igual a 6,5, considerando as provas e a dinâmica.
- 2ª) Se houver mais de um candidato com nota média maior ou igual a 6,5, será selecionado o que apresentar melhor desempenho na dinâmica.
- 3ª) Persistindo o empate, será escolhido o candidato que apresentar desempenho mais regular nas três provas.

Qual foi o candidato selecionado para essa vaga?

- 46 Na tabela seguinte constam os percentuais de atraso nos voos de uma companhia aérea, no primeiro semestre de certo ano, segundo os destinos:

**Percentual de atraso em voos**

Voos nacionais	Voos internacionais com até 5 horas de duração	Voos internacionais com mais de 5 horas de duração
12%	8%	7%

Dados elaborados pelo autor.

Preocupada com os índices apresentados, a companhia fez uma campanha para diminuir os percentuais de atraso nos voos em cada um dos três destinos. Os índices projetados para o segundo semestre são de 9% para voos nacionais e 4% para voos internacionais com até 5 horas de duração. Qual deverá ser o valor projetado para o percentual de atraso nos voos internacionais com mais de 5 horas de duração a fim de que a redução seja atingida, e o desvio padrão dos três índices do 2º semestre seja igual ao desvio padrão dos 3 índices do 1º semestre?

## ► Desvio médio

Outra medida de dispersão, geralmente menos usual que a variância (e o desvio padrão), é o **desvio médio absoluto**, ou, simplesmente, **desvio médio**.

Sejam  $x_1, x_2, \dots, x_n$  os valores assumidos por uma variável quantitativa  $X$  e  $\bar{x}$  a média aritmética desses valores. O desvio médio (indicaremos por DM) é definido por:

$$DM = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{1}{n} \cdot \sum_{i=1}^n |x_i - \bar{x}|$$

Observe que DM representa a média aritmética dos desvios entre  $x_i$  ( $i = 1, 2, \dots, n$ ) e  $\bar{x}$ , expressos pelo módulo da diferença entre um valor que a variável assume ( $x_i$ ) e a média ( $\bar{x}$ ).

### EXEMPLO 5

As temperaturas mínimas diárias de duas cidades da serra catarinense, em uma mesma semana de inverno, estão representadas abaixo.

	Segunda	Terça	Quarta	Quinta	Sexta
Cidade A	-1 °C	0 °C	2 °C	-1 °C	-3 °C
Cidade B	-1 °C	-2 °C	1 °C	1 °C	-2 °C

Vamos usar o desvio médio para decidir em qual cidade as temperaturas oscilaram menos em relação à média semanal.

- Para a cidade **A**, temos:

$$\bar{x} = \frac{-1 + 0 + 2 + (-1) + (-3)}{5} = -\frac{3}{5} = -0,6$$

$$DM = \frac{|-1 - (-0,6)| + |0 - (-0,6)| + |2 - (-0,6)| + |-1 - (-0,6)| + |-3 - (-0,6)|}{5}$$

$$DM = \frac{|-0,4| + |0,6| + |2,6| + |-0,4| + |-2,4|}{5} = \frac{0,4 + 0,6 + 2,6 + 0,4 + 2,4}{5} = 1,28$$

- Para a cidade **B**, temos:

$$\bar{x} = \frac{(-1) + (-2) + 1 + 1 + (-2)}{5} = -\frac{3}{5} = -0,6$$

$$DM = \frac{|-1 - (-0,6)| + |-2 - (-0,6)| + |1 - (-0,6)| + |1 - (-0,6)| + |-2 - (-0,6)|}{5}$$

$$DM = \frac{|-0,4| + |-1,4| + |1,6| + |1,6| + |-1,4|}{5} = \frac{0,4 + 1,4 + 1,6 + 1,6 + 1,4}{5} = 1,28$$

Coincidentemente, as duas cidades apresentaram igual temperatura mínima média e igual desvio médio, o que nos mostra que não houve diferença na média das dispersões das temperaturas (em torno da média) para as duas cidades, considerando o desvio médio como medida.

Para "desempatar", poderíamos, por exemplo, calcular a amplitude das temperaturas nas duas cidades: cidade **A**:  $2\text{ °C} - (-3\text{ °C}) = 5\text{ °C}$ ; cidade **B**:  $1\text{ °C} - (-2\text{ °C}) = 3\text{ °C}$ . Concluiríamos, desse modo, que em **B** os dados estão menos dispersos (ou variam menos), formando um conjunto mais homogêneo.



## EXERCÍCIOS



**47** Calcule o desvio médio para cada um dos seguintes conjuntos de dados:

- a) 2 — 4 — 6  
 b) 2 — 3 — 5 — 4 — 8 — 8  
 c) 20 — 25 — 15 — 35 — 30

**48** Um país hipotético é formado por duas regiões, **A** e **B**, cada uma com cinco cidades de mesma população. Foi feito um levantamento para saber o grau de satisfação da população de cada cidade, em relação à respectiva administração regional. No quadro, constam notas de 0 a 10 para medir a satisfação dos habitantes:

<b>Região A</b>	7,0	4,5	5,5	5,0	3,0
<b>Região B</b>	5,0	8,5	3,0	1,0	7,5

Calcule o desvio médio absoluto para cada região, determinando em qual delas as opiniões são menos divergentes.

**49** Na tabela seguinte, encontram-se os valores de uma gratificação dada aos funcionários de uma pequena empresa no fim do ano:

**Gratificação de fim de ano**

Gratificação (em reais)	Número de funcionários
200,00	8
450,00	12
800,00	5
1 500,00	3
2 500,00	2

Dados elaborados pelo autor.

Calcule o desvio médio das gratificações recebidas, arredondando os valores obtidos para o número inteiro mais próximo.

## ▶ Medidas de centralidade e dispersão para dados agrupados

Os salários dos 23 funcionários de um estabelecimento comercial estão representados na tabela seguinte:

**Salários**

Salários (em reais)	Número de funcionários
1 000 — 2 500	4
2 500 — 4 000	12
4 000 — 5 500	7

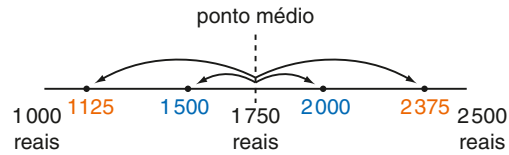
Dados elaborados pelo autor.

Qual é a média salarial dos funcionários deste estabelecimento?

Quando a variável em estudo apresenta seus valores agrupados em classes ou intervalos, não dispomos de informações para saber como esses valores estão distribuídos em cada faixa.

Para que se possa calcular a média (e outras medidas de centralidade e dispersão) desses valores, costuma-se fazer a suposição de que, em cada intervalo, os valores estão distribuídos de forma simétrica em relação ao **ponto médio** (indicado por  $x_i$ ) do intervalo. Ao considerar, por exemplo, o primeiro intervalo, 1 000 — 2 500 (cujo ponto médio é  $\frac{1\,000 + 2\,500}{2} = 1\,750$ ), uma possível distribuição simétrica dos quatro salários é dada a seguir.

Salários: R\$ 1 125,00; R\$ 1 500,00; R\$ 2 000,00 e R\$ 2 375,00



Observe que há uma compensação entre valores equidistantes dos extremos do intervalo (1 000 e 2 500), de modo que a média de cada par desses valores coincide com o ponto médio  $x_i$ . Vejamos: 1 125 e 2 375 são equidistantes dos extremos do intervalo; a média entre eles é  $\frac{1\,125 + 2\,375}{2} = 1\,750 = x_i$ .

O mesmo raciocínio se aplica ao par de valores 1 500 e 2 000.

Na prática, essa suposição é equivalente a admitir que todos os quatro salários desse intervalo são iguais a R\$ 1 750,00, que é o ponto médio.

Estendendo esse raciocínio aos demais intervalos, é possível calcular a média salarial dos funcionários desse estabelecimento:

Salários (em reais)	Ponto médio do intervalo	Número de funcionários
1 000 – 2 500	1 750	4
2 500 – 4 000	3 250	12
4 000 – 5 500	4 750	7

$$\bar{x} = \frac{1\,750 \cdot 4 + 3\,250 \cdot 12 + 4\,750 \cdot 7}{4 + 12 + 7} = \frac{79\,250}{23} \approx 3\,445,65$$

Arredondando para o inteiro mais próximo, temos que a média salarial dos funcionários do estabelecimento é R\$ 3 446,00.

## ► Cálculo do desvio padrão

O cálculo das medidas de dispersão (variância e desvio padrão) em relação à média está apoiado na mesma suposição: dentro de cada intervalo, os valores distribuem-se de forma simétrica em torno do ponto médio ( $x_i$ ). Na prática, admitimos que todos os valores do intervalo coincidem com  $x_i$ .

Salários (em reais)	Ponto médio do intervalo ( $x_i$ )	Desvio quadrático ( $x_i - \bar{x}$ ) <sup>2</sup>	Número de funcionários
1 000 – 2 500	1 750	(1 750 – 3 446) <sup>2</sup>	4
2 500 – 4 000	3 250	(3 250 – 3 446) <sup>2</sup>	12
4 000 – 5 500	4 750	(4 750 – 3 446) <sup>2</sup>	7

A variância ( $\sigma^2$ ), como sabemos, é a média aritmética desses desvios, ponderada pelas respectivas frequências absolutas (número de funcionários) de cada intervalo:

$$\sigma^2 = \frac{4 \cdot (1\,750 - 3\,446)^2 + 12 \cdot (3\,250 - 3\,446)^2 + 7 \cdot (4\,750 - 3\,446)^2}{23}$$

$$\sigma^2 = \frac{11\,505\,664 + 460\,992 + 11\,902\,912}{23} = \frac{23\,869\,568}{23} \approx 1\,037\,807$$

Desse modo, a variância dos salários é 1 037 807 (reais)<sup>2</sup> e o desvio padrão dos salários é:

$$\sigma = \sqrt{1\,037\,807 \text{ (reais)}^2} \approx 1\,018,73 \text{ reais}$$

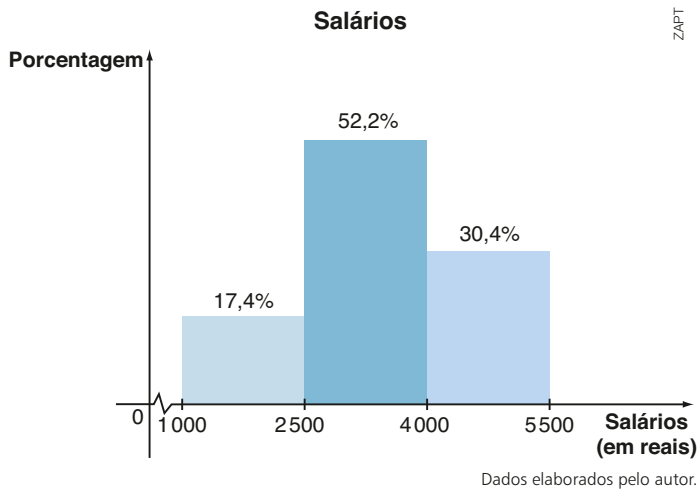
## ► Determinação da classe modal

Definimos **classe modal** como a classe que apresenta maior frequência absoluta. No exemplo, a classe modal é 2 500–4 000, pois há 12 valores percentuais a esse intervalo (as outras frequências são menores: 4 e 7).

## ► Cálculo da mediana

Lembremos, inicialmente, que a mediana de uma relação de valores é um valor que separa essa relação em duas partes com a mesma quantidade de valores, sendo que, em uma das partes, todos os valores são menores ou iguais à mediana e, na outra parte, todos os valores são maiores ou iguais à mediana.

Observe, no histograma seguinte, as porcentagens aproximadas de cada intervalo:



Da leitura do gráfico, notamos que:

- ao final do primeiro intervalo encontram-se 17,4% do total de valores;
- ao final dos dois primeiros intervalos, encontram-se acumulados 69,6% do total de valores (17,4% + 52,2% = 69,6%).

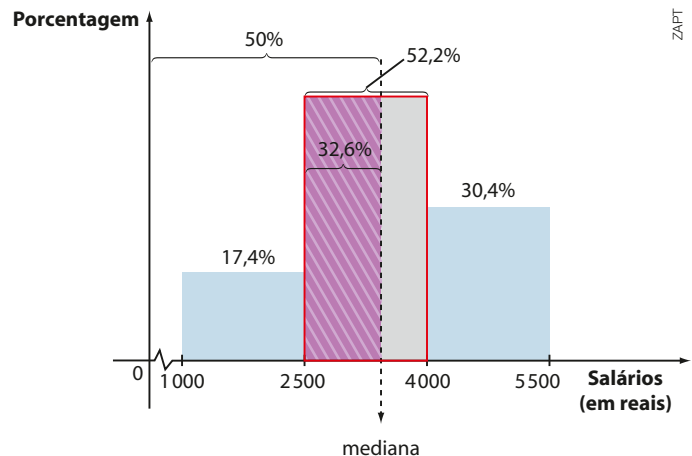
Com base nas observações anteriores, concluímos que a mediana se encontra no segundo intervalo. Do limite inferior do primeiro intervalo (1 000) até a mediana concentram-se 50% do total de valores (17,4% + 32,6% = 50%).

Observando que, no segundo intervalo, o retângulo roxo e o retângulo destacado com fio vermelho possuem a mesma altura, temos que a área de cada um desses retângulos (expressa como porcentagem da área total sob o histograma) é diretamente proporcional à medida de sua base, isto é:

$$\frac{Me - 2500}{32,6\%} = \frac{4000 - 2500}{52,2\%} \Rightarrow$$

$$\Rightarrow Me \approx 3436,78$$

Logo, a mediana é aproximadamente 3 436,78 reais.





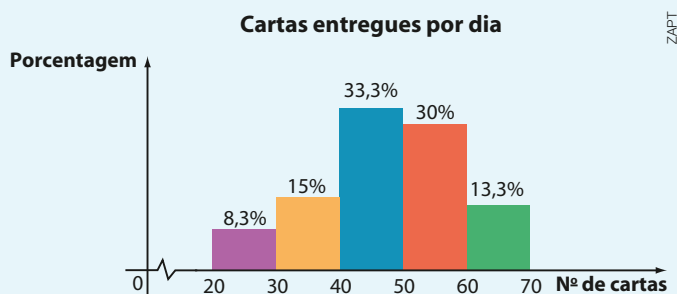
## EXERCÍCIO RESOLVIDO

- 4 Durante 60 dias, anotou-se o número de cartas entregues, diariamente, em um edifício residencial. Os resultados são mostrados na tabela e no histograma seguintes.

**Cartas entregues por dia**

Número de cartas	Frequência absoluta	Porcentagem (%)
20 – 30	5	8,3
30 – 40	9	15
40 – 50	20	33,3
50 – 60	18	30
60 – 70	8	13,3

Dados elaborados pelo autor.



Determine as três medidas de centralidade correspondentes ao número de cartas diariamente entregues no edifício.

### Solução:

- Média ( $\bar{x}$ ):

Usando os pontos médios de cada intervalo, temos:

$$\bar{x} = \frac{25 \cdot 5 + 35 \cdot 9 + 45 \cdot 20 + 55 \cdot 18 + 65 \cdot 8}{5 + 9 + 20 + 18 + 8} = \frac{2850}{60} = 47,5$$

- Mediana (Me):

A mediana encontra-se na 3ª classe (de 40 a 50), pois a porcentagem acumulada ao final das três primeiras classes já ultrapassa 50% ( $8,3\% + 15\% + 33,3\% = 56,6\%$ ). Ao final das duas primeiras classes, a porcentagem de observações acumulada é de  $8,3\% + 15\% = 23,3\%$ . Isso mostra que do valor 40 até a mediana estão concentradas  $50\% - 23,3\% = 26,7\%$  das observações. Podemos então estabelecer a seguinte proporção:

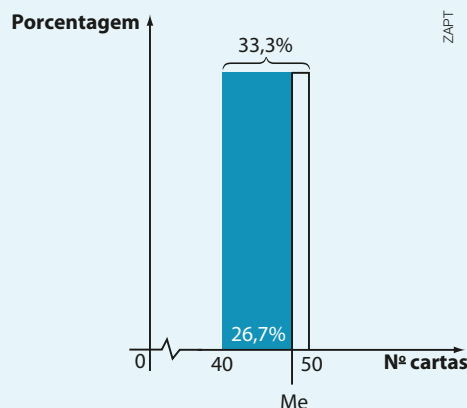
$$\frac{50 - 40}{33,3\%} = \frac{Me - 40}{26,7\%} \Rightarrow Me \approx 48$$

|
└

porcentagem referente ao intervalo 40 – 50
porcentagem referente ao intervalo 40 – Me (retângulo colorido)

- Classe modal:

A classe modal corresponde ao intervalo que reúne a maior porcentagem de valores. Nesse caso, a classe modal é o intervalo  $[40; 50[$ .

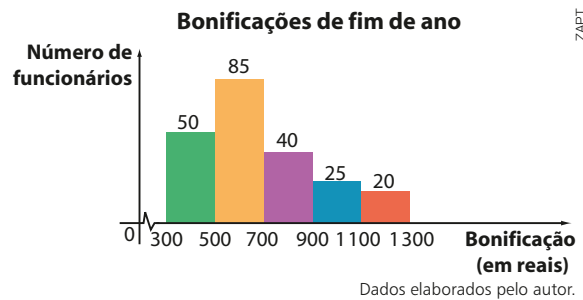




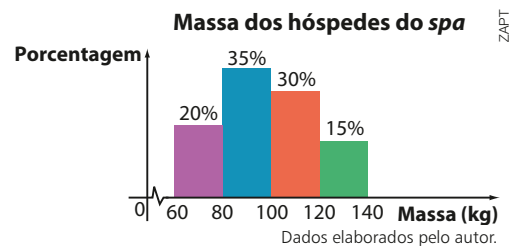
## EXERCÍCIOS



- 50** No gráfico ao lado está representada a distribuição de bonificações de fim de ano concedidas aos funcionários de um estabelecimento comercial.
- Qual é o número de funcionários do estabelecimento?
  - Qual é o valor médio das bonificações?
  - Qual é a classe modal das bonificações?
  - Escolhido ao acaso um funcionário desse estabelecimento, qual é a probabilidade de que ele tenha recebido ao menos R\$ 700,00 de bonificação?



- 51** No histograma ao lado estão representadas as massas de 200 clientes que se hospedaram durante uma semana em um *spa*:
- Quantos hóspedes tinham menos de 120 kg?
  - Qual a massa média desses clientes?
  - Qual a massa mediana desses clientes?
  - Qual o desvio padrão das massas dessa distribuição?



- 52** Na tabela seguinte constam os valores doados em um dia, na arrecadação de uma campanha beneficente:

**Campanha beneficente**

Doações (em reais)	Número de doações
5 a 15	1 250
16 a 26	1 083
27 a 37	762
38 a 48	541
49 a 59	509
60 a 70	321

Dados elaborados pelo autor.

- Qual é a menor quantia que pôde ter sido arrecadada nesse dia?
  - Qual é a maior quantia que pôde ter sido arrecadada nesse dia?
- 53** Uma companhia de ônibus registrou a taxa de ocupação (em %) de suas viagens entre Goiânia e Brasília durante 40 dias. Os resultados são mostrados a seguir:
- 43 — 66 — 54 — 75 — 78 — 61 — 48 — 50 — 53 — 60 — 60 — 86 — 61 — 60 — 55 — 62 —  
 45 — 57 — 61 — 40 — 32 — 49 — 52 — 48 — 69 — 70 — 68 — 80 — 82 — 79 — 39 — 48 —  
 84 — 76 — 36 — 61 — 91 — 81 — 65 — 55
- Considerando classes de amplitude 10 a partir de 32%, construa um histograma correspondente.
  - Calcule a taxa média, a taxa mediana e a classe modal de ocupação, considerando os dados agrupados.
  - Qual é o desvio padrão da taxa de ocupação, considerando os dados agrupados?

- 54 As temperaturas máximas diárias registradas no mês de janeiro em uma cidade estão dadas na tabela seguinte:

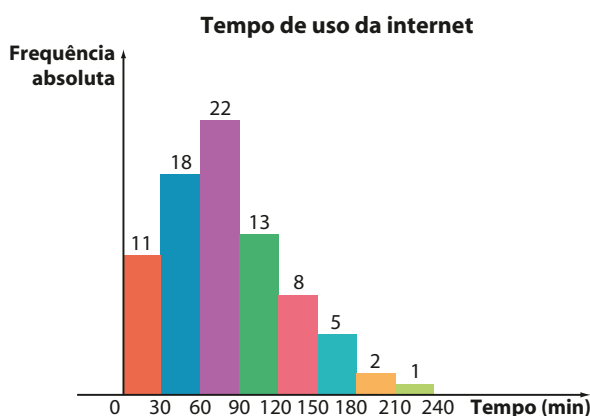
**Temperaturas máximas diárias  
registradas em janeiro**

Temperatura máxima	Número de dias
25 °C – 28 °C	9
28 °C – 31 °C	11
31 °C – 34 °C	7
34 °C – 37 °C	4
<b>Total</b>	<b>31</b>

Dados elaborados pelo autor.

Determine:

- a) a média, a mediana e a classe modal das temperaturas;  
 b) a variância e o desvio padrão das temperaturas.
- 55 Um provedor de internet mediu o tempo (em minutos) de uso diário da rede por seus assinantes, por meio de uma amostragem. Com os dados obtidos na pesquisa construiu-se o seguinte histograma:



Dados elaborados pelo autor.

- a) Que porcentagem do total de assinantes fica entre meia hora e uma hora e meia na rede?  
 b) Qual é a média e a mediana do tempo de uso da internet?  
 c) A partir do histograma anterior, faça um outro histograma agrupando os tempos de hora em hora. Calcule média, mediana e desvio padrão.



## DESAFIO

Seja  $x_1, x_2, \dots, x_n$  uma relação de dados numéricos correspondentes aos valores assumidos por uma variável quantitativa. Explique o que acontece com a média ( $\bar{x}$ ), a variância ( $\sigma^2$ ) e o desvio padrão ( $\sigma$ ) desses valores se cada  $x_i, i \in \{1, 2, \dots, n\}$  é:

- a) aumentado em duas unidades;  
 b) multiplicado por 2;  
 c) reduzido em 20%.