GELSON IEZZI
SAMUEL HAZZAN
DAVID DEGENSZAJN

FUNDAMENTOS DE MATEMÁTICA ELEMENTAR

Matemática comercial
Matemática financeira
Estatística descritiva

11

Atual Editora

NOVAS QUESTÕES DE VESTIBULARES

CAPÍTULO III

Estatística Descritiva

I. Introdução

Imagine que, um mês antes de uma eleição presidencial, a federação das indústrias de determinado estado encomendou a um instituto especializado uma pesquisa cujo objetivo consistiu em detectar a intenção de voto do eleitor e levantar o perfil socioeconômico dos eleitores de cada um dos candidatos.

O que o instituto fez?

- Primeiramente, dimensionou uma amostra da população e fez a coleta de dados por meio de uma pesquisa de campo. A escolha da amostra é, em geral, complexa, pois deve-se levar em conta, entre outros fatores, o tempo e o custo da pesquisa, o número de eleitores de cada cidade do estado, a camada social à qual o entrevistado pertence, o local onde será feita a entrevista. É imprescindível que a amostra seja representativa, a fim de não haver comprometimento na análise dos resultados.
- Num segundo momento, organizou em tabelas os dados brutos coletados, construiu gráficos para apresentar os resultados obtidos e divulgou-os nos meios de comunicação. É preciso também associar ao conjunto de informações medidas de tendência **central** e medidas de **variabilidade** (ou dispersão dos dados em relação a valores centrais).
- Por fim, fez a análise confirmatória dos dados, isto é, verificou a margem de erro com que os resultados da amostra refletiram, de fato, a intenção de votos de toda a população de eleitores.

A ciência que se dedica a esse trabalho é a **Estatística**.

Os levantamentos estatísticos costumam ser divulgados em jornais, revistas, televisão, Internet, etc. e quase sempre têm relação direta com a vida das pessoas, pois envolvem assuntos como saúde, comportamento, bem-estar e desenvolvimento humano, economia, demografia, pesquisas de mercado, educação, entre muitos outros.

Sobre as etapas mencionadas, podemos dizer que a primeira diz respeito às técnicas de **Amostragem**, a segunda compete à **Estatística Descritiva** e a última é objeto de estudo da **Inferência Estatística**.

A Estatística Descritiva é utilizada também para se organizar e resumir informações relativas a uma população inteira, como ocorre, por exemplo, nos censos demográficos efetuados pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Neste capítulo, nos ocuparemos do estudo de aspectos relacionados à Estatística Descritiva.

II. Variável

A entidade representativa dos moradores de um bairro queria traçar um perfil dos frequentadores de um parque ali situado. Uma equipe de pesquisa de rua, contratada para realizar o trabalho, elaborou questões a fim de reunir as informações procuradas. Numa manhã de quarta-feira, 20 pessoas foram entrevistadas e cada uma respondeu a questões para identificar sexo, idade (arredondada para o inteiro mais próximo), número de vezes que frequenta o parque por semana, estado civil, meio de transporte utilizado para chegar ao parque, tempo de permanência no parque e renda familiar mensal. Os resultados são mostrados na tabela a seguir:

Sexo	Idade	Frequência semanal	Estado civil	Meio de transporte	Tempo de permanência	Renda familiar mensal (em salários mínimos)
Masculino	26	2	casado	carro	30 min	13,3
Masculino	23	1	solteiro	ônibus	35 min	11,8
Feminino	41	5	viúva	a pé	2h50min	8,9
Masculino	49	3	separado	a pé	45 min	13,9
Feminino	19	5	solteira	carro	1 h	11,6
Feminino	20	4	solteira	a pé	1h20min	16,0
Masculino	27	3	solteiro	carro	45 min	19,5
Masculino	38	3	casado	a pé	2h15min	9,3
Masculino	27	2	separado	ônibus	1h30min	10,2

Sexo	Idade	Frequência semanal	Estado civil	Meio de transporte	Tempo de permanência	Renda familiar mensal (em salários mínimos)
Feminino	50	7	casada	a pé	45 min	12,4
Masculino	52	2	solteiro	a pé	1h40min	10,7
Feminino	48	4	casada	a pé	1h15min	14,7
Masculino	28	4	casado	a pé	1 h	16,6
Masculino	36	1	casado	carro	1h30min	12,5
Feminino	31	3	solteira	ônibus	2 h	8,2
Masculino	56	3	viúvo	a pé	30 min	15,4
Feminino	41	6	solteira	carro	2h30min	18,8
Masculino	44	1	casado	ônibus	50 min	12,1
Feminino	29	2	separada	a pé	40 min	5,0
Masculino	31	3	casado	ônibus	2h45min	7,6

Cada um dos aspectos investigados – os quais permitirão fazer a análise desejada – é denominado **variável**.

Algumas variáveis, como sexo, estado civil e meio de transporte utilizado para chegar ao parque, apresentam como resposta um atributo, qualidade ou preferência do entrevistado. Variáveis dessa natureza recebem o nome de **variáveis qualitativas**. Se considerarmos, por exemplo, a variável meio de transporte utilizado, dizemos que **carro**, **ônibus** e **a pé** correspondem às realizações ou valores assumidos por essa variável.

Outras variáveis, como idade, frequência semanal, tempo de permanência e renda familiar mensal, apresentam como resposta um número. Variáveis desse tipo são denominadas **variáveis quantitativas**. Podemos classificá-las em dois grupos:

- **Variáveis quantitativas discretas**: são aquelas cujos valores são obtidos por **contagem** e representados por elementos de um conjunto finito ou enumerável. No exemplo, a variável frequência semanal é discreta, e seus valores são 1, 2, 3, 4, 5, 6 ou 7.
- **Variáveis quantitativas contínuas**: são aquelas cujos valores são obtidos por **mensuração** e representados por valores pertencentes a um intervalo **real**. As variáveis idade, tempo de permanência e renda familiar mensal são contínuas e seus valores se distribuem em determinado intervalo real. A variável tempo de permanência, por exemplo, tem seus valores (em horas) pertencentes ao intervalo [0,5; 3[.

EXERCÍCIOS

255. Ao se cadastrar em um *site* de comércio eletrônico, o usuário deve preencher um questionário com estas oito perguntas:

1. Você tem computador em casa?
2. Quantas vezes por semana você acessa a Internet?
3. Numa escala de zero a 10, qual seu índice de confiança na segurança do comércio eletrônico?
4. Quantos cartões de crédito você possui?
5. A residência em que vive é própria ou alugada?
6. Qual é o provedor que você utiliza para acessar a rede?
7. Qual é o tempo médio de acesso à Internet?
8. Já comprou algum produto via Internet?

Cada uma das questões anteriores define uma variável. Classifique-as como qualitativas ou quantitativas.

256. Num cursinho pré-vestibular, os estudantes inscritos responderam a um questionário no qual constavam, entre outras, as seguintes questões:

1. Qual é a área da carreira universitária pretendida?
2. Você cursou o ensino médio em escola particular ou pública?
3. Qual é a renda familiar mensal?
4. Qual é o grau de escolaridade do chefe da família?
5. Qual é a sua disciplina favorita?
6. Quantas vezes você já fez cursinho?
7. Você é usuário da Internet?
8. Quanto tempo de estudo diário pretende dedicar ao cursinho?

Em relação às variáveis definidas pelas questões acima, responda:

a) Quantas são classificadas como qualitativas?
b) Dê três possíveis realizações da variável definida pela questão 4.

257. Uma pesquisa realizada na plataforma de embarque de um terminal rodoviário tinha como objetivo conhecer o perfil do usuário dos fins de semana. Os 200 entrevistados responderam às seguintes questões:

1. Qual seu estado civil?
2. Você possui veículo próprio?
3. Quantas vezes por mês você utiliza este terminal?

4. Qual é a principal razão desta viagem: lazer, negócios ou visita à família?
5. Qual é, aproximadamente, o tempo de viagem até o destino final?
6. Em relação aos serviços deste terminal, você está: satisfeito, parcialmente satisfeito ou insatisfeito?
7. Qual é a quantia mensal que você costuma gastar neste terminal (incluindo passagens, alimentação, entretenimento, etc.)?

Classifique cada uma das variáveis determinadas por essas questões.

III. Tabelas de frequência

A simples leitura dos dados brutos da tabela anteriormente apresentada não nos fornece as condições necessárias à determinação do perfil do frequentador do parque, uma vez que as informações não estão devidamente organizadas.

O primeiro procedimento que possibilita uma leitura mais resumida dos dados é a construção de tabelas de frequência.

Para cada variável estudada, contamos o número de vezes que ocorre cada um de seus valores (ou realizações). O número obtido é chamado **frequência absoluta** e é indicado por n_i (cada valor assumido pela variável aparece um determinado número de vezes, o que justifica o uso do índice *i*). Vejamos:

Dos 20 entrevistados, encontramos os seguintes resultados para a frequência absoluta dos valores assumidos pela variável estado civil:

- separado ($n_1 = 3$);
- solteiro ($n_2 = 7$);
- casado ($n_3 = 8$);
- viúvo ($n_4 = 2$).

Note que:

$$n_1 + n_2 + n_3 + n_4 = \sum_{i=1}^{4} n_i = 20$$

Em geral, quando os resultados de uma pesquisa (ou estudo) são divulgados em jornais e revistas, os valores referentes à frequência absoluta aparecem acompanhados do **número total** de valores colhidos, a fim de tornar a análise mais significativa.

Poderíamos, por exemplo, repetir a pesquisa do parque algum tempo depois e construir uma amostra com 30 entrevistados em vez dos 20 participantes da pesquisa inicial. Para comparar os resultados obtidos nas duas amostras seria preciso levar em conta que elas têm "tamanhos" diferentes.

Definimos, então, para cada valor assumido por uma variável, a **frequência relativa** (f_i) como a razão entre a frequência absoluta (n_i) e o número total de dados (*n*), isto é:

$$\boxed{f_i = \frac{n_i}{n}}$$

Observações:

1ª) Como $n_i \leqslant n$, segue que, para cada *i*, $0 \leqslant f_i \leqslant 1$. Por esse motivo, é comum a frequência relativa ser expressa em porcentagem.

2ª) A soma das frequências relativas dos valores assumidos por determinada variável é sempre igual a 1.

De fato:

$$\sum_i f_i = \sum_i \frac{n_i}{n} = \frac{1}{n}\sum_i n_i = \frac{1}{n} \cdot n = 1$$

Exemplo:

Para a variável estado civil da tabela anteriormente apresentada, construímos a seguinte tabela de frequência:

Estado civil	Frequência absoluta (n_i)	Frequência relativa (f_i)	Porcentagem (%)
Separado	3	$\frac{3}{20} = 0{,}15$	15
Solteiro	7	$\frac{7}{20} = 0{,}35$	35
Casado	8	$\frac{8}{20} = 0{,}40$	40
Viúvo	2	$\frac{2}{20} = 0{,}10$	10
Total	**20**	**1,0**	**100**

A construção das tabelas de frequência para as variáveis sexo, frequência semanal de visita ao parque e meio de transporte utilizado é análoga.

Em alguns casos, porém, pode ocorrer que os valores assumidos por uma variável pertençam a determinado intervalo real, não havendo praticamente repetição (coincidência) de valores. Isso ocorre com as variáveis idade, tempo de permanência no parque e renda familiar mensal. Esta última tem seus valores variando no intervalo [5, 20[. Nesse caso, construímos uma tabela de frequência em que os dados estarão agrupados em classes (ou intervalos) de valores.

Observações:

1ª) Vamos convencionar que cada intervalo construído é fechado à esquerda e aberto à direita, isto é, a notação $a \vdash b$ refere-se ao intervalo real [a, b[, que inclui *a* e não inclui *b*, isto é:

$$[a, b[= \{x \in \mathbb{R} \mid a \leqslant x < b\}$$

2ª) A amplitude do intervalo a ⊢ b é dada pela diferença b – a. (No exemplo que será fornecido a seguir, a amplitude de cada uma das classes da renda familiar é igual a 3.)

3ª) Não há regras fixas para a construção dos intervalos usados para agrupar as informações a partir dos dados brutos. Dependendo da natureza dos dados, podemos ter um número maior ou menor de classes. Recomenda-se, no entanto, sempre que possível, construir classes de **mesma amplitude**. Além disso, convém evitar classes de amplitude muito grande ou muito pequena, a fim de que a análise não fique comprometida.

Exemplos:

1º) Considerando a variável renda mensal familiar, é possível agrupar os dados brutos nas seguintes classes:

Renda familiar mensal (em salários mínimos)	**Frequência absoluta (n_i)**	**Frequência relativa (f_i)**	**Porcentagem (%)**
5 ⊢ 8	2	$\frac{2}{20} = 0{,}10$	10
8 ⊢ 11	5	$\frac{5}{20} = 0{,}25$	25
11 ⊢ 14	7	$\frac{7}{20} = 0{,}35$	35
14 ⊢ 17	4	$\frac{4}{20} = 0{,}20$	20
17 ⊢ 20	2	$\frac{2}{20} = 0{,}10$	10
Total	**20**	**1,0**	**100**

2º) Para a variável tempo de permanência no parque, construímos uma tabela de frequência em que as informações estão agrupadas em intervalos de amplitude igual a 30.

Tempo de permanência (em minutos)	**Frequência absoluta (n_i)**	**Frequência relativa (f_i)**	**Porcentagem (%)**
30 ⊢ 60	8	$\frac{8}{20} = 0{,}40$	40
60 ⊢ 90	4	$\frac{4}{20} = 0{,}20$	20

Tempo de permanência (em minutos)	Frequência absoluta (n_i)	Frequência relativa (f_i)	Porcentagem (%)
90 ⊢ 120	3	$\frac{3}{20} = 0{,}15$	15
120 ⊢ 150	2	$\frac{2}{20} = 0{,}10$	10
150 ⊢ 180	3	$\frac{3}{20} = 0{,}15$	15
Total	**20**	**1,0**	**100**

A leitura da tabela permite concluir que:
- a maioria (60% dos entrevistados) permanece menos de 90 minutos no parque;
- três em cada quatro entrevistados ficam no parque por menos de duas horas (note que 40% + 20% + 15% = 75%).

EXERCÍCIOS

Os exercícios 258 a 260 referem-se à situação da tabela apresentada nas páginas 79 e 80.

258. Construa uma tabela de frequência para a variável sexo.

259. Construa uma tabela de frequência para a variável frequência semanal de visita ao parque.

260. Com os dados referentes à idade agrupados em classes de intervalo, cada um com amplitude igual a 10, construa uma tabela de frequência.

261. Em uma pesquisa socioeconômica sobre itens de conforto, perguntou-se a cada um dos 800 entrevistados: Quantos aparelhos de TV em cores há em sua casa?

Os resultados aparecem na tabela:

Nº de aparelhos	Frequência absoluta	Frequência relativa	Porcentagem (%)
0	20	▲	▲
1	▲	▲	▲
2	▲	0,6	▲
3	▲	▲	7,5
4	30	▲	▲

a) Complete a tabela.

b) Suponha que levantamentos posteriores mostraram que os resultados dessa amostra representam, em termos da frequência relativa, a distribuição do número de aparelhos de TV de toda a população. No universo de 680 000 domicílios, qual o número daqueles em que há exatamente 1 aparelho?

262. Os dados seguintes referem-se ao tempo de espera (em minutos) de 30 clientes em uma fila de banco, em um dia de grande movimento:

23 — 19 — 7 — 21 — 16 — 13 — 11 — 16 — 33 — 22

17 — 15 — 12 — 18 — 25 — 20 — 14 — 16 — 12 — 10

8 — 20 — 16 — 14 — 19 — 23 — 36 — 30 — 28 — 35

Construa uma tabela de frequência, agrupando as informações em classes de amplitude igual a 5, a partir do menor tempo encontrado.

263. A tabela abaixo informa os tipos de lazer preferidos por 80 garotos da 1ª série do ensino médio de um colégio.

Lazer	Frequência absoluta	Frequência relativa
Jogar futebol com os amigos	48	a
Computador e *videogame*	b	c
Paquerar no *shopping*	d	e
Viajar para a praia	f	g
Total	**80**	**1,00**

Complete a tabela, sabendo que c é o dobro de e, que é o quíntuplo de g.

264. Vinte e cinco jovens de até 15 anos foram selecionados para participar de um programa desenvolvido pela Secretaria de Esportes de uma cidade cujo objetivo consiste na formação de futuros jogadores de vôlei. As alturas dos jovens (em metro) são dadas a seguir:

1,82 — 1,77 — 1,79 — 1,74 — 1,73 — 1,81 — 1,82 — 1,69 — 1,71

1,78 — 1,78 — 1,88 — 1,72 — 1,65 — 1,75 — 1,78 — 1,73

1,82 — 1,84 — 1,74 — 1,76 — 1,79 — 1,83 — 1,76 — 1,70

a) A partir da menor altura encontrada, agrupe os dados em classes de amplitude 5 cm e faça a tabela de frequência correspondente.

b) Em visita ao centro de treinamento, um técnico estrangeiro sugeriu que pelo menos 48% dos jovens deveriam ter estatura superior ou igual a 1,80 m. Quantos jovens nessas condições devem ser incorporados ao atual grupo, de acordo com tal sugestão? Use os dados agrupados no item *a*.

265. A tabela seguinte informa os valores de 160 empréstimos solicitados a um banco por pessoas físicas durante uma semana.

Valor do empréstimo (em R$)	Frequência absoluta	Frequência relativa
200 ⊢ 400	*a*	*b*
400 ⊢ 600	60	*c*
600 ⊢ 800	*d*	*e*
800 ⊢ 1 000	*f*	0,05
1 000 ⊢ 1 200	*g*	*h*
Total	**160**	**1,00**

Complete a tabela, sabendo que 52,5% dos empréstimos representavam valores maiores ou iguais a R$ 600,00 e que, entre eles, $\frac{2}{3}$ eram inferiores a R$ 800,00.

266. (UF-GO) A tabela abaixo foi extraída da Pesquisa Nacional por Amostra de Domicílio/2001 do IBGE. Ela mostra as classes de rendimento mensal no Estado de Goiás e o número de pessoas de 10 anos ou mais de idade em cada classe.

Classe de rendimento mensal	Pessoas de 10 anos ou mais de idade		
	Total	Homens	Mulheres
Total	4 141 696	2 005 447	2 136 249
Até $\frac{1}{2}$ salário mínimo	210 438	62 010	148 428
Mais de $\frac{1}{2}$ a 1 salário mínimo	696 875	299 431	397 444
Mais de 1 a 2 salários mínimos	816 385	498 301	318 084
Mais de 2 a 3 salários mínimos	354 673	251 875	102 798
Mais de 3 a 5 salários mínimos	257 695	172 865	84 830
Mais de 5 a 10 salários mínimos	186 355	125 954	60 401
Mais de 10 a 20 salários mínimos	75 830	55 911	19 919
Mais de 20 salários mínimos	41 446	33 409	8 037
Sem rendimento	1 501 999	505 691	996 308

Analise essa tabela e julgue os itens a seguir:

1) O número de pessoas que ganham mais de 5 salários mínimos é inferior a 8% do total de pessoas.
2) A razão entre o número de mulheres e de homens que ganham até 1 salário mínimo é maior que a razão entre o número de mulheres e de homens com rendimento superior a 1 salário mínimo.

3) Mais de 60% das pessoas sem rendimento são mulheres.
4) Mais da metade das pessoas não possui rendimento ou ganha até 1 salário mínimo.

IV. Representação gráfica

Os gráficos constituem um importante instrumento de análise e interpretação de um conjunto de dados.

Diariamente é possível encontrar representações gráficas nos mais variados veículos de comunicação (jornais, revistas, televisão, Internet), associadas a assuntos diversos do nosso dia a dia, como resultados de pesquisas de opinião, saúde e desenvolvimento humano, economia, esportes, cidadania, etc.

A importância dos gráficos está ligada sobretudo à facilidade e rapidez na absorção e interpretação das informações por parte do leitor e também às inúmeras possibilidades de ilustração e resumo dos dados apresentados.

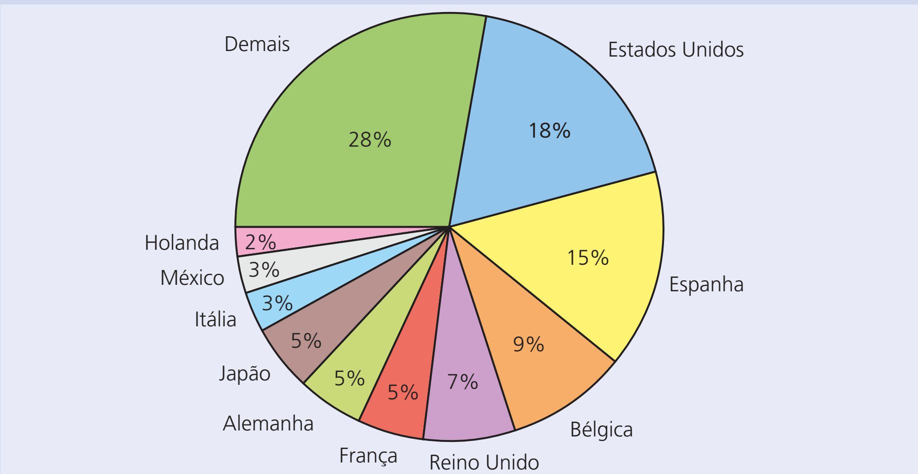
Estudaremos, neste capítulo, quatro tipos de representações gráficas: o gráfico de setores (ou "*pizza*"), o gráfico de barras (verticais ou horizontais), o histograma e o gráfico de linhas (poligonal).

Em relação aos tipos de gráficos citados, trabalharemos nos itens de V a VIII com sua construção, leitura e interpretação. É importante destacar que esses gráficos podem ser feitos utilizando-se planilhas ou *softwares* estatísticos.

Gráfico de setores

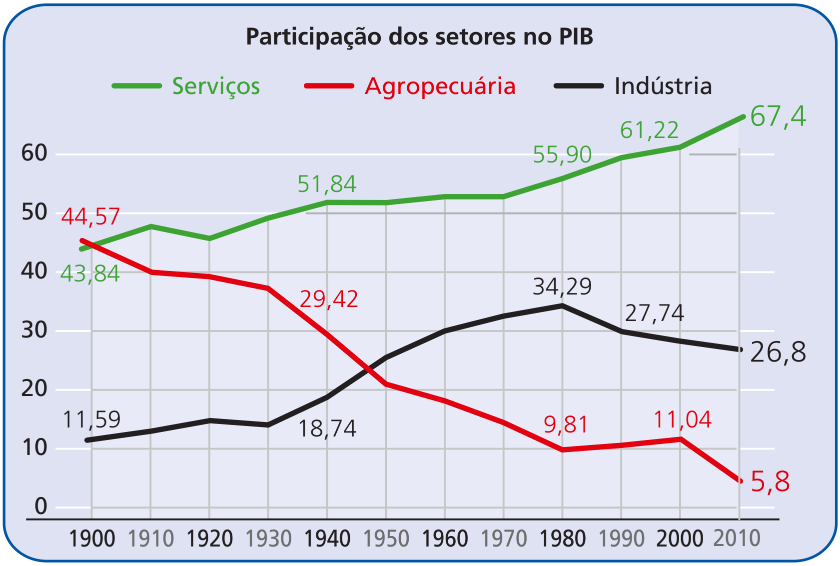
IED por país (%) investimentos estrangeiros diretos no Brasil

(Censo de Capitais Estrangeiros no país em 2011, ano-base 2010)



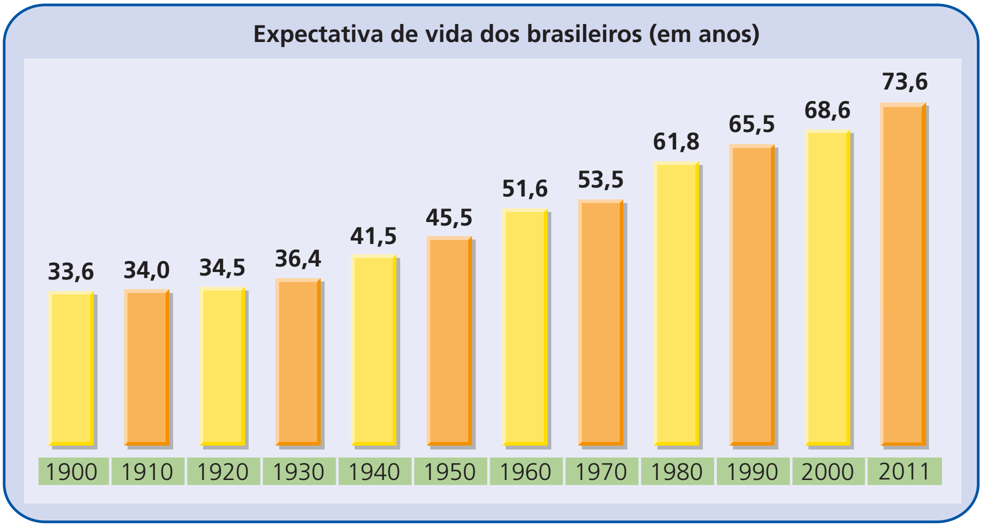
Fonte: Banco Central do Brasil e Ministério da Fazenda. Disponível em: <www.fazenda.gov.br>. Acesso em: 2 abr. 2013.

Gráfico poligonal



Fonte: *Almanaque Abril*, 2012.

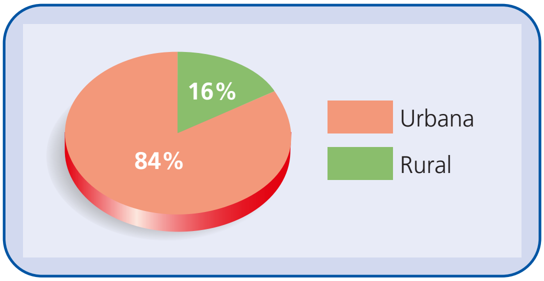
Gráfico de barras



Fonte: *Almanaque Abril*, 2012.

V. Gráfico de setores

O gráfico seguinte informa a distribuição da população brasileira que vive no campo (zona rural) e nas cidades (zona urbana).



Fonte: *Almanaque Abril*, 2012.

Para representar essa distribuição, dividimos um círculo em duas partes (setores circulares), uma com ângulo de medida proporcional à porcentagem da população rural e outra com ângulo de medida proporcional à porcentagem da população urbana.

Temos, então, a seguinte proporção:

- população rural:

$$\left.\begin{array}{lcl} 100\% & \text{———} & 360° \\ 16\% & \text{———} & x \end{array}\right\} \quad x = 57{,}6° = 57°36'$$

- população urbana:

$$\left.\begin{array}{lcl} 100\% & \text{———} & 360° \\ 84\% & \text{———} & y \end{array}\right\} \quad y = 302{,}4° = 302°24'$$

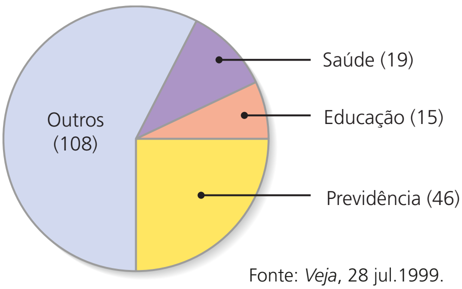
(Poderíamos simplesmente fazer também: $360° - 57{,}6° = 302{,}4°$.)

Com o auxílio de um transferidor, construímos o gráfico acima, que é chamado de **gráfico de setores** ou de "*pizza*".

De modo geral, quando uma variável assume *k* valores distintos, dividimos um círculo em *k* setores circulares cujas medidas dos ângulos são proporcionais às frequências correspondentes a cada um desses valores.

EXERCÍCIOS

267. (Vunesp-SP) O gráfico, publicado pela revista *Veja* de 28/7/1999, mostra como são divididos os 188 bilhões de reais do orçamento da União entre os setores de saúde, educação, previdência e outros.



Fonte: *Veja*, 28 jul.1999.

Se os 46 bilhões de reais gastos com a previdência fossem totalmente repassados aos demais setores, de modo que 50% fossem destinados à saúde, 40% à educação e os 10% restantes aos outros, determine o aumento que o setor de saúde teria:

a) em reais;

b) em porcentagem, em relação à sua dotação inicial, aproximadamente.

268. Uma pesquisa realizada com 800 pessoas às vésperas de um feriado prolongado tinha como pergunta principal: “O que você pretende fazer nestes quatro dias?”. Os resultados são dados na tabela seguinte:

Intenção	Número de pessoas
Descansar em casa	240
Viajar	360
Passear na própria cidade	160
Trabalhar	40

Faça um gráfico de setores para representar esses resultados.

269. Os gráficos seguintes mostram a disposição dos alunos das turmas da 3ª série do ensino médio para fazer cursinho pré-vestibular paralelamente a frequentar as aulas do colégio.

Turmas da manhã

Turmas da tarde

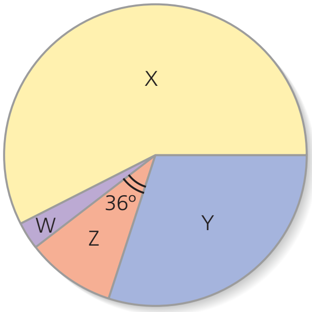
Sabendo que as turmas da manhã contam com 340 alunos e as da tarde com 280 alunos, determine:

a) o número total de alunos que não pretendem fazer cursinho;

b) a diferença entre o número de alunos do vespertino e do matutino que pretendam fazer cursinho o ano inteiro.

270. Em uma cidade, o mercado de leite é disputado por quatro marcas: X, Y, Z e W. Os resultados de uma sondagem a propósito da marca preferida, realizada com 400 consumidores, estão parcialmente apresentados na tabela e no gráfico seguintes.

Marca de preferência	Número de pessoas
X	230
Y	120
Z	▲
W	▲

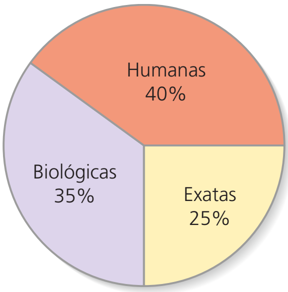


Determine:

a) a diferença entre o número de consumidores que preferem Z a W;

b) a diferença entre os ângulos correspondentes a X e Y.

271. Uma psicóloga realizou com os alunos da 1ª série do ensino médio de um colégio um estudo sobre orientação profissional. Após algumas dinâmicas e entrevistas, condensou as informações sobre a intenção de carreira dos alunos no gráfico abaixo.

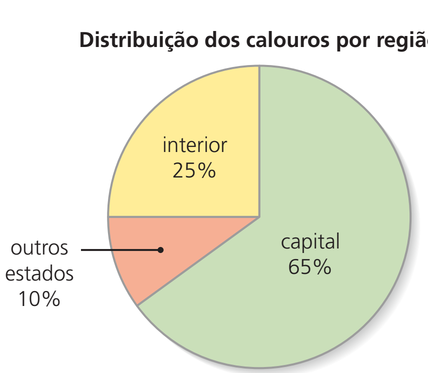


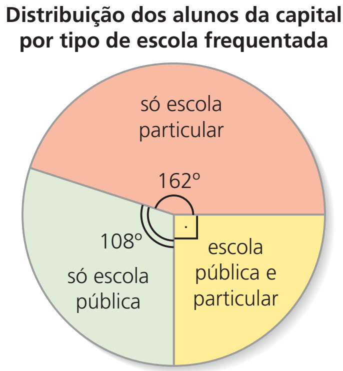
Quando os mesmos alunos estavam na 3ª série, a psicóloga repetiu o estudo com eles e notou que, em relação à sondagem anterior, $\frac{5}{16}$ dos interessados em Humanas migraram para Exatas e $\frac{3}{40}$ para Biológicas. Admitindo que não haja outras migrações:

a) construa o novo gráfico de setores correspondente, destacando os ângulos;

b) determine quantos alunos migraram de Humanas para Exatas, sabendo que o número dos participantes da dinâmica foi 400.

272. Uma universidade realizou um levantamento sobre a origem dos 4 800 novos alunos ingressantes. Os dados encontram-se resumidos nestes gráficos:





Com base nos gráficos, responda:

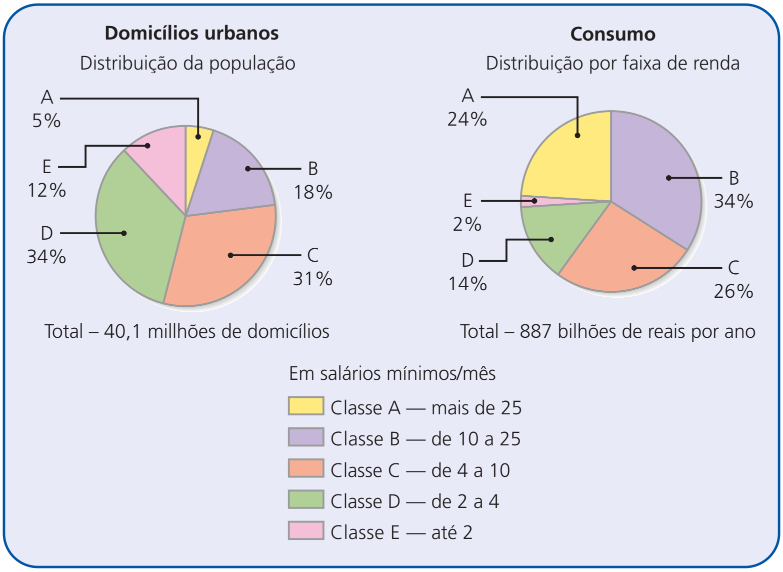
a) Qual é o número de calouros procedentes do interior?

b) Qual é o número de alunos da capital que estudaram nos dois tipos de escola (pública e particular)?

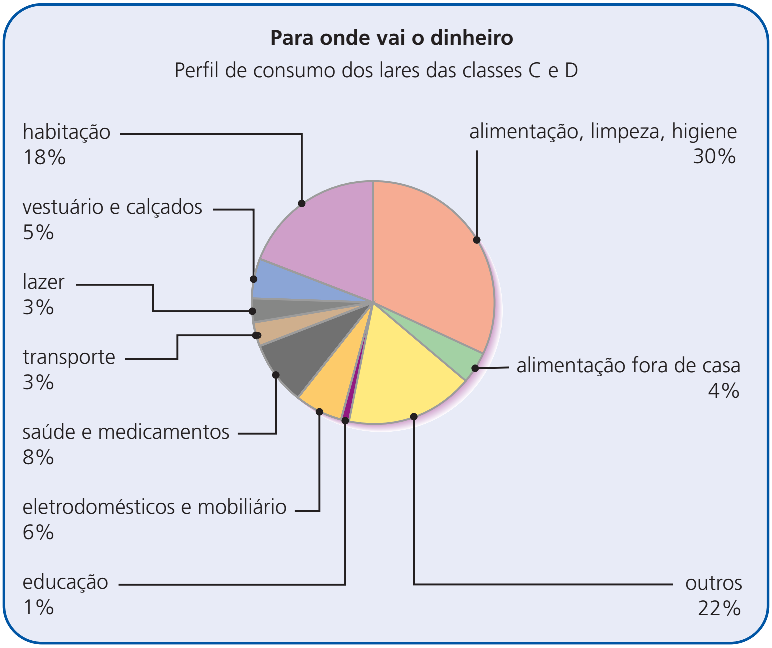
c) Qual é a porcentagem de calouros que estudaram apenas em escolas particulares da capital?

d) Qual é o número de calouros que já frequentaram a escola pública na capital?

273. Observe os gráficos a seguir e faça o que se pede.



Fonte: *Exame*, 1º out. 2003.



Fonte: *Exame*, 1 out. 2003.

a) Complete as afirmações corretamente:

1) As classes A e B juntas, embora representem apenas ▲% do total de domicílios urbanos, detêm ▲% do total consumido por ano pelos brasileiros.

2) O número de domicílios urbanos das classes D e E reunidos é da ordem de ▲ milhões.

3) A classe C é composta de aproximadamente ▲ milhões de domicílios urbanos e está representada no gráfico por um setor de ▲ graus. O consumo correspondente a essa classe gira em torno de ▲ bilhões de reais por ano.

b) Em relação ao consumo das classes C e D, assinale V (verdadeiro) ou F (falso) em cada item e justifique a classificação:

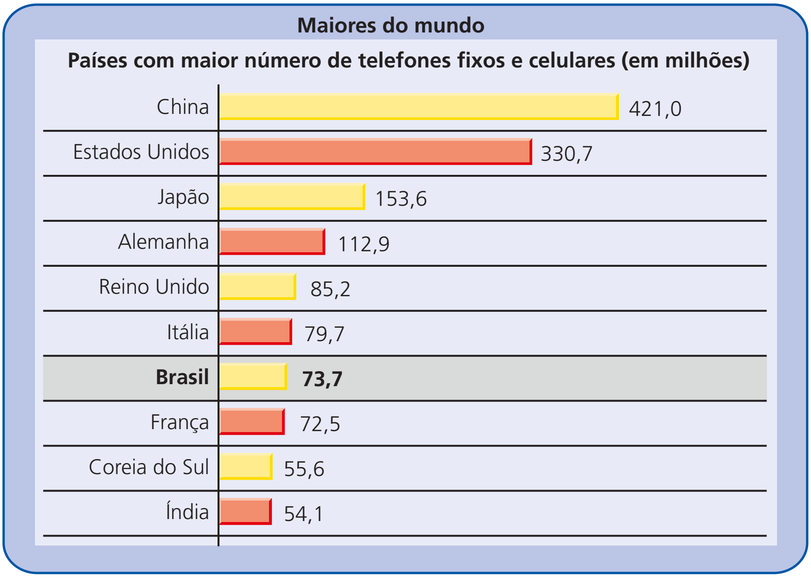
1) Alimentação, limpeza e higiene movimentam mais de 100 bilhões de reais por ano.

2) O total de gastos com saúde e medicamentos supera os 30 bilhões de reais por ano.

3) Os gastos com lazer de um único domicílio dessas classes são da ordem de 410 reais por ano.

VI. Gráfico de barras

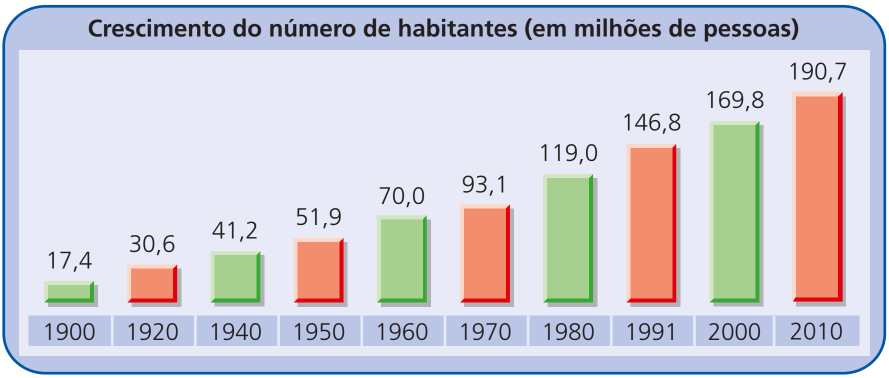
O gráfico abaixo relaciona os países onde há maior número de telefones (fixos e celulares, somados) e as quantidades correspondentes a cada um.



Fonte: *O Estado de S. Paulo*, 6 jul. 2003.

Ao lado do nome de cada país há uma barra cujo comprimento é **proporcional** ao número de telefones. Nessa escala, cada centímetro equivale a aproximadamente 70 milhões de telefones.

Esse tipo de gráfico recebe o nome de **gráfico de barras horizontais**. Para construí-lo, basta estabelecer uma escala conveniente para definir o tamanho da barra a ser usada para representar a frequência de cada ocorrência da variável em estudo.



Fonte: IBGE.

No gráfico anterior está representado o aumento da população brasileira em um século.

A cada ano corresponde uma coluna cujo comprimento é proporcional ao número de habitantes. Na escala utilizada, cada meio centímetro equivale a aproximadamente 35 milhões de habitantes. Esse tipo de gráfico é chamado de **gráfico de barras verticais**.

EXERCÍCIOS

274. O funcionário da bilheteria de um estádio de futebol classificou durante quinze minutos os torcedores que compareceram ao jogo segundo o critério: pagante (P), convidado (C) e menor com acompanhante (M).

Os dados brutos são apresentados a seguir:

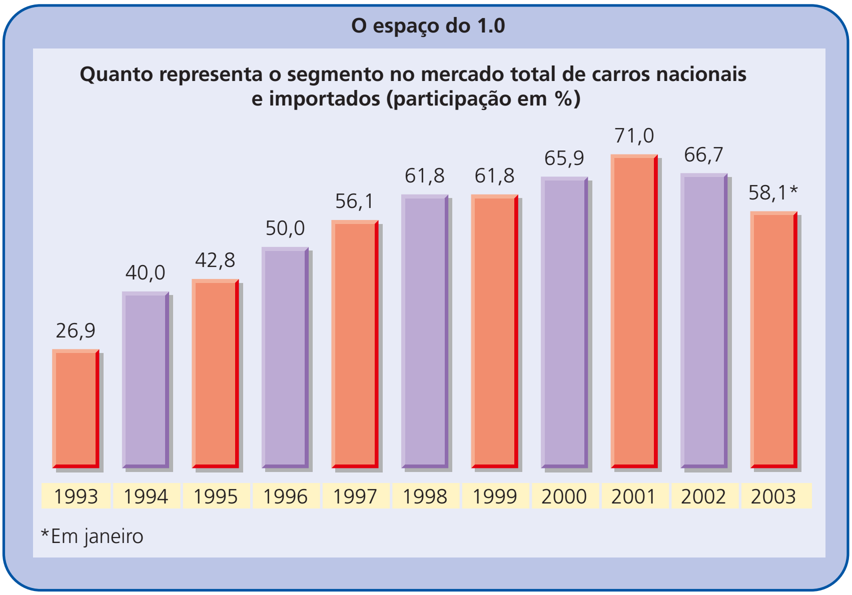
P — P — P — P — C — P — M — M — M — P — P — P — P — P

C — P — P — P — M — M — C — M — P — P — P — P — C — P

C — C — P — P — P — P — M — C — C — P — P — P

Faça um gráfico de barras horizontais para representar a distribuição percentual do público registrado pelo funcionário.

275.



Fonte: *O Estado de S. Paulo*, 2 mar. 2003.

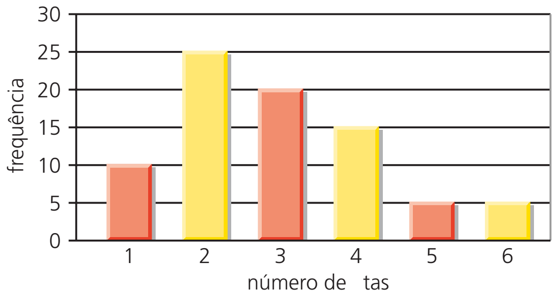
Considerando o gráfico, assinale V (verdadeira) ou F (falsa) nas afirmações, justificando as falsas:

a) A participação dos carros populares vem sempre aumentando desde 1993.

b) Sabe-se que em 1999 foram comercializados, no Brasil, 1,2 milhão de veículos. Assim, o número de carros populares comercializados no país foi 650 mil.

c) A participação dos carros populares correspondeu, no mínimo, à metade dos veículos comercializados nesse período, exceto nos três primeiros anos.

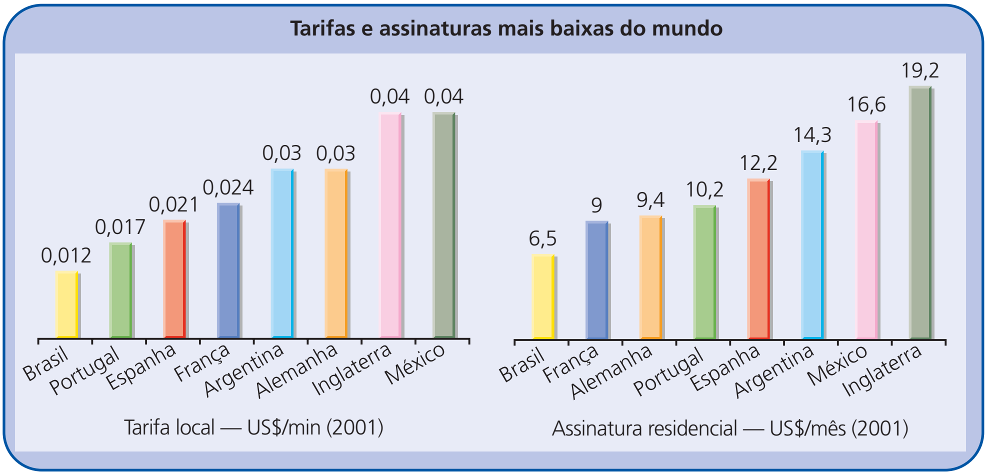
276. (FGV-SP) No gráfico abaixo está representado, no eixo das abscissas, o número de fitas de vídeo alugadas por semana numa videolocadora e, no eixo das ordenadas, a correspondente frequência (isto é, a quantidade de pessoas que alugaram o correspondente número de fitas).



a) Qual a porcentagem de pessoas que alugaram 4 ou mais fitas?

b) Se cada fita é alugada por R$ 4,00, qual a receita semanal da videolocadora?

277. Considerando os dois gráficos apresentados, responda às perguntas a seguir:



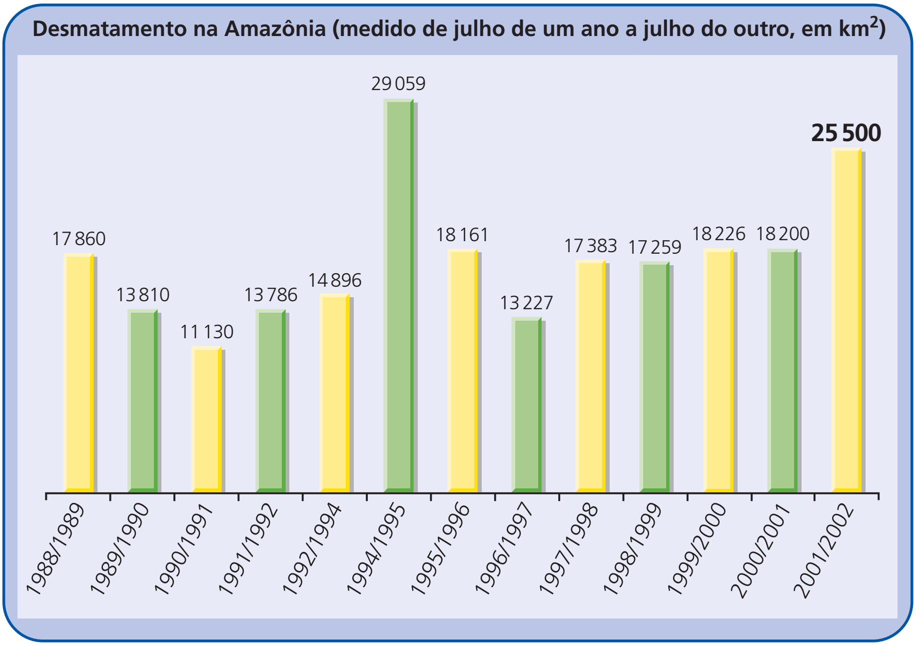
Fonte: *O Estado de S. Paulo*, 26 maio 2003.

a) Onde é mais caro falar ao telefone?

b) Qual é o valor da conta telefônica de um usuário que conversou 1 hora em um mês na França? E na Argentina?

c) Caso a tarifa telefônica local no Brasil aumente 150% e a da França não sofra alterações, a partir de quantos minutos mensais de uso de telefone a conta no Brasil passa a ser mais cara que a conta na França?

278. Cosiderando o gráfico, faça o que se pede.



Fonte: *O Estado de S. Paulo*, 26 jun. 2003.

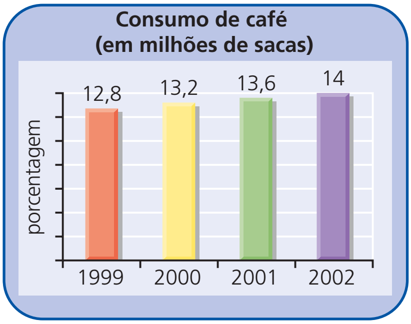
a) Determine a ordem de grandeza do número total de quilômetros quadrados desmatados no período.

b) Determine a quantos campos de futebol de 100 m de comprimento e 70 m de largura corresponde o total desmatado calculado no item *a*.

c) Houve um período de anos consecutivos em que foi registrada pequena variação na área desmatada. Identifique-o.

d) Sabendo que a área da Amazônia Legal é da ordem de 4,9 milhões de quilômetros quadrados, determine o percentual correspondente à área da floresta desmatada em todo o período.

279. (UF-PE) O consumo anual de café em estabelecimentos comerciais no Brasil, de 1999 a 2002, está ilustrado no gráfico a seguir.



Admitindo esses dados, analise as alternativas a seguir, justificando:

a) O consumo cresceu linearmente de 2000 a 2002.

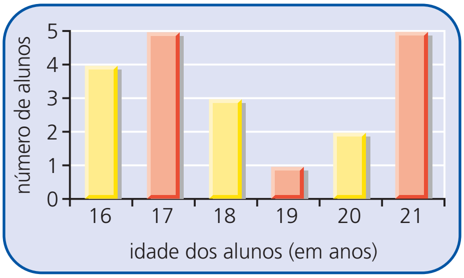
b) Entre 2000 e 2002 o crescimento percentual foi superior a 6%.

c) O crescimento percentual em 2001 foi igual ao crescimento percentual em 2002 (crescimento relativo ao ano anterior).

d) Em 2001 o crescimento percentual (em relação a 2000) foi inferior a 4%.

e) A média anual de consumo foi superior a 13 milhões de sacas.

280. (Vunesp-SP) Num curso de Inglês, a distribuição das idades dos alunos é dada pelo gráfico ao lado.

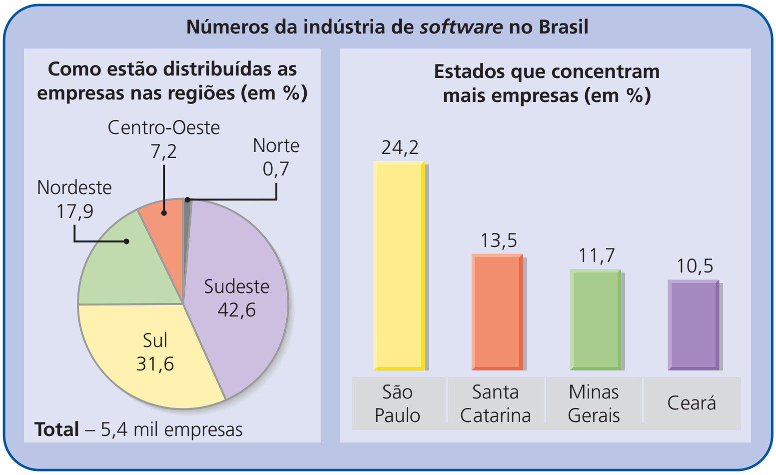


Com base nos dados do gráfico, determine:

a) o número total de alunos do curso e o número de alunos com no mínimo 19 anos;

b) escolhido um aluno ao acaso, qual a probabilidade de sua idade ser no mínimo 19 anos ou ser exatamente 16 anos.

281. Considerando o gráfico, responda às perguntas a seguir.



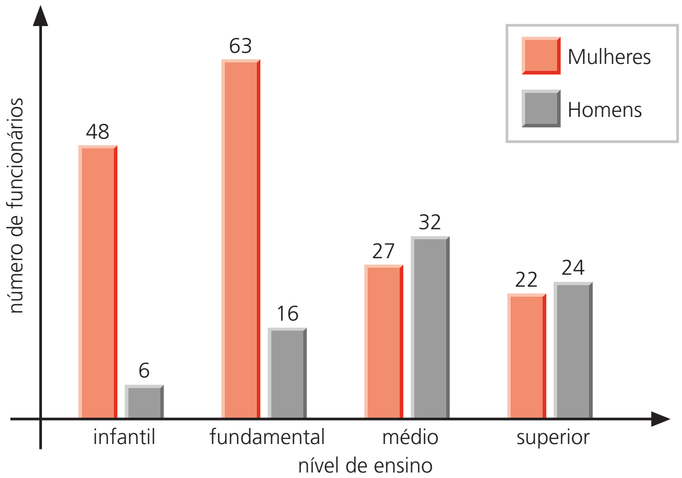
Fonte: *O Estado de S. Paulo*, 12 maio 2003.

a) Qual a medida aproximada do ângulo do setor que representa cada região?

b) Que número representa as empresas de *software* instaladas no Sudeste?

c) Qual a participação percentual de Santa Catarina e São Paulo na região em que cada estado se situa?

282. O gráfico abaixo mostra a distribuição dos funcionários de uma escola integrada (que oferece cursos desde o ensino infantil até o ensino superior) por ensino e por sexo.



Com base no gráfico, assinale V (verdadeira) ou F (falsa) nas proposições seguintes e justifique as falsas:

a) O número de mulheres que trabalham na escola representa mais de $\frac{2}{3}$ do total de funcionários.

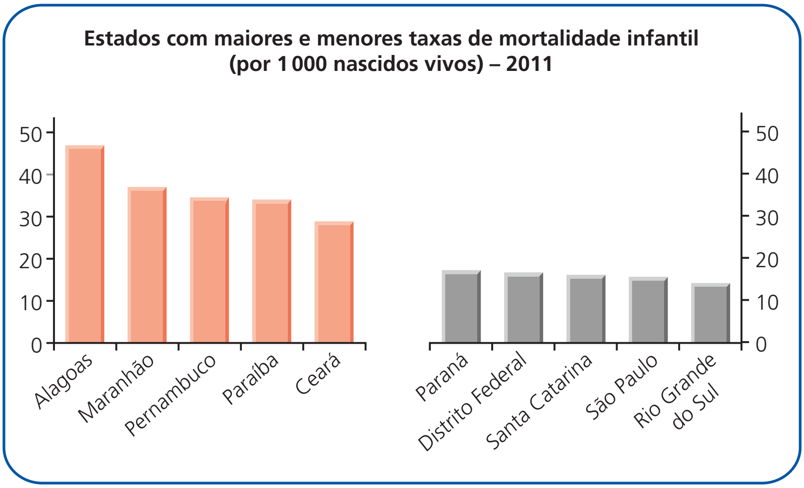
b) O número de homens que trabalham na faculdade supera o número total de homens que trabalham no ensino infantil e fundamental.

c) No ensino fundamental os homens correspondem a menos de 15% do total de funcionários.

d) O número de mulheres que trabalham no ensino fundamental é 150% maior que o número de mulheres que trabalham no ensino médio.

e) Para que as mulheres representem mais de 55% dos funcionários que trabalham no ensino médio é necessário contratar pelo menos mais 11 funcionárias. (Admita que não haverá saída de nenhum funcionário.)

283. Considerando o gráfico, classifique cada afirmação como verdadeira (V) ou falsa (F).



Fonte: *Almanaque Abril*, 2002.

a) A mortalidade infantil em Alagoas praticamente coincide com a mortalidade infantil verificada nos Estados da região Sul, juntos.

b) Uma queda de 20% na taxa de mortalidade infantil da Paraíba reduz essa taxa a menos de 30 mortes por 1 000 nascidos vivos.

c) A taxa de mortalidade infantil do Ceará é aproximadamente o triplo da de São Paulo.

d) A taxa de mortalidade infantil de Pernambuco, expressa em termos percentuais, é maior que 3%.

284. Na tabela abaixo estão relacionados os 30 municípios brasileiros que atingiram os maiores índices de desenvolvimento humano municipal (IDHM), de acordo com o censo de 2000.

	Município	**IDHM**
1º	São Caetano do Sul (SP)	0,919
2º	Águas de São Pedro (SP)	0,908
3º	Niterói (RJ)	0,886
4º	Florianópolis (SC)	0,875
5º	Santos (SP)	0,871
6º	Bento Gonçalves (RS)	0,870

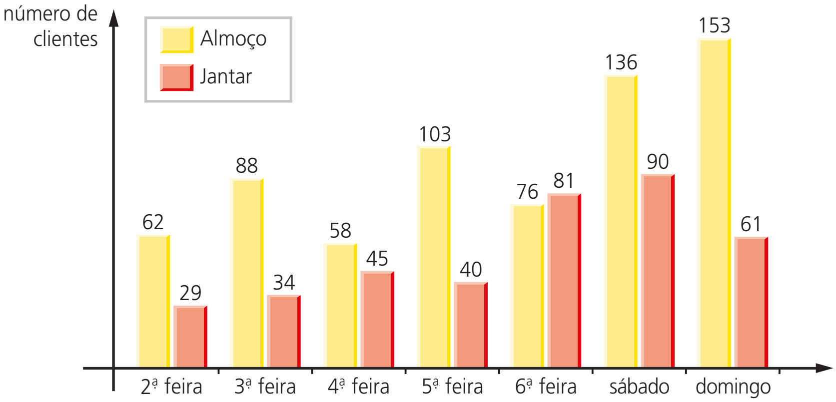
	Município	IDHM
7º	Balneário Camboriú (SC)	0,867
8º	Joaçaba (SC)	0,866
9º	Porto Alegre (RS)	0,865
10º	Fernando de Noronha* (PE)	0,862
11º	Carlos Barbosa (RS)	0,858
12º	Caxias do Sul (RS)	0,857
13º	Joinville (SC)	0,857
14º	Vinhedo (SP)	0,857
15º	Jundiaí (SP)	0,857
16º	Selbach (RS)	0,856
17º	Curitiba (PR)	0,856
18º	Vitória (ES)	0,856
19º	Luzerna (SC)	0,855
20º	Blumenau (SC)	0,855
21º	Ribeirão Preto (SP)	0,855
22º	Lacerdópolis (SC)	0,854
23º	Santana de Parnaíba (SP)	0,853
24º	Campinas (SP)	0,852
25º	Ivoti (RS)	0,851
26º	Videira (SC)	0,851
27º	Quatro Pontes (PR)	0,851
28º	Saltinho (SP)	0,851
29º	Veranópolis (RS)	0,850
30º	Jaraguá do Sul (SC)	0,850

*Distrito Estadual
Fonte: *Folha de S. Paulo*, 3 out. 2003.

a) Sabendo que o índice do 1º colocado é x% maior que o índice do 2º colocado e y% maior que o índice do 30º colocado, determine *x* e *y*.

b) Faça um gráfico de barras para representar o número de municípios pertencentes a cada Estado relacionado na tabela.

285. O gráfico seguinte mostra o número de clientes que uma churrascaria atendeu durante certa semana.



Os preços praticados por esse estabelecimento são:

almoço: de 2ª a 6ª feira → R$ 13,00

sábado e domingo → R$ 18,00

jantar: todos os dias → R$ 12,00

Qual foi o faturamento da churrascaria nessa semana?

VII. Histograma

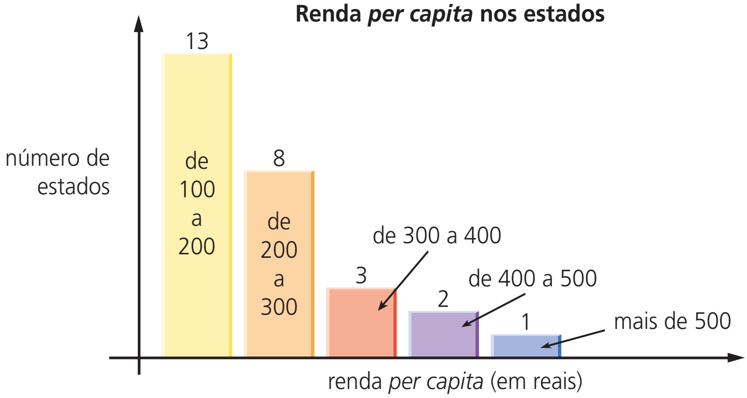
Na tabela seguinte, extraída do *Atlas do Desenvolvimento Humano no Brasil*, está relacionada à renda *per capita* média em cada estado (dados do Censo de 2000), expressa em reais.

Distrito Federal	605,4	Minas Gerais	276,6	Rio Grande do Norte	176,2
Santa Catarina	348,7	Espírito Santo	289,6	Ceará	156,2
São Paulo	442,7	Amapá	211,4	Acre	180,7
Rio Grande do Sul	357,7	Roraima	232,5	Bahia	160,2
Rio de Janeiro	413,9	Rondônia	233,8	Sergipe	163,5
Paraná	321,4	Pará	168,6	Paraíba	150,2
Mato Grosso do Sul	287,5	Amazonas	173,9	Piauí	129,0
Goiás	286,0	Tocantins	172,6	Alagoas	139,9
Mato Grosso	288,1	Pernambuco	183,8	Maranhão	110,4

Fonte: *Folha de S.Paulo*, 3 out. 2003.

Agrupando esses valores em cinco classes de intervalos — 100 $\vdash$ 200, 200 $\vdash$ 300, 300 $\vdash$ 400, 400 $\vdash$ 500 e mais de 500 —, é possível construir uma tabela de frequência. Para representar graficamente essas informações, construímos um gráfico semelhante ao de barras verticais, usando como abscissa os limites das classes de intervalos e como ordenada a frequência (absoluta ou relativa).

Renda *per capita* nos estados



Esse tipo de gráfico é denominado **histograma**.

EXERCÍCIOS

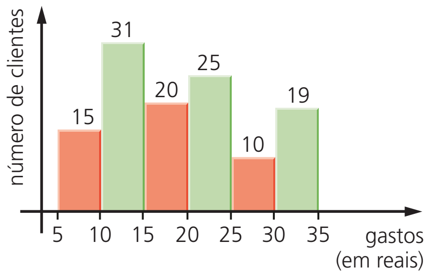
286. Os dados seguintes referem-se à participação percentual da indústria na composição do Produto Interno Bruto (PIB) dos estados brasileiros.

AC	17,4%	CE	36,0%	MG	41,6%	PR	36,9%	RS	31,4%
AL	44,7%	DF	11,1%	MS	23,6%	RJ	32,0%	SC	36,3%
AM	41,0%	ES	30,4%	MT	11,7%	RN	41,9%	SE	28,5%
AP	24,0%	GO	21,2%	PA	29,1%	RO	6,6%	SP	41,0%
BA	36,7%	MA	32,0%	PB	20,5%	RR	26,3%	TO	4,9%
				PE	25,6%				
				PI	21,4%				

Fonte: *Almanaque Abril*, 2002.

Faça um histograma representativo dessa situação, agrupando os dados em intervalos de amplitude 10.

287. O histograma ao lado mostra os gastos dos clientes de um supermercado registrados em um caixa expresso durante uma manhã.



a) Que porcentagem do total de clientes gastou pelo menos 20 reais?

b) Que porcentagem do total de clientes gastou menos de 15 reais?

c) Estime a menor quantia possível que pôde ter sido arrecadada nesse caixa na manhã considerada.

288. O departamento pessoal de uma pequena fábrica relacionou o tempo de serviço e o salário de seus 30 funcionários. Tais dados encontram-se na tabela a seguir.

Funcionário	Tempo de serviço (meses)	Salário (R$)
1	20	832,00
2	16	641,00
3	6	1 105,00
4	7	432,00
5	10	592,00
6	14	617,00
7	18	720,00
8	26	864,00
9	18	803,00
10	16	851,00
11	13	692,00
12	8	1 625,00
13	17	2 143,00
14	21	1 294,00
15	23	967,00

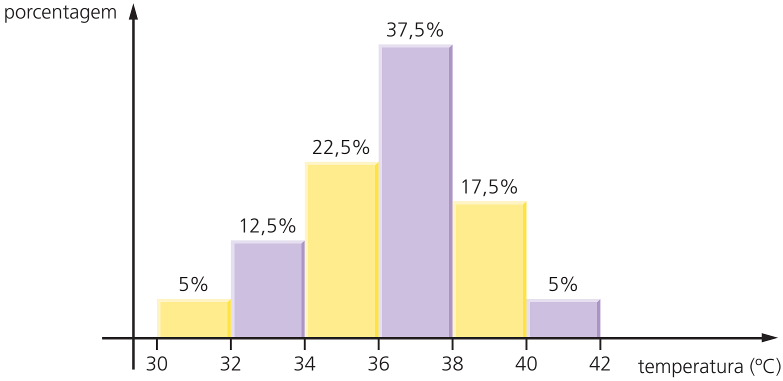
Funcionário	Tempo de serviço (meses)	Salário (R$)
16	9	873,00
17	11	556,00
18	25	831,00
19	5	886,00
20	10	1 427,00
21	13	1 061,00
22	17	1 317,00
23	8	1 248,00
24	19	960,00
25	15	820,00
26	9	749,00
27	7	861,00
28	4	639,00
29	11	603,00
30	15	1 512,00

a) Faça um histograma para representar a distribuição do tempo de serviço dos funcionários, utilizando intervalos de amplitude 4, a partir do menor valor encontrado.

b) Faça um histograma para representar a distribuição dos salários, utilizando intervalos de amplitude 200, a partir do valor 400.

c) Refaça o histograma do item *b*, supondo a contratação de 8 novos funcionários, cada um com salário de R$ 700,00.

289. O histograma seguinte mostra as temperaturas máximas diárias registradas em 80 dias durante um verão na cidade do Rio de Janeiro.



a) Em quantos dias a temperatura máxima manteve-se abaixo dos 38 °C?

b) Em quantos dias a temperatura máxima variou de 36 °C a 42 °C?

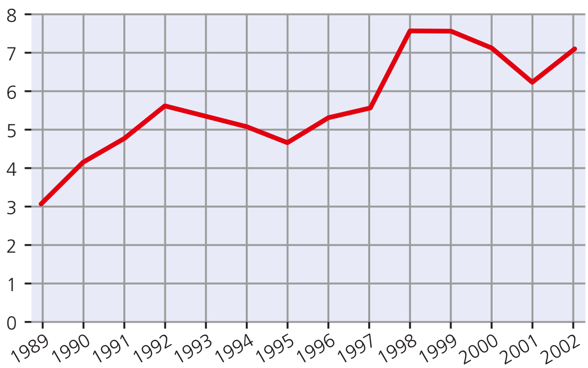
c) O dono de uma barraca de praia disse que o carioca costuma tomar 1 litro de cerveja na praia por dia quando a temperatura está abaixo de 32 °C e que, para cada 2 °C de aumento da temperatura, esse consumo sobe 10% (em relação ao consumo anterior). Se um carioca foi à praia nesses 80 dias, quantos litros de cerveja consumiu ao todo, de acordo com essa previsão?

VIII. Gráfico de linhas (poligonal)

O gráfico seguinte mostra a evolução da taxa de desemprego no Brasil no período de 1989 a 2002. A cada ano está associada certa taxa de desemprego.

Desse modo, ficam determinados diversos pontos no gráfico; unindo-os por segmentos de reta, obtemos o chamado **gráfico de linhas** ou **gráfico de curva poligonal**. É importante lembrar que esse tipo de gráfico define uma função entre as variáveis (taxa e anos) envolvidas. Dizemos que a taxa de desemprego é função do tempo.

Evolução do desemprego no Brasil
(% da população economicamente ativa)



Fonte: *Almanaque Abril* – atualidades de vestibular, 2004.

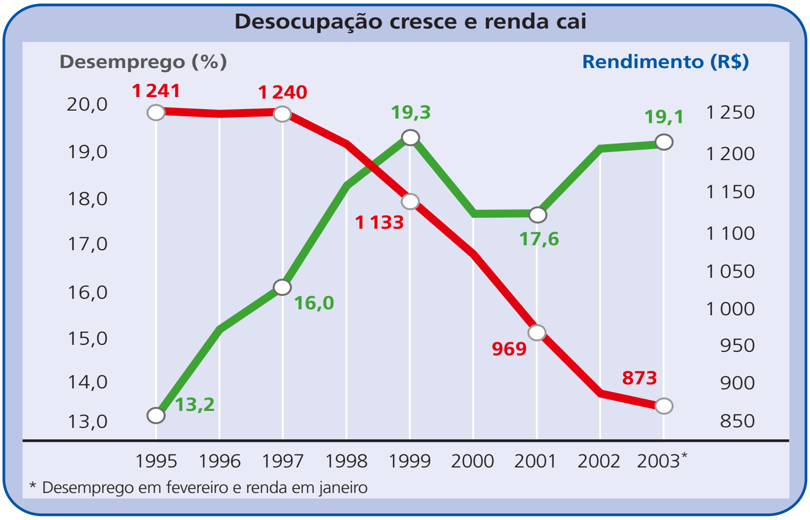
A leitura do gráfico nos permite concluir que:

- A taxa de desemprego aumentou de 1989 a 1992, teve ligeira queda de 1992 a 1995 e a partir daí cresceu até 1998. De 1998 a 1999, manteve-se praticamente constante, caindo a partir daí até 2001, quando houve retomada de crescimento.
- Nos últimos cinco anos, a taxa de desemprego manteve-se acima de 6% da população economicamente ativa.
- Considerando-se dois anos consecutivos, pode-se dizer que o maior aumento do desemprego ocorreu de 1997 a 1998, com acréscimo de aproximadamente 2 pontos percentuais na taxa.

O gráfico de linhas é muito usado quando se quer representar o comportamento de uma variável cujos valores diminuem ou aumentam no decorrer do tempo de maneira contínua.

EXERCÍCIOS

290. Observe os dados apresentados no gráfico e responda.



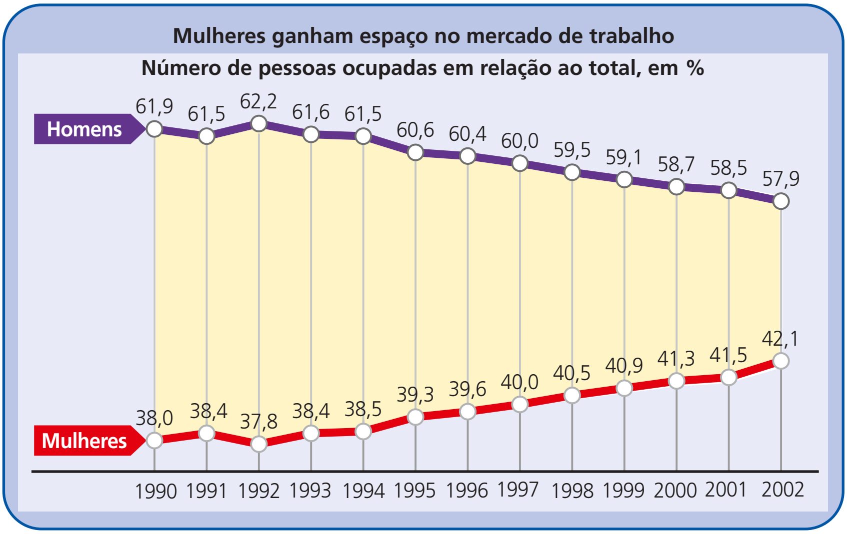
Fonte: *O Estado de S. Paulo*, 28 abr. 2004.

a) Identifique os períodos de crescimento e decrescimento relativos às duas variáveis do gráfico.

b) Qual foi a perda percentual nos rendimentos no período de 1995 a 2003?

c) Suponha que, em 2003, o total dos rendimentos dos trabalhadores empregados tivesse sido dividido igualmente entre toda a população economicamente ativa a fim de que todos tivessem renda. Quantos reais **a menos** um trabalhador empregado passaria a receber?

291. Observe o gráfico a seguir:



Fonte: *Folha de S. Paulo*,18 nov. 2002.

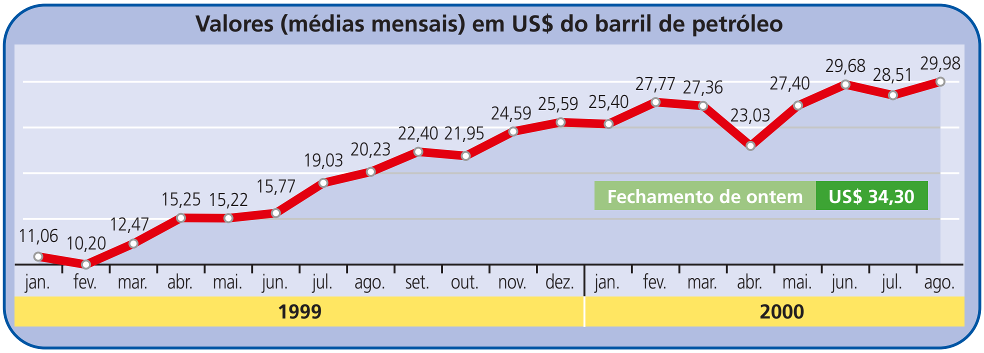
Responda:

a) A partir de que ano é possível afirmar que a participação masculina tornou-se decrescente e a feminina crescente?

b) Em um grupo de 500 000 trabalhadores, no ano de 2001, qual era a diferença entre o número de homens e de mulheres?

c) Em que anos a diferença entre a participação masculina e a feminina não excedeu 20 pontos percentuais? Em que anos ela ultrapassou 23 pontos percentuais?

292. (UF-MT, adaptado) Observe a figura:



Adaptado de: *Jornal do Brasil*, 7 set. 2000.

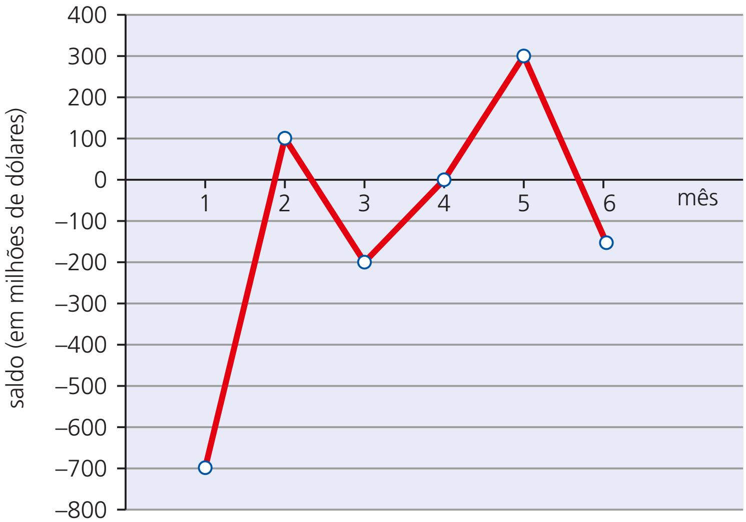
A partir das informações dadas e utilizando a aproximação de duas casas decimais, julgue os itens:

0) No período considerado, a variação do menor valor do barril de petróleo para o maior foi de 193,92%.

1) A média aritmética dos valores do barril de petróleo dos meses relativos ao segundo trimestre de 1999 é US$ 15,41.

2) Se a variação (em dólar) do valor do barril de petróleo de julho de 2000 a agosto de 2000 se mantivesse constante para os meses seguintes, o valor do barril ultrapassaria US$ 40,00 em fevereiro de 2001.

293. (UF-AL) O saldo da balança comercial de um país é a diferença entre os valores de suas exportações e importações. O gráfico mostra o saldo da balança comercial brasileira no primeiro semestre de 1999, em números aproximados.



De acordo com o gráfico, assinale V ou F nas proposições seguintes:

a) O valor das importações superou o das exportações em janeiro.

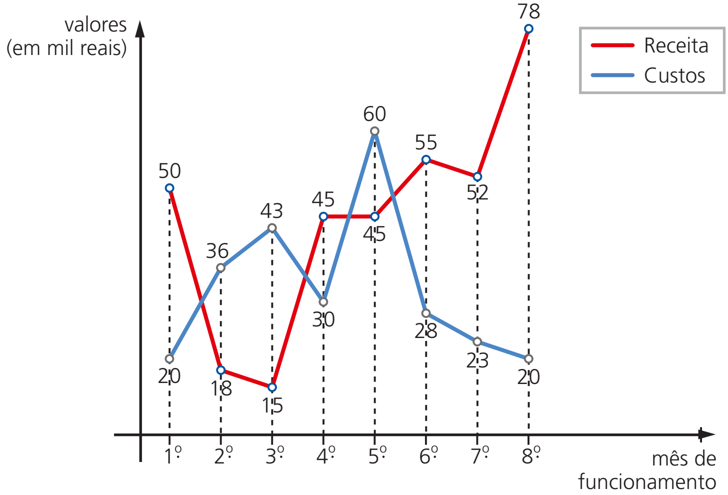
b) O valor das exportações superou o das importações em março.

c) O valor das exportações do país vem aumentando em 1999.

d) O saldo da balança comercial em junho é de aproximadamente –150 000 dólares.

e) O saldo acumulado da balança comercial no primeiro semestre é de aproximadamente –650 000 000 dólares.

294. O gráfico seguinte mostra o desempenho de uma pequena fábrica nos oito primeiros meses de funcionamento:



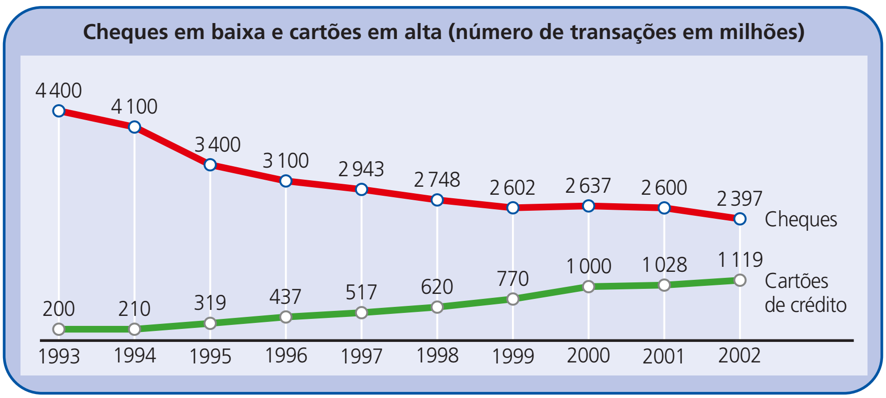
Com base no gráfico, responda:

a) Em que meses a empresa operou no "vermelho", isto é, os custos superaram a receita?

b) Qual foi a receita total da fábrica nesse período?

c) Faça um gráfico de linhas para representar a evolução do lucro da fábrica mês a mês nesse período; em seguida calcule o lucro total no período.

295. O gráfico abaixo mostra queda nas operações com cheques e avanço nas operações com cartões de crédito. Os valores referem-se às quantidades de transações efetuadas (em milhões).



Fonte: *Veja*,10 set. 2003.

Assinale V (verdadeira) ou F (falsa) nas afirmativas seguintes, justificando as falsas:

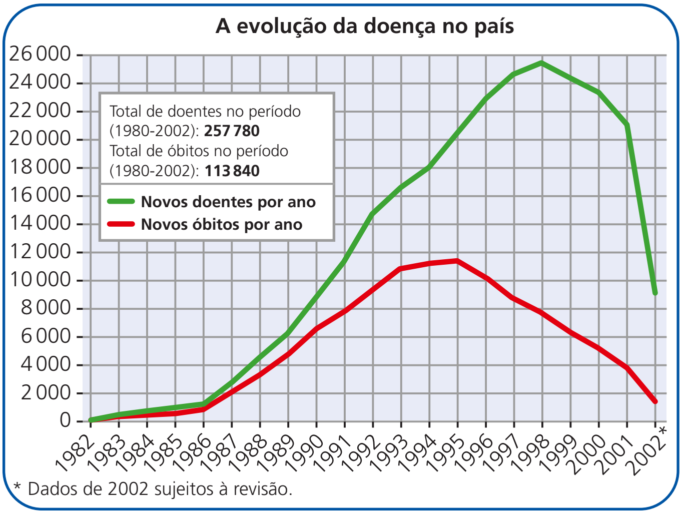
a) As transações efetuadas com cartões aumentaram a cada ano no período considerado.

b) De 1993 a 2002 registrou-se queda de aproximadamente 45% nas transações efetuadas com cheques.

c) O crescimento percentual das transações com cartões aumentou 560% no período de 1993 a 2002.

d) Considerando os dados de 2001 e 2002, pode-se dizer que a queda percentual nas operações com cheques correspondeu ao ganho percentual nas operações com cartões, com uma margem de erro de até 2 pontos percentuais.

296. Segundo o *Almanaque Abril* — atualidades de vestibular, de 2004, graças "a uma eficiente campanha de combate à Aids, que incluiu a distribuição gratuita de remédios, o Brasil conseguiu domar a epidemia e seu programa se tornou um exemplo para a comunidade mundial". Com base no gráfico sobre a evolução da doença no país, apresentado pela publicação, faça o que se pede.



Fonte: *Almanaque Abril* – atualidades de vestibular, 2004.

a) Identifique os períodos de crescimento e decrescimento das variáveis em estudo.

b) Faça uma estimativa do número total de óbitos dos últimos cinco anos.

c) A queda na mortalidade por Aids deve-se principalmente à distribuição gratuita de coquetéis antirretrovirais aos infectados. De acordo com o gráfico, a partir de que ano teve início esse programa?

d) A partir de que ano houve queda tanto no número de novos casos como no número de óbitos?

IX. Medidas de centralidade e variabilidade

Nos itens anteriores, vimos como resumir um conjunto de dados em tabelas de frequência e também como representá-los graficamente.

Agora, a partir dos valores assumidos por uma variável quantitativa, vamos estabelecer medidas correspondentes a um resumo da distribuição de tais valores.

Estabeleceremos um valor **médio** ou **central** e um valor indicativo do grau de **variabilidade** ou **dispersão** em torno do valor central.

Como valores centrais, vamos estudar a **média**, a **mediana** e a **moda**.

Como medida da variabilidade, vamos estudar a **variância**, o **desvio padrão** e o **desvio médio**.

X. Média aritmética

Seja *x* uma variável quantitativa e $x_1, x_2, ..., x_n$ os valores assumidos por *x*. Define-se a **média aritmética** de *x* – indicada por $\bar{x}$ – como a divisão da soma de todos esses valores pelo número de valores, isto é:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + ... + x_n}{n}$$

Exemplos:

1º) Um aluno, preparando-se para o exame vestibular, fez 12 simulados no cursinho ao longo do ano. Em cada simulado, o número de questões era oitenta. Os valores seguintes correspondem às pontuações obtidas nesses exames:

56 – 52 – 61 – 53 – 48 – 68

49 – 59 – 61 – 62 – 60 – 55

Qual é a média aritmética desses valores?

Temos:

$$\bar{x} = \frac{\sum_{i=1}^{12} x_i}{12} = \frac{56 + 52 + ... + 60 + 55}{12} = \frac{684}{12} = 57$$

A nota média obtida por esse aluno é 57 pontos. Qual é o significado desse valor?

Caso o aluno apresentasse a mesma pontuação (desempenho) em todos os simulados, essa pontuação deveria ser 57 pontos a fim de que fosse obtida a pontuação total de 684 pontos, equivalente à soma dos pontos obtidos efetivamente nas 12 provas.

Observe que em nenhum simulado ocorreu a pontuação média, que é 57 pontos.

Isso sugere que, ao calcularmos a média aritmética de um conjunto de valores, podemos obter um resultado que não coincide com nenhum dos valores que a variável assume.

2º) A média aritmética de um conjunto formado por 10 elementos é igual a 8. Acrescentando-se a esse conjunto o número 41, qual será a nova média?

Sejam $x_1, x_2, \ldots, x_{10}$ os elementos desse conjunto.

Temos:

$$\bar{x} = 8 \Rightarrow \frac{\sum_{i=1}^{10} x_i}{10} = 8 \Rightarrow \sum_{i=1}^{10} x_i = 80$$

Ao acrescentarmos o número 41 ao conjunto, a soma de todos os seus elementos será 80 + 41 = 121 e a nova média ($\bar{x}'$) será dada por:

$$\bar{x}' = \frac{\left(\sum_{i=1}^{10} x_i\right) + 41}{10 + 1} = \frac{80 + 41}{11} = 11$$

Propriedades:

Vamos estudar agora duas propriedades da média aritmética.

Sejam $x_1, x_2, \ldots, x_n$ os valores assumidos por uma variável *x* e $\bar{x}$ a média aritmética correspondente.

> Se a cada x_i (i = 1, 2, ..., n) adicionarmos uma constante real *c*, a média aritmética fica adicionada de *c* unidades.

Essa propriedade pode ser facilmente demonstrada.

Consideremos que os novos valores assumidos por essa variável sejam: $x_1 + c, x_2 + c, \ldots, x_n + c$.

A nova média ($\bar{x}'$) é dada por:

$$\bar{x}' = \frac{\sum_{i=1}^{n} (x_i + c)}{n} = \frac{(x_1 + c) + (x_2 + c) + \ldots + (x_n + c)}{n} =$$

$$= \frac{(x_1 + x_2 + \ldots + x_n)}{n} + \frac{\overbrace{(c + c + \ldots + c)}^{n \text{ vezes}}}{n} + \frac{\sum_{i=1}^{n} x_i}{n} = \frac{n \cdot c}{n}$$

isto é:

$$\overline{x}' = \overline{x} + c$$

Se multiplicarmos cada x_i ($i = 1, 2, ..., n$) por uma constante real *c*, a média aritmética fica multiplicada por *c*.

Para demonstrar essa segunda propriedade, consideremos que os novos valores assumidos por essa variável sejam: $cx_1, cx_2, ..., cx_n$.

A nova média ($\overline{x}'$) é dada por:

$$\overline{x}' = \frac{\sum_{i=1}^{n}(cx_i)}{n} = \frac{cx_1 + cx_2 + \ldots + cx_n}{n} = \frac{c \cdot (x_1 + x_2 + \ldots + x_n)}{n} = c \cdot \frac{\sum_{i=1}^{n} x_i}{n}$$

isto é:

$$\overline{x}' = c \cdot \overline{x}$$

XI. Média aritmética ponderada

Seja *x* uma variável quantitativa que assume os valores $x_1, x_2, ..., x_k$ com **frequências absolutas** respectivamente iguais a $n_1, n_2, ..., n_k$. A **média aritmética ponderada** de *x* – indicada por $\overline{x}$ – é definida como a divisão da soma de todos os produtos $x_i \cdot n_i$ ($i = 1, 2, ..., k$) pela soma das frequências, isto é:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots + x_k \cdot n_k}{n_1 + n_2 + \ldots + n_k}$$

Lembrando que a frequência relativa (f_i) é definida por $\dfrac{n_i}{\sum_{i=1}^{k} n_i}$, é possível também expressar a média por:

$$\overline{x} = \sum_{i=1}^{k} x_i \cdot f_i = x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_k \cdot f_k$$

Exemplos:

1º) Um feirante possuía 50 kg de maçã para vender em uma manhã. Começou a vender as frutas por R$ 2,50 o quilo e, com o passar das horas, reduziu o preço em duas ocasiões para não haver sobras. A tabela seguinte informa a quantidade de maçãs vendidas em cada período, bem como os diferentes preços cobrados pelo feirante.

Período	Preço por quilo (em reais)	Número de quilos de maçã vendidos
Até às 10 h	2,50	32
Das 10 h às 11 h	2,00	13
Das 11 h às 12 h	1,40	5

Naquela manhã, por quanto foi vendido, em média, o quilo da maçã?
Sendo $\overline{p}$ o preço médio do quilo da maçã, temos:

$$\overline{p} = \frac{\overbrace{2{,}50 + 2{,}50 + \ldots + 2{,}50}^{32 \text{ vezes}} + \overbrace{2{,}0 + \ldots + 2{,}0}^{13 \text{ vezes}} + \overbrace{1{,}40 + 1{,}40 + \ldots + 1{,}40}^{5 \text{ vezes}}}{32 + 13 + 5}$$

isto é:

$$\overline{p} = \frac{2{,}50 \cdot 32 + 2{,}00 \cdot 13 + 1{,}40 \cdot 5}{50} = \frac{113}{50} \cong 2{,}26 \text{ reais}$$

Ou seja, 2,26 reais é o preço médio do quilo de maçãs vendido.

Dizemos que se trata de uma média aritmética ponderada dos preços, em que o "fator de ponderação" (que também pode ser chamado de "peso") corresponde à quantidade de maçãs vendidas (frequência absoluta) em cada período.

2º) A fim de arrecadar recursos para a festa de formatura, cada formando recebeu uma rifa com 20 números para vendê-los a seus conhecidos. Encerrado o prazo combinado, foi feito o levantamento de quantos números cada um vendeu e constatou-se que 10% dos formandos venderam 10 números, 30% venderam 15 números e os demais conseguiram vender todos os números. Qual foi a média de números da rifa que cada formando vendeu?

A variável (x) em questão é a quantidade de números vendidos. Os valores assumidos por *x* são 10, 15 e 20, com frequências relativas iguais a 0,10, 0,30 e 0,60, respectivamente.

Segue que a média ($\overline{x}$) é:

$$\overline{x} = \sum_{i=1}^{3} x_i f_i = 10 \cdot 0{,}10 + 15 \cdot 0{,}30 + 20 \cdot 0{,}60 = 17{,}5$$

Isso significa que, em média, os formandos venderam 17,5 números da rifa.

EXERCÍCIOS

297. Calcule, em cada caso, a média aritmética dos valores:

a) 18 – 21 – 25 – 19 – 20 – 23 – 21

b) 35 – 36 – 37 – 38 – 39 – 40

c) 7 – 7 – 7 – 8 – 8 – 8 – 9 – 9 – 10 – 10 – 10 – 10

d) 0,5 – 0,5 – 0,5 – 0,5 – 0,25 – 0,25

e) a – a – a – a – a – b – b – b – c – c

f) 43 – 49 – 52 – 41 – 47 – 50 – 53 – 44

298. Um ônibus de excursão partiu com 40 turistas a bordo, dos quais 8 reservaram a viagem com antecedência e pagaram, cada um, R$ 300,00. Os demais pagaram, cada um, R$ 340,00 pela viagem. Qual foi o preço médio que cada turista pagou nessa excursão?

299. Sejam A = {x, 6, 3, 4, 5} e B = {9, 1, 4, 8, x, 6, 11, 3}.

a) Determine *x* para que as médias aritméticas dos dois conjuntos sejam iguais.

b) Determine os possíveis valores inteiros de *x* de modo que $\overline{x}_A$ não ultrapasse 4 e $\overline{x}_B$ seja, no mínimo, igual a 5.

300. Para que valores de *a* as médias aritméticas de $\{-3, a, 10, 9\}$ e $\{-2, 3, a^2, -5\}$ coincidem?

301. *x* é uma variável que assume os valores:

$$11 - 8 - 7 - a - 16 - 10$$

Determine *a* de modo que:

a) $\overline{x} = 11$ b) $12 \leqslant \overline{x} < 13$ c) $\overline{x} < 0$

302. Os dados na tabela abaixo referem-se ao número de unidades de um livro didático vendidas, mês a mês, nos dois primeiros anos após seu lançamento.

Mês	1º ano	2º ano
Janeiro	2460	3152
Fevereiro	2388	2963
Março	2126	2049
Abril	1437	1614
Maio	931	1024
Junho	605	898

Mês	1º ano	2º ano
Julho	619	910
Agosto	421	648
Setembro	742	937
Outubro	687	702
Novembro	1043	1051
Dezembro	1769	2016

a) Do 1º para o 2º ano de vendas, a média mensal de livros vendidos aumentou em x unidades. Qual é o valor de x?

b) Do 1º para o 2º ano de vendas, a média mensal de livros vendidos aumentou em y%. Qual é o valor de y?

303. Os dados seguintes referem-se às quantidades mensais de CDs do cantor X vendidos durante um ano.

$$3\,000 - 4\,000 - 3\,500 - 5\,200 - 6\,700 - 5\,000$$

$$8\,500 - 7\,600 - 6\,500 - 6\,400 - 7\,000 - 5\,400$$

Em quantos meses as vendas mensais superaram a média de CDs vendidos?

304. A média aritmética de 80 números é igual a 40,5. Adicionando-se a esse conjunto de valores o número 243, qual será a nova média aritmética?

305. A média aritmética de uma lista formada por 55 números é igual a 28. Adicionando-se dois números a essa relação, a média aumenta em 2 unidades. Determine-os, sabendo que um deles é o triplo do outro.

306. A média aritmética de 45 números é igual a 6. Ao acrescentarmos o número x a esses valores, a média aumenta em 50%.

a) Qual é o valor de x?

b) Qual é a média aritmética dos números $\frac{x}{2}, \frac{x}{4}, \frac{x}{6}, \frac{x}{8}, \frac{x}{12}$?

307. Uma prova de Conhecimentos Gerais foi aplicada em duas turmas, A e B, com n e m alunos, respectivamente. A média das notas da turma A foi 6,8 e a da turma B foi 5,2. Juntando as notas das duas turmas, a média geral foi 5,8.

a) Intuitivamente, responda: O que é maior: n ou m?

b) Determine n e m, sabendo que a diferença entre eles é igual a 14.

308. A média de "pesos" de 25 clientes hospedadas em um spa era de 84 kg. A elas juntou-se um grupo de n amigas. Curiosamente, cada amiga desse grupo "pesava" 90 kg. Determine o valor de n, sabendo que a média de "pesos" de todas as clientes hospedadas no spa aumentou 1 quilograma.

309. A média aritmética de 15 números é 26. Retirando-se um deles, a média dos demais passa a ser 25. Qual foi o número retirado?

310. A média aritmética de n números é 29. Retirando-se o número 24, a média aumenta para 30. Qual é o valor de n?

311. Determine n a fim de que a média aritmética dos números 2^n, 2^{n+1}, 2^{n+2} e 2^{n+3} seja igual a 60.

312. A média aritmética de 7 números inteiros é 4. Determine-os, sabendo que eles formam uma P. A. crescente de razão 6.

313. Calcule a média aritmética entre os números reais log 2, log 3, log 4 e log 5, sabendo que log 1,2 $\cong$ 0,08.

314. A média aritmética de 10 números, x_1, x_2, ..., x_{10}, é 4. Qual será a nova média se:

a) x_1 for aumentado de 4 unidades e x_2 aumentado de 8 unidades?

b) x_1 for subtraído de 10 unidades e x_2 aumentado de 6 unidades?

315. A tabela ao lado mostra o salário médio dos trabalhadores de três cidades, A, B e C, que compõem uma região metropolitana.

Cidade	Salário médio (em reais)
A	530,00
B	600,00
C	700,00

Determine o salário médio na região metropolitana se:

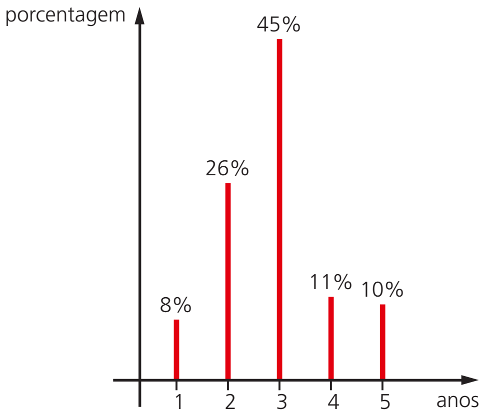
a) A, B e C têm o mesmo número de trabalhadores;

b) A tem 200 000 trabalhadores, B tem 300 000 e C tem 500 000;

c) A tem o dobro de trabalhadores de B, que tem o triplo de trabalhadores de C.

316. Na situação do exercício anterior, suponha que A concentre 70% dos trabalhadores da região metropolitana. Determine o percentual de trabalhadores que vivem em B e C, respectivamente, a fim de que o salário médio dos trabalhadores da região seja R$ 560,00.

317. O gráfico seguinte informa a distribuição do tempo de serviço (em anos) dos funcionários de uma pequena empresa.



Qual é o tempo médio de trabalho dos funcionários dessa empresa?

318. Sejam $x_1, x_2, \ldots, x_n$ os *n* valores assumidos por uma variável quantitativa e $\bar{x}$ a média aritmética correspondente a tais valores. Estabeleça uma relação entre a nova média ($\bar{x}'$) e $\bar{x}$ em cada caso a seguir:

a) Cada x_i (i = 1, 2, ..., n) é aumentado de duas unidades.

b) Cada x_i (i = 1, 2, ..., n) é multiplicado por três.

c) Cada x_i (i = 1, 2, ..., n) é diminuído de cinco unidades.

d) Cada x_i (i = 1, 2, ..., n) é multiplicado por –2 e ao resultado são acrescentadas três unidades.

e) Cada x_i (i = 1, 2, ..., n) é subtraído de $\bar{x}$ unidades.

319. A tabela ao lado mostra o número de gols por partida registrado nas duas primeiras rodadas de um campeonato brasileiro.

a) Qual foi a média de gols por partida registrada nas duas primeiras rodadas?

b) A rodada seguinte previa a realização de *n* jogos no sábado e a dos demais no domingo. Em cada um dos jogos de sábado foram marcados 3 gols. Com isso, a média de gols do campeonato (computadas as duas primeiras rodadas e os jogos de sábado) elevou-se para 2,5 gols por partida. Qual é o valor de *n*?

Nº de gols	Frequência absoluta
0	5 jogos
1	6 jogos
2	8 jogos
3	4 jogos
4	5 jogos
5	3 jogos
6	1 jogo

320. A média dos salários dos funcionários de uma loja é de R$ 806,00. Qual será a nova média salarial se:

a) cada funcionário receber um aumento de R$ 120,00?

b) cada funcionário receber um aumento de 20%?

321. Uma prova foi aplicada em duas turmas, A e B, e as médias obtidas foram 7,2 e 6,3, respectivamente. Se cada aluno da turma A tivesse obtido *n* pontos a menos e cada aluno da turma B tivesse obtido *n* pontos a mais, as médias das duas turmas seriam iguais. Qual é o valor de *n*?

322. Em uma empresa, a média salarial é R$ 930,00. Pretende-se dar a cada funcionário um aumento de 5% e um abono de R$ 80,00. Qual será a nova média de salários na empresa se:

a) o aumento for dado antes do abono?

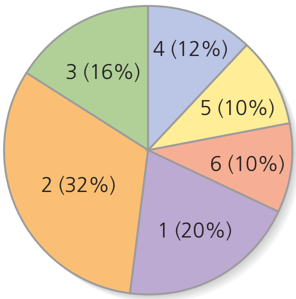
b) o aumento for dado após a incorporação do abono ao salário?

323. É comum encontrarmos produtos com conteúdo líquido menor que o declarado nas embalagens. Em uma pequena cidade, doces de leite são vendidos em copos de vidro em cujos rótulos consta a informação relativa ao "peso" de 200 g. Dois fabricantes, A e B, fornecem doces com conteúdo real médio de 190 g e 195 g, respectivamente. Um supermercado comprou um total de n copos (somadas as duas marcas) de doce de leite, e verificou-se que o conteúdo médio líquido do lote era 193,5 g.

Determine o número de copos comprados de cada fabricante, sabendo que um deles vendeu 40 copos a mais que o outro.

324. (UFF-RJ) Cada um dos 60 alunos da turma A obteve, na avaliação de um trabalho, nota 5 ou nota 10. A média aritmética dessas notas foi 6. Determine quantos alunos obtiveram nota 5 e quantos obtiveram nota 10.

325. (Unicamp-SP) O gráfico a seguir, em forma de pizza, representa as notas obtidas em uma questão pelos 32 000 candidatos presentes à primeira fase de uma prova de vestibular. Ele mostra, por exemplo, que 32% desses candidatos tiveram nota 2 nessa questão.



Pergunta-se:

a) Quantos candidatos tiveram nota 3?

b) É possível afirmar que a nota média, nessa questão, foi menor ou igual a 2? Justifique sua resposta.

326. Em uma fábrica, a média salarial de determinado setor, que emprega 20 funcionários, é 832 reais. Um deles, que ganhava 950 reais, foi afastado, e foram contratados 2 novos funcionários, um com salário de 780 reais e o outro com salário de 920 reais. Qual é o número inteiro mais próximo da nova média de salários nesse setor?

327. Em uma classe de educação infantil, a média de idade das 25 crianças é 4 anos e 3 meses. Qual é o número de crianças com 4 anos e 9 meses que devem ingressar nessa classe a fim de elevar essa média para 4 anos e 4 meses?

328. Um programa beneficente veiculado em um canal de TV tinha como objetivo arrecadar fundos para crianças carentes. O telespectador poderia escolher entre 10, 20 ou 50 reais e ligar para o número correspondente ao valor escolhido a fim de fazer a doação. Na primeira hora, 50 000 pessoas fizeram doações, das quais 48% contribuíram com o valor mínimo, 37% com o valor intermediário e cada uma das demais com o valor maior.

a) Qual foi a média de doações da primeira hora?

b) Na hora seguinte, 30 000 pessoas contribuíram para a campanha, das quais $\frac{1}{3}$ colaborou com o valor mínimo. Determine o valor doado pelas demais pessoas, sabendo que a doação média das duas primeiras horas foi R$ 22,80.

329. Um grupo de 20 nadadores, cuja média de altura é 1,88 m, está treinando para uma competição. Se um grupo de 7 atletas cuja média de altura é 1,92 m se juntar ao primeiro grupo, qual será a média de altura dos 27 atletas?

330. A média aritmética dos números $a_1, a_2, a_3, ..., a_{14}, a_{15}$ é 24. Qual é a média aritmética dos números $a_1 + 1, a_2 + 2, ..., a_{14} + 14, a_{15} + 15$?

331. (Fuvest-SP) Numa classe com vinte alunos as notas do exame final podiam variar de 0 a 100 e a nota mínima para aprovação era 70. Realizado o exame, verificou-se que oito alunos foram reprovados. A média aritmética das notas desses oito alunos foi 65, enquanto a média dos aprovados foi 77.

Após a divulgação dos resultados, o professor verificou que uma questão havia sido mal formulada e decidiu atribuir 5 pontos a mais para todos os alunos. Com essa decisão, a média dos aprovados passou a ser 80 e a dos reprovados 68,8.

a) Calcule a média aritmética das notas da classe toda antes da atribuição dos 5 pontos extras.

b) Com a atribuição dos 5 pontos extras, quantos alunos, inicialmente reprovados, atingiram nota para aprovação?

332. A média aritmética dos números $x_1, x_2, ..., x_n$ é p. Determine a média aritmética dos números $x_1 - 1, x_2 - 1, x_3 - 1, x_4 - 1, ..., x_n + (-1)^n$, considerando que:

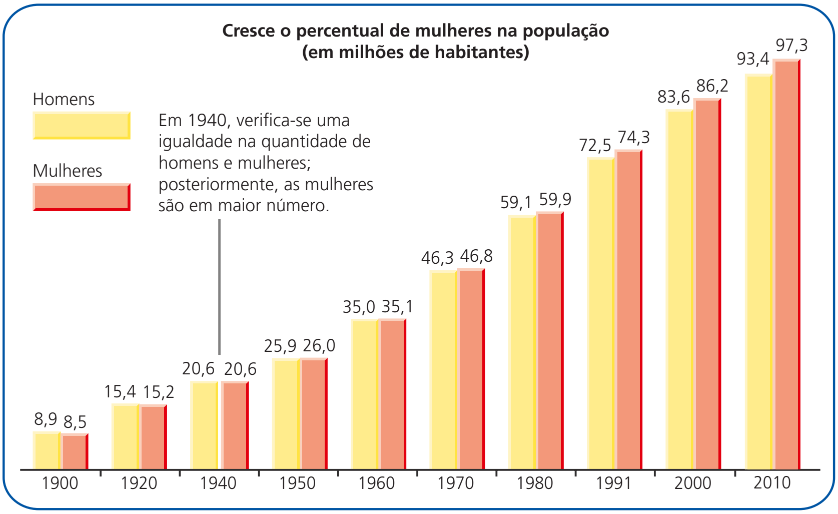
a) n é par;

b) n é ímpar.

333. (UF-GO) Em um time de futebol, o jogador mais velho entre os onze titulares foi substituído por um jogador de 16 anos. Isso fez com que a média de idade dos 11 jogadores diminuísse 2 anos.

Calcule a idade do jogador mais velho que foi substituído.

334.



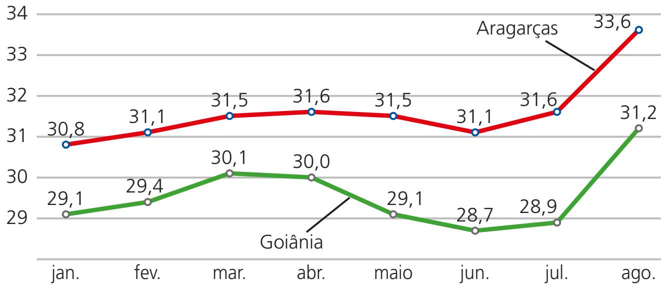
Fonte: *Almanaque Abril/2012*.

Calcule o percentual da população feminina e da população masculina relativo a cada ano constante no gráfico. Em seguida, utilizando apenas uma casa após a vírgula, determine, relativamente a cada sexo:

a) a média desses percentuais no período considerado;

b) a média desses percentuais de 1940 a 2010.

335. (UF-GO) O gráfico abaixo representa as temperaturas médias mensais nas cidades de Goiânia e Aragarças (considerada a cidade mais quente do Estado de Goiás), no período de janeiro a agosto de 2001.



Fonte: *O Popular*, 11 set. 2001.

Com base nesse gráfico, julgue como verdadeira (V) ou falsa (F) cada uma das afirmações a seguir:

a) Em Goiânia, a temperatura média no mês de agosto é 4% superior à temperatura média no mês de abril.

b) Em Goiânia, a média das temperaturas médias mensais no período de janeiro a agosto é igual à temperatura média do mês de junho.

c) No período de janeiro a agosto, a amplitude (diferença entre o maior e o menor valor) da temperatura média mensal, em Goiânia, é maior do que em Aragarças.

d) No período de janeiro a agosto, a diferença das temperaturas médias mensais entre Aragarças e Goiânia é máxima no mês de maio.

336. (UF-MS) Suprimindo-se um dos elementos do conjunto {1, 2, 3, ..., 201}, a média aritmética dos elementos restantes é 101,45. Sendo *m* o elemento suprimido, calcule o valor de m + 201.

337. Considere um conjunto de dados formado por *n* valores. Adicionando-se a esse conjunto o número 119, a média aumenta 4 unidades em relação à média inicial; retirando-se do conjunto original o número 54, a média diminui 1 unidade em relação à média inicial.

a) Qual é o valor de *n*?

b) Qual é a média aritmética inicial do conjunto de dados?

XII. Mediana

Em 2002, a população brasileira era constituída por aproximadamente 175 milhões de habitantes.

A área da superfície do território brasileiro é 8 514 204,8 km^2.

Assim, a densidade demográfica nesse ano era:

$$\frac{175 \text{ milhões de habitantes}}{8{,}514 \text{ milhões de km}^2} \cong 20{,}6 \text{ habitantes/km}^2$$

Na tabela seguinte, constam os valores (expressos em habitantes por km^2) das densidades demográficas dos 26 estados, além do Distrito Federal.

Estado	Densidade demográfica
Acre	3,7
Alagoas	101,3
Amapá	3,3
Amazonas	1,8
Bahia	23,2
Ceará	50,9
Distrito Federal	352,2
Espírito Santo	67,2
Goiás	14,7

Estado	Densidade demográfica
Maranhão	17,0
Mato Grosso	2,8
Mato Grosso do Sul	5,8
Minas Gerais	30,5
Pará	5,0
Paraíba	61,1
Paraná	48,0
Pernambuco	80,3
Piauí	11,3

Estado	Densidade demográfica
Rio de Janeiro	328,0
Rio Grande do Norte	52,2
Rio Grande do Sul	36,1
Rondônia	5,8
Roraima	1,5
Santa Catarina	56,1
São Paulo	149,0
Sergipe	81,1
Tocantins	4,2

Fonte: *Almanaque Abril*, 2002.

Calculando a média das densidades relacionadas anteriormente, encontramos:

$$\bar{x} = \frac{\sum_{i=1}^{27} x_i}{27} = \frac{1\,594,1}{27} \cong 59,04 \text{ habitantes/km}^2$$

Observe que esse valor é quase o triplo do valor encontrado para a densidade demográfica da população brasileira.

O cálculo da média aritmética ficou muito afetado por lugares com altíssima concentração populacional, como o Distrito Federal e o Estado do Rio de Janeiro, cujos valores – 352,2 e 328,0, respectivamente – destoam fortemente dos valores dos demais estados.

Calculemos a nova média, eliminando esses dois valores:

$$\bar{x}' = \frac{1\,594,1 - 352,2 - 328,0}{25} = \frac{913,9}{25} \cong 36,6 \text{ habitantes/km}^2$$

Observe que esse valor já está mais próximo ao correspondente à densidade demográfica brasileira.

Se eliminarmos o estado de São Paulo, que também tem uma alta densidade demográfica (149 habitantes/km^2), a nova média será:

$$\bar{x}'' = \frac{913,9 - 149}{24} = \frac{764,9}{24} \cong 31,9 \text{ habitantes/km}^2$$

Conforme podemos notar, esse novo valor está ainda mais próximo da real densidade demográfica brasileira.

Como vimos, a média aritmética pode ser muito afetada quando encontramos valores discrepantes em um conjunto de dados, podendo se tornar uma medida de centralidade pouco representativa do resumo dos dados.

Para contornar questões dessa natureza, definiremos, a seguir, uma medida de centralidade mais resistente aos valores discrepantes (em inglês, chamados *outliers*) denominada **mediana**.

Sejam $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ os n valores ordenados de uma variável x.

A **mediana** desse conjunto de valores – indicada por Me – é definida por:

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

Essa definição garante que a mediana seja um valor que divide o conjunto de dados em duas partes nas quais o número de elementos é o mesmo e de modo que o número de valores menores ou iguais à mediana seja igual ao número de valores maiores ou iguais a ela.

Exemplos:

1º) Vejamos como encontrar a mediana dos dados referentes à introdução sobre densidade demográfica.

É preciso inicialmente ordenar os valores (usaremos as siglas dos estados e a ordem crescente):

RR – AM – MT – AP – AC – TO – PA – RO – MS –
PI – GO – MA – BA – MG – RS – PR – CE – RN –
SC – PB – ES – PE – SE – AL – SP – RJ – DF

Como n é ímpar (n = 27), segue que:

$$\text{Me} = x_{\left(\frac{27+1}{2}\right)} = x_{14}$$

ou seja, a mediana é a densidade demográfica do 14º estado na sequência acima, que é Minas Gerais; portanto, Me = 30,5 habitantes/km^2. Note que essa medida de centralidade é mais representativa que a média (59,04 habitantes/km^2).

O cálculo da média só fica próximo ao da mediana quando eliminamos os estados com alta densidade demográfica (DF, RJ e SP).

2º) Os números seguintes indicam a quantidade de faltas de um aluno durante o ano letivo nas dez disciplinas do seu curso:

3 – 4 – 9 – 6 – 3 – 8 – 2 – 4 – 5 – 6

Para encontrar o número mediano de faltas do aluno, ordenamos esses valores:

$$2 - 3 - 3 - 4 - \boxed{4 - 5} - 6 - 6 - 8 - 9$$

Como *n* é par (n = 10), temos:

$$Me = \frac{x_5 + x_6}{2} = \frac{4 + 5}{2} = 4{,}5 \text{ faltas}$$

XIII. Moda

Seja *x* uma variável quantitativa que assume os valores $x_1, x_2, \ldots, x_k$, com frequências absolutas iguais a $n_1, n_2, \ldots$, nk, respectivamente. Se o máximo entre $n_1, n_2, \ldots, n_k$ é igual a n_j, $j \in \{1, 2, \ldots, k\}$, dizemos que a moda – indicada por Mo – é igual ao valor x_j.

Ou seja:

> A moda de um conjunto de valores corresponde ao valor que ocorre mais vezes.

Exemplos:

Vamos determinar a moda dos seguintes conjuntos de valores.

1º) 6 – 9 – 12 – 9 – 4 – 5 – 9

A moda é Mo = 9, pois há três valores iguais a 9.

2º) 12 – 13 – 19 – 13 – 14 – 12 – 16

Há duas modas, 12 e 13, pois cada um desses valores ocorre com maior frequência (duas vezes). Dizemos que se trata de uma distribuição **bimodal**.

3º) 4 – 29 – 15 – 13 – 18 – 20 – 21 – 26 – 9

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

EXERCÍCIOS

338. Calcule a moda e a mediana de cada um dos seguintes conjuntos de valores:

a) 9 – 8 – 8 – 7 – 10 – 12 – 11 – 8 – 8 – 7 – 6 – 14 – 10

b) 0 – 0 – 0 – 1 – 1 – 1 – 1 – 2 – 2 – 2 – 2 – 2 – 3 – 3 – 3 – 3 – 3 – 3

c) 40 – 44 – 42 – 23 – 36 – 40

d) 0,6 – 0,7 – 0,7 – 0,5 – 0,8 – 0,6 – 0,4 – 0,9

339. Determine as medidas de centralidade (média, mediana e moda) correspondentes aos percentuais relacionados na tabela a seguir:

Os 20 municípios com menor taxa de analfabetismo no Brasil (%)

	Município	Taxa de analfabetismo
1º	São João do Oeste (SC)	0,9
2º	Morro Reuter (RS)	1,6
3º	Harmonia (RS)	1,8
4º	Pomerode (SC)	1,9
5º	Bom Princípio (RS)	1,9
6º	São Vendelino (RS)	1,9
7º	Feliz (RS)	1,9
8º	Lagoa dos Três Cantos (RS)	2,0
9º	Salvador das Missões (RS)	2,2
10º	Ivoti (RS)	2,3
11º	Quatro Pontes (PR)	2,4
12º	Vale Real (RS)	2,5
13º	Timbó (SC)	2,6
14º	Dois Irmãos (RS)	2,6
15º	Jaraguá do Sul (SC)	2,6
16º	São José do Hortêncio (RS)	2,7
17º	Teutônia (RS)	2,7
18º	Blumenau (SC)	2,8
19º	Linha Nova (RS)	2,8
20º	Nova Petrópolis (RS)	2,8

Fonte: *O Estado de S. Paulo*, 5 jun. 2003.

340. A tabela seguinte relaciona os países com maior consumo anual de peixe.

Os maiores consumidores		
	País	**Quantidade de peixe consumido (milhões de toneladas)**
1º	China	30
2º	Japão	8
3º	Estados Unidos	6
4º	Índia	4
5º	Indonésia	4
6º	Rússia	3
7º	Coreia do Sul	2
8º	Filipinas	2
9º	França	2
10º	Espanha	2

Fonte: *Veja*, 9 jul. 2003.

a) Calcule a média, a mediana e a moda dos dados apresentados. Por que a média é bem maior que as outras duas medidas?

b) Sabendo que a população da China é 1,285 bilhão de habitantes e a da Espanha é 39,9 milhões de habitantes, mostre que o consumo *per capita* anual na Espanha é maior que o dobro do consumo *per capita* na China. (Dados extraídos de: *Almanaque Abril*, 2002.)

341. As tabelas seguintes informam o número de jornais diários em circulação na região metropolitana das capitais brasileiras.

Cidade	Jornais em circulação
Aracaju	3
Belém	3
Belo Horizonte	6
Boa Vista	3
Brasília	2
Campo Grande	2
Cuiabá	3
Curitiba	8
Florianópolis	3

Cidade	Jornais em circulação
Fortaleza	4
Goiânia	2
João Pessoa	3
Macapá	2
Maceió	3
Manaus	4
Natal	3
Palmas	3
Porto Alegre	3

Cidade	Jornais em circulação
Porto Velho	3
Recife	4
Rio Branco	4
Rio de Janeiro	11
Salvador	4
São Luís	2
São Paulo	16
Teresina	5
Vitória	2

Fonte: *Almanaque Abril*, 2002.

a) Intuitivamente, responda: Qual medida de centralidade – a média ou a mediana – é mais adequada para representar esses valores?

b) Calcule a média, a moda e a mediana.

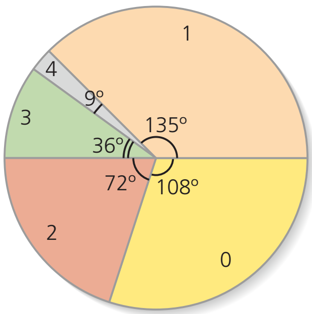
c) Elimine os dois estados com maior número de jornais e recalcule a média.

342. Um instituto de pesquisa fez um levantamento dos preços por quilo de vários produtos em um sacolão. Os resultados estão na tabela ao lado.

Qual é a média, a moda e a mediana do preço por quilo dos produtos à venda nesse sacolão?

Preço (em reais)	Frequência (%)
2,00	30
3,00	40
4,00	20
6,00	10

343. O gráfico abaixo informa a distribuição do número de filhos de 800 funcionários de uma empresa.



a) Quantos funcionários têm exatamente 2 filhos?

b) Qual é a mediana do número de filhos?

c) Qual é a moda do número de filhos?

344. A tabela ao lado informa o número de defeitos, por peça, encontrados durante uma inspeção feita em um lote de 80 peças que chegou a um porto.

Número de defeitos por peça	Número de peças
0	12
1	20
2	24
3	16
4	8

a) Considerando o número de defeitos por peça, qual é a mediana dos valores encontrados?

b) Qual será a nova mediana se forem acrescentadas a esse lote 18 peças, cada uma com exatamente 1 defeito?

c) Adicionando-se ao lote inicial n peças, cada uma com 3 defeitos, o valor da mediana passa a ser 3. Qual é o menor valor possível de n?

345. Os valores ordenados abaixo referem-se ao número de desistências mensais de reservas solicitadas a uma companhia aérea.

$$48 - 52 - 58 - 63 - 68 - x - 76 - 82 - y - 96 - 98 - 102$$

a) Sabendo que a mediana desses valores é 73 e que a média é 75, quais são os valores de x e de y?

b) Supondo que em cada um dos 5 meses seguintes o número de desistências tenha variado entre 50 e 60, qual será o valor da mediana relativa a esses 17 meses?

346. Considere a sequência decrescente:

$$2^n, 2^{n-1}, \ldots, 2^{n-5} \text{ (em que } n \text{ é um número natural)}$$

Sabendo que a mediana dos elementos dessa sequência é 6, determine:

a) o valor de n;

b) a média aritmética dos elementos dessa sequência.

347. (UnB-DF) A tabela adiante apresenta o levantamento das quantidades de peças defeituosas para cada lote de 100 unidades fabricadas em uma linha de produção de autopeças durante um período de 30 dias úteis.

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Nº de peças defeituosas	6	4	3	4	2	4	3	5	1	2	1	5	4	1	3

Dia	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Nº de peças defeituosas	7	5	6	4	3	2	6	3	5	2	1	3	2	5	7

Considerando S a série numérica de distribuição de frequências de peças defeituosas por lote de 100 unidades, julgue os itens a seguir:

1) A moda da série S é 5.
2) Durante o período de levantamento desses dados, o percentual de peças defeituosas ficou, em média, abaixo de 3,7%.
3) Os dados obtidos nos 10 primeiros dias do levantamento geram uma série numérica de distribuição de frequências com a mesma mediana da série S.

348. Uma pesquisa realizada com 280 pessoas fez o levantamento da frequência anual de visitas ao dentista. Os resultados aparecem na tabela ao lado.

Responda:

a) Qual é o número mediano de visitas?

b) Quantas pessoas dessa amostra que visitam o dentista uma única vez por ano deveriam passar a visitá-lo duas vezes por ano a fim de que a mediana passasse a ser 1,5 visita?

Número de visitas ao dentista por ano	Número de pessoas
0	63
1	105
2	39
3	47
4	16
5 ou mais	10
Total	**280**

XIV. Variância

Em certo país, o governo financia um programa de assistência às famílias de baixa renda. Cada família recebe, de cinco em cinco semanas, a quantia de 100 UM (unidades monetárias) para comprar produtos de alimentação em estabelecimentos conveniados. O coordenador desse projeto selecionou em uma pequena cidade quatro famílias e acompanhou a distribuição dos gastos semana a semana.

Observe a tabela:

	1ª semana	2ª semana	3ª semana	4ª semana	5ª semana	Total (valor do benefício)
Família I	20 UM	20 UM	20 UM	20 UM	20 UM	100 UM
Família II	20 UM	24 UM	20 UM	16 UM	20 UM	100 UM
Família III	12 UM	28 UM	24 UM	20 UM	16 UM	100 UM
Família IV	36 UM	32 UM	20 UM	8 UM	4 UM	100 UM

Como cada família gasta 100 UM no período de cinco semanas, a média semanal de gastos é $\frac{100}{5} = 20$ UM.

A leitura dos valores da tabela mostra um comportamento diferente de cada família na utilização do benefício concedido pelo governo: a família I, por exemplo, gasta sempre a mesma quantia por semana para comprar alimentos; já a família IV faz gastos que oscilam entre 4 e 36 UM por semana.

Desse modo, se essa análise for limitada à média semanal de gastos, estarão sendo omitidas informações importantes em relação à homogeneidade ou heterogeneidade dos gastos semanais de cada família. Para revelar o grau de variabilidade de um conjunto de dados há necessidade de uma medida específica, a **variância**, definida a seguir.

Seja *x* uma variável quantitativa que assume os valores $x_1, x_2, \ldots, x_n$ e $\bar{x}$ a média aritmética correspondente a esses valores.

A variância desses valores – indicada por Var(x) ou σ^2 – é definida por:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2}{n}$$

Notemos que cada termo do numerador corresponde ao quadrado da diferença entre um valor observado e o valor médio. Essa diferença traduz o quanto um valor observado se distancia do valor médio, sendo, portanto, uma medida do grau de variabilidade dos dados em estudo.

Vamos calcular a variância dos gastos semanais das quatro famílias (lembre que a média semanal, para cada família, é 20 UM).

- Família I:

$$\sigma^2 = \frac{(20-20)^2 + (20-20)^2 + (20-20)^2 + (20-20)^2 + (20-20)^2}{5} = 0$$

- Família II:

$$\sigma^2 = \frac{(20-20)^2 + (24-20)^2 + (20-20)^2 + (16-20)^2 + (20-20)^2}{5} = \frac{16+16}{5} =$$

$$= 6{,}4 \text{ UM}^2$$

- Família III:

$$\sigma^2 = \frac{(12-20)^2 + (28-20)^2 + (24-20)^2 + (20-20)^2 + (16-20)^2}{5} =$$

$$= \frac{64+64+16+16}{5} = 32 \text{ UM}^2$$

• Família IV:

$$\sigma^2 = \frac{(36-20)^2 + (32-20)^2 + (20-20)^2 + (8-20)^2 + (4-20)^2}{5} =$$

$$= \frac{256 + 144 + 144 + 256}{5} = 160 \text{ UM}^2$$

O aumento no valor da variância nesses cálculos revela uma variabilidade crescente de gastos semanais em relação à média (20 UM).

Nessa situação, a **unidade de variância** é UM^2 e o **gasto semanal médio** é expresso em UM, o que gera uma incompatibilidade. Para uniformizar as unidades, definiremos mais adiante o desvio padrão σ.

12. Propriedades:

Como a média aritmética, a variância também apresenta duas propriedades importantes.

Seja *x* uma variável quantitativa que assume os valores $x_1, x_2, ..., x_n$. Considere $\bar{x}$ a média aritmética e σ^2 a variância correspondente.

> Se a cada x_i $(i = 1, 2, ..., n)$ for adicionada uma constante real *c*, a variância não se altera.

Essa propriedade pode ser demonstrada da seguinte maneira.

Sendo $(\sigma^2)'$ a nova variância, mostremos que $(\sigma^2)' = \sigma^2$.

Consideremos $x'_1 = x_i + c$ $(i = 1, 2, ..., n)$ os novos valores assumidos pela variável.

Conforme vimos no item **Média aritmética**, a nova média $\bar{x}'$ é dada por $\bar{x}' = \bar{x} + c$.

Da definição de variância, segue:

$$(\sigma^2)' \frac{\sum_{i=1}^{n} (x'_i - \bar{x}')^2}{n} = \frac{\sum_{i=1}^{n} [(x_i + c) - (\bar{x} + c)^2}{n} = \frac{\sum_{i=1}^{n} (x_i + c - \bar{x} - c)^2}{n} =$$

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

> Se cada x_i $(i = 1, 2, ..., n)$ for multiplicado por uma constante real *c*, a variância fica multiplicada por c^2.

Vamos demonstrar essa segunda propriedade.

Sendo $(\sigma^2)'$ a nova variância, os novos valores que a variável x assume são:

$$x'_1 = c \cdot x_i \ (i = 1, 2, \ldots, n), \text{ a saber: } c \cdot x_i, c \cdot x_2, \ldots, c \cdot x_n$$

De acordo com o item **Média aritmética**, a nova média $\overline{x'}$ é dada por $\overline{x'} = c \cdot \overline{x}$.

Temos:

$$(\sigma^2)' \ \frac{\sum\limits_{i=1}^{n} (x'_i - \overline{x}')^2}{n} = \frac{\sum\limits_{i=1}^{n} (c \cdot x_i - c \cdot \overline{x})^2}{n} = \frac{\sum\limits_{i=1}^{n} c^2 \cdot (x_i - \overline{x})^2}{n} =$$

$$= c^2 \cdot \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n} = c^2 \cdot \sigma^2$$

XV. Desvio padrão

Sejam $x_1, x_2, \ldots, x_n$ os valores assumidos por uma variável x. Chamamos **desvio padrão** de x – indicado por DP(x) ou σ – a raiz quadrada da variância de x.

$$\sigma = \sqrt{\frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n}} = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n}}$$

Exemplo:

Na situação considerada na introdução do estudo da variância (ver p. 126), o desvio padrão dos gastos de cada família é dado por:

- família I: $\sigma^2 = 0 \ \Rightarrow \sigma = 0$ UM
- família II: $\sigma^2 = 6{,}4 \text{ UM}^2 \ \Rightarrow \sigma = \sqrt{6{,}4 \text{ UM}^2} \cong 2{,}53$ UM
- família III: $\sigma^2 = 32 \text{ UM}^2 \ \Rightarrow \sigma = \sqrt{32 \text{ UM}^2} \cong 5{,}66$ UM
- família IV: $\sigma^2 = 160 \text{ UM}^2 \ \Rightarrow \sigma = \sqrt{160 \text{ UM}^2} \cong 12{,}65$ UM

Observação:

Das duas propriedades descritas para a variância (ver p. 128), decorrem as seguintes consequências imediatas:

1ª) Quando adicionamos uma constante a cada elemento de um conjunto de valores, o desvio padrão não se altera.

2ª) Quando multiplicamos cada elemento de um conjunto de valores por uma constante real c, o desvio padrão fica multiplicado por c.

13. Outra expressão para variância e para desvio padrão

É possível encontrar para variância e desvio padrão outras expressões equivalentes às das definições apresentadas e que poderão ser úteis na resolução de alguns exercícios.

Vejamos:

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

Desenvolvendo o produto notável, temos:

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

Como $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$, podemos escrever:

$$\sigma^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^{n} x_i^2 - 2 \cdot \frac{\sum_{i=1}^{n} x_i}{n} \cdot \sum_{i=1}^{n} x_i + n \cdot \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n^2} \right] \Rightarrow$$

$$\Rightarrow \sigma^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^{n} x_i^2 - \frac{2}{n} \left(\sum_{i=1}^{n} x_i \right)^2 + \frac{1}{n} \cdot \left(\sum_{i=1}^{n} x_i \right)^2 \right]$$

isto é:

$$\boxed{\sigma^2 = \frac{1}{n} \cdot \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n} \right]}$$

, que é expressão para a variância.

Como o desvio padrão corresponde à raiz quadrada da variância, temos:

$$\boxed{\sigma^2 = \sqrt{\frac{1}{n} \cdot \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i \right)^2}{n} \right]}}$$

, que é expressão para o desvio padrão.

Exemplo:

Os dados seguintes referem-se aos gastos mensais com ônibus e metrô (expressos em reais) que um estudante universitário tem durante um semestre:

42 – 50 – 54 – 48 – 56 – 59

Aplicando a expressão anterior para o cálculo do desvio padrão (σ) das despesas, temos:

$$\sum_{i=1}^{6} x_i = 42 + 50 + 54 + 48 + 56 + 59 = 309$$

e

$$\sum_{i=1}^{6} x_i^2 = 42^2 + 50^2 + 54^2 + 48^2 + 56^2 + 59^2 = 16\,101$$

Daí:

$$\sigma^2 = \frac{1}{6} \cdot \left[16\,101 - \frac{309^2}{6}\right] = 31{,}25\ (\text{reais})^2 \Rightarrow \sigma \cong 5{,}59 \text{ reais}$$

14. Observação geral sobre variância e desvio padrão

As expressões das medidas de dispersão apresentadas no item **Variância** referem-se à **variância** (e **desvio padrão**) **populacional**.

Nos casos em que os dados são coletados a partir de uma amostra da população, obtém-se como medida de dispersão a chamada **variância amostral**, representada por S^2 e dada por:

$$\boxed{S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}$$, sendo $\bar{x}$ a média amostral.

Observe que as expressões que definem σ^2 e S^2 diferem apenas pelo denominador: σ^2 apresenta como denominador *n* e S^2 apresenta como denominador $n - 1$. O motivo dessa diferença exige conhecimentos mais aprofundados que os fornecidos nesta obra introdutória.

Fica, então, convencionado que, nos exercícios seguintes, salvo observações contrárias, deve-se considerar sempre a **variância populacional** σ^2.

EXERCÍCIOS

349. Calcule o desvio padrão dos seguintes conjuntos de valores:

a) 2 – 3 – 4 – 5 – 6

b) 2 – 2 – 3 – 4 – 4

c) (–2) – (–1) – (–1) – 0 – 1 – 3

d) $\frac{1}{2} - \frac{1}{8} - \frac{1}{4} - \frac{1}{5} - \frac{1}{10}$

e) 70 – 65 – 60 – 60 – 65 – 68 – 72 – 60

350. A tabela seguinte informa a participação percentual dos estados da região Nordeste no produto interno bruto (PIB) nacional.

Alagoas	0,9%	Maranhão	1,0%	Piauí	0,5%
Bahia	4,4%	Paraíba	0,7%	Rio Grande do Norte	0,9%
Ceará	1,8%	Pernambuco	2,3%	Sergipe	0,5%

Fonte: *Almanaque Abril*, 2002.

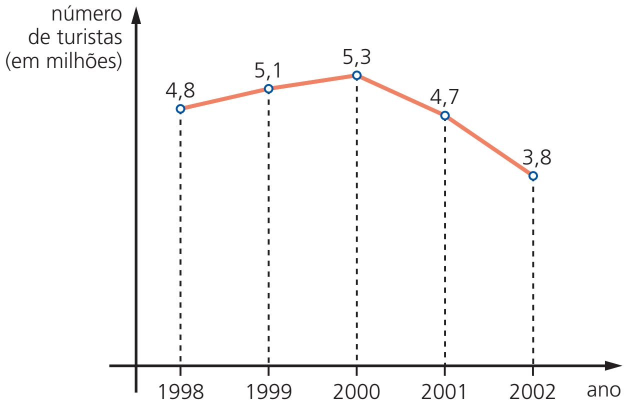
a) Calcule a média ($\overline{x}$) e o desvio padrão (σ) dos percentuais acima.

b) Quantos estados têm participação pertencente ao intervalo

$$\left[\overline{x} - \frac{1}{2}\sigma, \overline{x} + \frac{1}{2}\sigma\right]?$$

351. O gráfico ao lado mostra os números relativos aos turistas estrangeiros que estiveram no Brasil no período de 1998 a 2002.

Qual é o desvio padrão dos dados apresentados?



Fonte: *Veja*, 16/4/2003.

352. Os dados seguintes referem-se às porcentagens da população de países sul-americanos que vivem em áreas urbanas.

Argentina	90%
Bolívia	63%
Brasil	81%
Chile	86%
Colômbia	74%

Equador	65%
Paraguai	56%
Peru	73%
Uruguai	91%
Venezuela	87%

Fonte: *Almanaque Abril*, 2002.

a) Calcule a média e o desvio padrão dos percentuais acima.

b) Elimine os dois países com menores percentuais. O que ocorrerá com o desvio padrão? Faça os cálculos para confirmar sua resposta.

353. Um conjunto é formado por três elementos: 8, 10 e *x*. Determine os possíveis valores de *x* para os quais a variância desses elementos é igual a $\frac{26}{3}$.

354. Sejam $x_1, x_2, \ldots, x_k$ os *k* valores distintos assumidos por uma variável *x*, com frequências absolutas iguais a $n_1, n_2, \ldots, n_k$, respectivamente. Encontre uma expressão para a variância desses valores.

Solução

Os valores estão assim distribuídos:

$$\underbrace{x_1, x_1, \ldots, x_1}_{n_1 \text{ vezes}}, \underbrace{x_2, x_2, \ldots, x_2}_{n_2 \text{ vezes}}, \ldots, \underbrace{x_k, x_k, \ldots, x_k}_{n_k \text{ vezes}}$$

A média aritmética é dada por:

$$\overline{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n_1 + n_2 + \ldots + n_k}$$

Usando a definição de variância:

$$\sigma^2 = \frac{1}{\sum\limits_{i=1}^{k} n_i} \Big\{ \underbrace{(x_1 - \overline{x})^2 + \ldots + (x_1 - \overline{x})^2}_{n_1 \text{ vezes}} + \underbrace{(x_2 - \overline{x})^2 + \ldots + (x_2 - \overline{x})^2}_{n_2 \text{ vezes}} + \ldots +$$

$$+ \underbrace{(x_k - \overline{x})^2 + \ldots + (x_k - \overline{x})^2}_{n_k \text{ vezes}} \Big\}$$

$$\sigma^2 = \frac{1}{\sum\limits_{i=1}^{k} n_i} \cdot \left\{ n_1 \cdot (x_1 - \overline{x})^2 + n_2 \cdot (x_2 - \overline{x})^2 + \ldots + n_k \cdot (x_k - \overline{x})^2 \right\}$$

que é expressão da variância em função da frequência absoluta.

Temos também:

$$\sigma^2 = f_1 \cdot (x_1 - \bar{x})^2 + f_2 \cdot (x_2 - \bar{x})^2 + \ldots + f_k \cdot (x_k - \bar{x})^2$$

em que $f_1, f_2, \ldots, f_k$ são as frequências relativas correspondentes.

355. A tabela ao lado informa a distribuição do número de cartões amarelos recebidos por um time durante os 45 jogos de um torneio:

Calcule o desvio padrão referente ao número de cartões recebidos.

Número de cartões	Número de jogos
0	5
1	19
2	10
3	7
4	4

Solução

Temos:

- $\bar{x} = \dfrac{0 \cdot 5 + 1 \cdot 19 + 2 \cdot 10 + 3 \cdot 7 + 4 \cdot 4}{5 + 19 + 10 + 7 + 4} = \dfrac{76}{45} \cong 1,69$
- $\sigma^2 = \dfrac{1}{45} \cdot [(0 - 1,69)^2 \cdot 5 + (1 - 1,69)^2 \cdot 19 + (2 - 1,69)^2 \cdot 10 + (3 - 1,69)^2 \cdot 7 +$

$+ (4 - 1,69)^2 \cdot 4]$

$\sigma^2 = \dfrac{1}{45} \cdot [14,28 + 9,05 + 0,96 + 12,01 + 21,34] \Rightarrow \sigma^2 = \dfrac{57,64}{45} = 1,28$

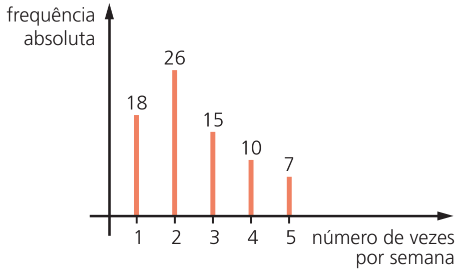
$\Rightarrow \sigma \cong 1,13$ cartão

356. Um professor aplicou um exercício em sua turma de 60 alunos e as notas possíveis eram zero, 0,5 ponto ou 1 ponto. Sabendo que 40% dos alunos não obtiveram pontuação, 35% conseguiram 0,5 ponto e o restante atingiu a pontuação máxima, determine:

a) a mediana dos pontos obtidos pelos alunos nessa atividade;

b) a variância correspondente aos pontos obtidos pelos alunos.

357. A Secretaria de Saúde de uma cidade está interessada em saber com que frequência semanal seus habitantes praticam atividades físicas. Para isso, uma equipe entrevistou *n* pessoas e os resultados encontram-se no gráfico a seguir:



a) Determine o valor de *n*.

b) Qual é a média das frequências de atividades físicas?

c) Qual é a moda e a mediana dos dados obtidos?

d) Qual é o desvio padrão dos dados obtidos?

358. Um conjunto de dados possui *n* valores ($n > 3$), dos quais três são iguais a 2 e os demais iguais a 5.

a) Determine, em função de *n*, a média aritmética desses elementos.

b) Determine o maior **valor inteiro** de *n* para o qual a variância desse conjunto de valores seja maior que 2.

359. (Unicamp-SP) Para um conjunto $X = \{x_1, x_2, x_3, x_4\}$, a média aritmética de *X* é definida por $\bar{x} = \dfrac{x_1 + x_2 + x_3 + x_4}{4}$ e a variância de X é definida por $v = \dfrac{1}{4}\left[(x_1 - \bar{x})^2 + \ldots + (x_4 - \bar{x})^2\right]$.

Dado o conjunto $X = \{2, 5, 8, 9\}$, pede-se:

a) Calcular a média aritmética de X.

b) Calcular a variância de X.

c) Quais elementos de X pertencem ao intervalo $\left[\bar{x} - \sqrt{v}, \bar{x} + \sqrt{v}\right]$?

360. (FGV-SP) Dados *n* valores $x_1, x_2, x_3, \ldots, x_n$, seja M sua média aritmética. Chama-se variância desses valores ao número σ^2 dado por:

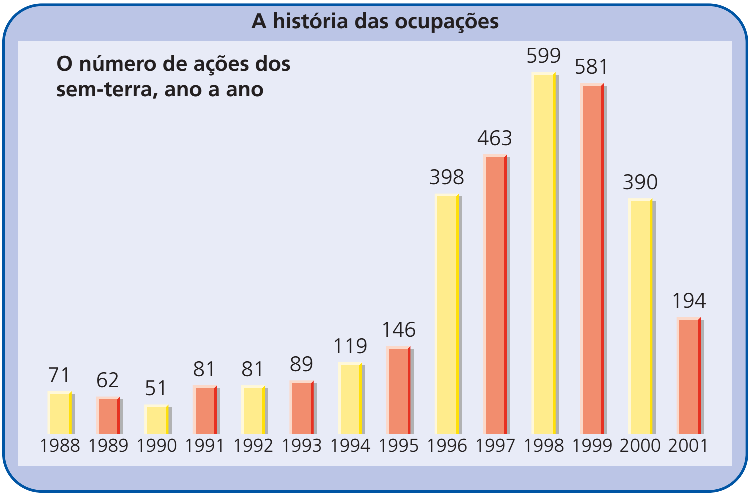
$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - M)^2}{n}$$

A raiz quadrada não negativa da variância chama-se desvio padrão.

a) Se em cada um de 10 meses consecutivos um fundo de investimentos render 1% ao mês, qual o desvio padrão dessas taxas de rendimento?

b) Se em cada um de 6 meses consecutivos o fundo render 1% ao mês e render 3% ao mês em cada um dos quatro meses seguintes, qual o desvio padrão dessas taxas de rendimento?

361. Observe os dados apresentados pelo gráfico:



Fonte: *O Estado de S. Paulo*, 6/3/2003.

a) Encontre o desvio padrão correspondente ao número anual de ações registradas no período de 1988 a 2001.

b) Considere os períodos de 1988 a 1993 e de 1996 a 2000. Calcule o desvio padrão correspondente a cada período. Por que se observa uma queda em relação ao desvio encontrado no item *a*?

362. O Departamento de Aviação Civil registrou durante cinco dias o percentual diário de voos de duas companhias aéreas, A e B, que decolaram sem atraso. Os dados estão relacionados a seguir:

Companhia A:

90% – 92% – 95% – 88% – 91%

Companhia B:

97% – 88% – 98% – 86% – 90%

a) Qual companhia apresentou percentual médio mais alto?

b) Qual companhia apresentou desempenho mais regular?

363. Seja o conjunto de valores 4, 1, 8, 7 e n. Qual é o valor de n que minimiza a variância desses valores? Qual é, nesse caso, o valor da variância?

364. Considere os seguintes conjuntos de valores:

$$A = \{3, 3, 3, 3, 4, 4\} \quad \text{e} \quad B = \{2, 2, 3, 3, 4, 4\}$$

Compare σ_A^2 com σ_B^2.

365. A tabela de frequências ao lado informa o número de filhos dos 80 funcionários de uma escola.

Número de filhos	Frequência absoluta
0	20
1	36
2	14
3	8
4	2

a) Qual é o desvio padrão correspondente ao número de filhos?

b) Suponha que cada funcionário dessa escola tenha um novo filho. Qual será o novo desvio padrão?

366. Sejam $x_1, x_2, ..., x_n$ os valores assumidos por uma variável x, e σ^2 a variância correspondente a tais valores. Determine a relação existente entre a nova variância $(\sigma^2)'$ e a variância original (σ^2) quando:

a) cada x_i ($i = 1, 2, ..., n$) é multiplicado por 2;

b) a cada x_i ($i = 1, 2, ..., n$) são adicionadas 3 unidades;

c) cada x_i ($i = 1, 2, ..., n$) é dividido por 5;

d) cada x_i é multiplicado por 4 e ao valor obtido são adicionadas 4 unidades;

e) de cada x_i subtraem-se 10 unidades.

367. Os saldos (x_i) em cadernetas de poupança de 1 000 clientes de um banco em uma pequena cidade são tais que:

$$\sum_i x_i = 322\,000 \quad \text{e} \quad \sum_i x_i^2 = 119\,309\,000, \text{ para } i \in \{1, 2, ..., 1\,000\}$$

a) Determine o saldo médio das cadernetas.

b) Qual é o desvio padrão correspondente aos saldos das cadernetas?

368. Uma pastelaria situada no centro de uma grande cidade funciona os sete dias da semana. Em certa semana, a receita média diária era R$ 1 200,00 e a soma dos quadrados das receitas diárias totalizava R$ 10 086 300,00. Qual foi o desvio padrão da receita diária registrada nessa semana?

369. Que número deve ser acrescentado ao conjunto de valores 2, 6, 5 e 7 a fim de que a variância aumente de 3,3 unidades?

370. Sejam $x_1, x_2, ..., x_n$ valores assumidos por uma variável, $\bar{x}$ a média aritmética e σ o desvio padrão. Suponha que de cada x_i ($i = 1, 2, ..., n$) subtraímos a média e dividimos a diferença obtida pelo desvio padrão. Qual será a nova média e o novo desvio padrão desse conjunto?

371. (UF-GO) Dados os números $a_1, a_2, ..., a_n$ e considerando a média aritmética M(x) dos n números $(a_1 - x)^2, (a_2 - x)^2, ..., (a_n - x)^2$, em que x é um número real qualquer:

a) determine x de modo que a média M(x) seja mínima;

b) determine o valor mínimo da média M(x), que é chamado de variância de $a_1, a_2, ..., a_n$.

372. Os dados seguintes referem-se à mortalidade infantil dos estados da região Nordeste e indicam o número de crianças que morrem no primeiro ano de vida entre 1 000 crianças nascidas vivas.

Alagoas	64,4	Maranhão	52,8	Piauí	44,4
Bahia	44,7	Paraíba	59,4	Rio Grande do Norte	47,9
Ceará	51,6	Pernambuco	57,5	Sergipe	44,5

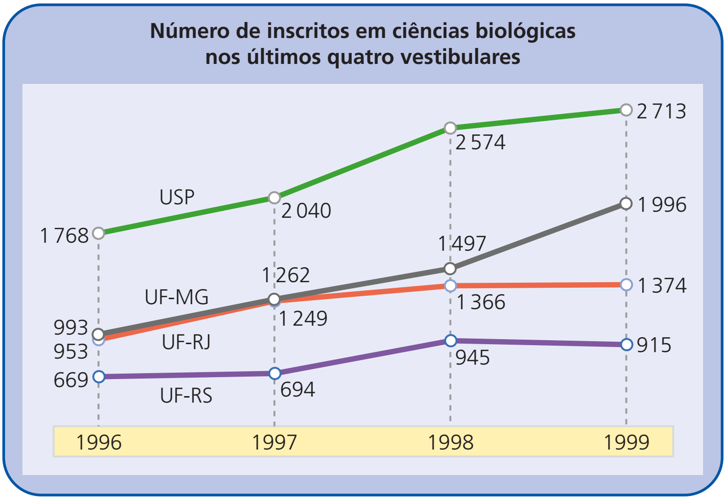
Fonte: *Almanaque Abril*, 2002.

a) Encontre a variância dos dados apresentados.

b) Admitindo-se que um programa do governo consiga reduzir 15% das taxas de mortalidade apresentadas, qual será o novo valor da variância?

373. (UnB-DF) Um novo *boom* desponta nas estatísticas dos últimos vestibulares. Desde o surgimento de Dolly, a polêmica ovelha clonada a partir da célula de um animal adulto, a carreira de ciências biológicas recebe cada vez mais candidatos e essa área firma-se como a ciência do próximo milênio.

O gráfico a seguir ilustra o número de inscritos nos últimos quatro vestibulares que disputaram as vagas oferecidas pela Universidade de São Paulo (USP) e pelas universidades federais do Rio de Janeiro (UF-RJ), de Minas Gerais (UF-MG) e do Rio Grande do Sul (UF-RS).



Adaptado de: *Época*, 26 abr. 1999.

Com base nessas informações, julgue os itens seguintes em V ou F, justificando:

1) De 1997 a 1998, o crescimento percentual do número de inscritos na USP foi maior que o da UF-RS.
2) Todos os segmentos de reta apresentados no gráfico têm inclinação positiva.
3) Durante todo o período analisado, a UF-MG foi a universidade que apresentou o maior crescimento percentual, mas não o maior crescimento absoluto.
4) Os crescimentos percentuais anuais na UF-RJ diminuíram a cada ano.
5) Considerando, para cada universidade representada no gráfico, a série numérica formada pelos números de inscritos em ciências biológicas nos últimos quatro vestibulares, a série da USP é a que apresenta a maior mediana, tendo desvio padrão maior que o da UF-RJ.

374. Sejam $x_1, x_2, \dots, x_n$ os *n* valores assumidos por uma variável quantitativa.
Uma medida de dispersão usual é o desvio médio, que é indicado por DM(x) e definido pela relação:

$$DM(x) = \frac{\sum\limits_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n}$$

Dado o conjunto de valores 2, 3, 5, 4 e 6, obtenha o desvio médio correspondente.

Solução

Temos:

- $\bar{x} = \dfrac{2 + 3 + 5 + 4 + 6}{5} = 4$
- $DM(x) = \dfrac{|2 - 4| + |3 - 4| + |5 - 4| + |4 - 4| + |6 - 4|}{5} = \dfrac{2 + 1 + 1 + 0 + 2}{5} = 1{,}2$

375. Calcule o desvio médio dos seguintes conjuntos de valores:

a) 9 – 10 – 10 – 10 – 10 – 12 – 12 – 15
b) 4 – 7 – 8 – 8 – 9 – 9 – 10 – 17
c) 3 – 3 – 3 – 4 – 4 – 5 – 6
d) 60 – 61 – 62 – 63 – 64

376. A expressão seguinte representa o numerador da expressão que define o desvio médio de uma variável:

$2 \cdot |8 - 10| + 3 \cdot |9 - 10| + n \cdot |10 - 10| + 2 \cdot |12 - 10| + m \cdot |13 - 10|$

a) Qual é a média dos valores dessa variável?
b) Se o desvio médio encontrado é 1,4, quais são os valores de *n* e *m*?

377. Os dados abaixo referem-se aos percentuais de matrículas feitas no ensino médio em escolas públicas nas regiões Sul e Sudeste.

Espírito Santo	85,3%
Minas Gerais	89,4%
Paraná	89,4%
Rio de Janeiro	79,6%

Rio Grande do Sul	86,0%
Santa Catarina	85,1%
Sergipe	86,6%

Fonte: *Almanaque Abril*, 2002.

a) Calcule o desvio médio desses percentuais.

b) Qual região, a Sul ou a Sudeste, apresenta dados mais homogêneos, considerando-se o desvio médio como medida de dispersão?

378. Responda:

a) O que aconteceria com o desvio médio se fossem retirados os módulos da definição?

b) Sejam $x_1, x_2, ..., x_n$ um conjunto de valores assumidos por uma variável. Mostre que a expressão do desvio médio e da variância coincidem quando todos os valores são iguais ou se $|x_i - \bar{x}| = 1, \forall i \in \{1, 2, ..., n\}$.

XVI. Medidas de centralidade e dispersão para dados agrupados

Em uma academia de ginástica deseja-se implantar um programa de racionamento de energia elétrica, que inclui, entre outras medidas, uma campanha de incentivo à redução do tempo de banho nos vestiários. Durante uma semana, registrou-se o tempo de duração dos banhos dos usuários.

Os dados coletados estão organizados na tabela:

Tempo de duração (em minutos)	Frequência absoluta
$1 \vdash 4$	18
$4 \vdash 7$	108
$7 \vdash 10$	270
$10 \vdash 13$	150
$13 \vdash 16$	54
Total	**600**

Como encontramos as medidas de centralidade (média, mediana e moda) e variabilidade (desvio padrão e variância) relativas a esses dados?

Quando as informações referentes a uma variável estão agrupadas em classes de valores (intervalos), não é possível saber como os valores estão distribuídos em cada faixa. Como recurso para associar medidas a esses dados, costuma-se fazer a suposição de que, em cada intervalo, os valores estão distribuídos homogeneamente, isto é, admite-se uma distribuição aproximadamente simétrica ao redor do ponto médio do intervalo. Assim, se um determinado intervalo contém *n* valores, há uma "compensação" entre valores equidistantes do ponto médio (x_i) da classe *i*, de modo que a média entre eles coincide com x_i.

Essas considerações nos levam a supor que as *n* observações do intervalo equivalem ao seu ponto médio.

15. Cálculo da média

Seja x_i o ponto médio de um determinado intervalo.

Retomando o exemplo da academia de ginástica que pretende implantar um programa de racionamento de energia elétrica, temos esta tabela:

Tempo de duração (em minutos)	Ponto médio (x_i)	Frequência absoluta (n_i)	Frequência relativa (f_i)
$1 \vdash 4$	$x_1 = 2,5$	$n_1 = 18$	$\frac{18}{600} = 0,03$
$4 \vdash 7$	$x_2 = 5,5$	$n_2 = 108$	$\frac{108}{600} = 0,18$
$7 \vdash 10$	$x_3 = 8,5$	$n_3 = 270$	$\frac{270}{600} = 0,45$
$10 \vdash 13$	$x_4 = 11,5$	$n_4 = 150$	$\frac{150}{600} = 0,25$
$13 \vdash 16$	$x_5 = 14,5$	$n_5 = 54$	$\frac{54}{600} = 0,09$

O tempo médio de banho é dado por:

$$\overline{x} = \frac{18 \cdot 2,5 + 108 \cdot 5,5 + 270 \cdot 8,5 + 150 \cdot 11,5 + 54 \cdot 14,5}{600} = 9,07 \text{ minutos}$$

(ou aproximadamente 9 minutos e 4 segundos)

Em geral, a média para dados agrupados é dada por:

$$\boxed{\bar{x} = \frac{\sum\limits_{i=1}^{k} x_i \cdot n_i}{\sum\limits_{i=1}^{k} n_i}}, \text{ sendo } \begin{cases} k \text{ o número de intervalos} \\ x_i \text{ o ponto médio da classe } i \\ n_i \text{ a frequência absoluta referente à classe } i \end{cases}$$

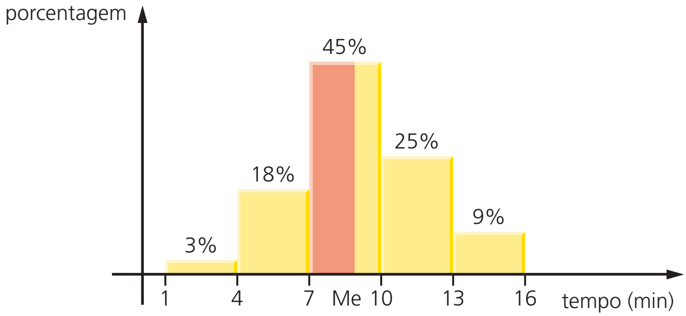
ou

$$\boxed{\bar{x} = \sum\limits_{i=1}^{k} x_i \cdot f_i}, \text{ sendo } f_i \text{ a frequência relativa referente à classe } i$$

16. Cálculo da mediana

Em variáveis contínuas que apresentam seus valores distribuídos em intervalos, admite-se que 50% dos dados encontram-se abaixo da mediana e 50% acima dela.

Nesses casos, para determinar a mediana, é importante, num primeiro momento, construir um histograma, usando a frequência relativa (ou porcentagem) de cada intervalo. Em relação ao exemplo da academia de ginástica, temos:



A mediana desse conjunto de dados é um valor pertencente ao intervalo 7 ⊢ 10, uma vez que a frequência acumulada das duas primeiras classes é 3% + 18% = 21% e das três primeiras classes é 3% + 18% + 45% = 66%.

Observe que, no terceiro intervalo, o retângulo sombreado e o retângulo “inteiro” (que define o intervalo) têm a mesma altura. Assim, a área de cada um desses retângulos (expressa como porcentagem da área total sob o histograma) é proporcional à medida de sua base.

Temos:

- retângulo sombreado $\begin{cases} \text{base: Me} - 7 \\ \text{área: } 50\% - 21\% \end{cases}$

• retângulo "inteiro" $\begin{cases} \text{base: } 10 - 7 \\ \text{área: } 45\% \end{cases}$

Segue, daí, a seguinte proporção:

$\dfrac{\text{Me} - 7}{50\% - 21\%} = \dfrac{3}{45\%} \Rightarrow \text{Me} \cong 8{,}93$ minutos (aproximadamente 8 minutos e 56 segundos)

17. Cálculo da classe modal

Suponha que os dados de uma variável contínua estejam distribuídos em classes de mesma amplitude.

A **classe modal** é dada pela classe que reúne a maior frequência (absoluta ou relativa).

No exemplo, a classe de maior frequência é a de 7 a 10 minutos, e ela concentra 270 valores (ou 45% dos dados da amostra).

Dizemos que a classe modal é o intervalo $7 \vdash 10$ (minutos).

18. Cálculo da variância e do desvio padrão

O cálculo da variância e do desvio padrão de uma variável que apresenta seus valores distribuídos em intervalos utiliza a mesma hipótese usada no cálculo da média: dentro de cada intervalo, os valores estão homogeneamente distribuídos.

Consideremos a situação de distribuição de salários de uma empresa com 200 funcionários, representada na tabela:

Faixa salarial (em salários mínimos)	Ponto médio (x_i)	Número de funcionários (frequência absoluta: n_i)
$2 \vdash 6$	4	45
$6 \vdash 10$	8	63
$10 \vdash 14$	12	36
$14 \vdash 18$	16	31
$18 \vdash 22$	20	17
$22 \vdash 26$	24	8

Temos:

- $\bar{x} = \dfrac{\sum_{i=1}^{6} x_i n_i}{\sum_{i=1}^{6} n_i} = \dfrac{4 \cdot 45 + 8 \cdot 63 + 12 \cdot 36 + 16 \cdot 31 + 20 \cdot 17 + 24 \cdot 8}{200} =$

$$= \frac{180 + 504 + 432 + 496 + 340 + 192}{200} = \frac{2\,144}{200} = 10{,}72 \text{ SM}$$

- Para cada intervalo, avaliamos o desvio quadrático do ponto médio correspondente em relação à média encontrada:

Intervalo	Ponto médio	Desvio quadrático
$2 \vdash 6$	4	$(4 - 10{,}72)^2 = 45{,}16$
$6 \vdash 10$	8	$(8 - 10{,}72)^2 = 7{,}39$
$10 \vdash 14$	12	$(12 - 10{,}72)^2 = 1{,}64$
$14 \vdash 18$	16	$(16 - 10{,}72)^2 = 27{,}88$
$18 \vdash 22$	20	$(20 - 10{,}72)^2 = 86{,}11$
$22 \vdash 26$	24	$(24 - 10{,}72)^2 = 176{,}36$

- Fazemos a média desses desvios, ponderando-os pelas frequências absolutas correspondentes, isto é:

$$\sigma^2 = \frac{45 \cdot 45{,}16 + 63 \cdot 7{,}39 + 36 \cdot 1{,}64 + 31 \cdot 27{,}88 + 17 \cdot 86{,}11 + 8 \cdot 176{,}36}{200}$$

$$\sigma^2 = \frac{2\,032{,}2 + 465{,}57 + 59{,}04 + 864{,}28 + 1\,463{,}87 + 1\,410{,}88}{200}$$

$$\sigma^2 = \frac{6\,295{,}84}{200} \cong 31{,}48 \text{ SM}^2$$

Logo, o desvio padrão é $\sigma = \sqrt{31{,}48} \Rightarrow \sigma \cong 5{,}61$ SM.

Em geral, quando uma variável apresenta seus valores distribuídos em *k* intervalos, a variância é dada por:

$$\boxed{\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^{k} n_i}}$$, sendo

x_i o ponto médio do intervalo *i*;
$\bar{x}$ a média aritmética;
n_i a frequência absoluta referente ao intervalo *i*.

Usando a frequência relativa $\left(f_i = \dfrac{n_i}{\sum_{i=1}^{k} n_i}\right)$, podemos escrever:

$$\boxed{\sigma^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot f_i}$$

EXERCÍCIOS

379. As alturas de um grupo de atletas de um clube estão relacionadas na tabela seguinte:

Altura (em metros)	Número de atletas
1,64 ⊢ 1,70	8
1,70 ⊢ 1,76	88
1,76 ⊢ 1,82	104
1,82 ⊢ 1,88	136
1,88 ⊢ 1,94	40
1,94 ⊢ 2,00	24
Total	**400**

a) Determine a média, a classe modal e a mediana dos dados apresentados.

b) Encontre a variância e o desvio padrão desses dados.

380. Os 200 funcionários de uma empresa foram submetidos a exames clínicos para avaliação de saúde. Na tabela seguinte, aparece o resultado do exame de dosagem de colesterol.

Colesterol (em mg/dℓde sangue)	Número de funcionários
140 ⊢ 180	21
180 ⊢ 220	45
220 ⊢ 260	73
260 ⊢ 300	34
300 ⊢ 340	27

a) Qual é a taxa mediana de colesterol, em mg, por dℓ de sangue?

b) O teste sugere que, se a taxa média de colesterol exceder 235 mg/dℓ de sangue, deve-se iniciar uma campanha de prevenção com os funcionários. Com base nesse exame, verifique se será necessário iniciar a campanha preventiva.

381. A seguir, são dados os percentuais da população dos estados brasileiros que vive em áreas urbanas.

Acre	66,4%
Alagoas	68,0%
Amapá	89,0%
Amazonas	74,8%
Bahia	67,1%
Ceará	71,5%
Distrito Federal	95,7%
Espírito Santo	79,5%
Goiás	87,9%

Maranhão	59,5%
Mato Grosso	79,4%
Mato Grosso do Sul	84,1%
Minas Gerais	82,0%
Pará	66,5%
Paraíba	71,1%
Paraná	81,4%
Pernambuco	76,5%
Piauí	62,9%

Rio de Janeiro	96,0%
Rio Grande do Norte	73,3%
Rio Grande do Sul	81,7%
Rondônia	64,1%
Roraima	76,1%
Santa Catarina	78,7%
São Paulo	93,4%
Sergipe	71,4%
Tocantins	74,3%

Fonte: *Almanaque Abril*, 2002.

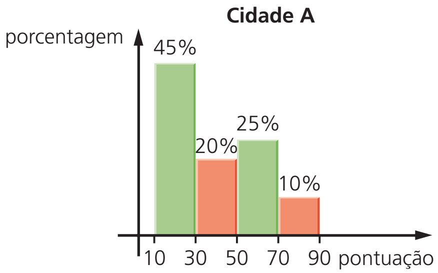
a) Agrupe essas informações em quatro intervalos, cada um com amplitude igual a 10, a partir do valor 59, e faça uma tabela de frequência.

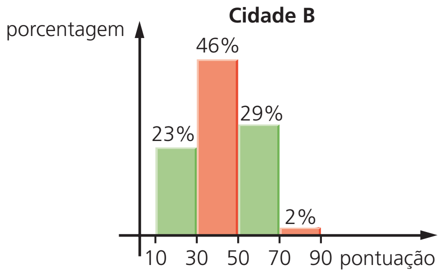
b) Utilizando os dados agrupados, calcule a média e o desvio padrão. Quantos valores não pertencem ao intervalo $[\bar{x} - \sigma, \bar{x} + \sigma]$?

382. Em um determinado Estado foi realizado nas suas duas maiores cidades, A e B, um levantamento sobre o grau de satisfação da população em relação à administração do governador. Um dos objetivos do levantamento era verificar se havia diferenças significativas quanto à opinião dos moradores das duas cidades. Cada entrevistado atribuiu uma nota de 0 a 100 para expressar sua satisfação.

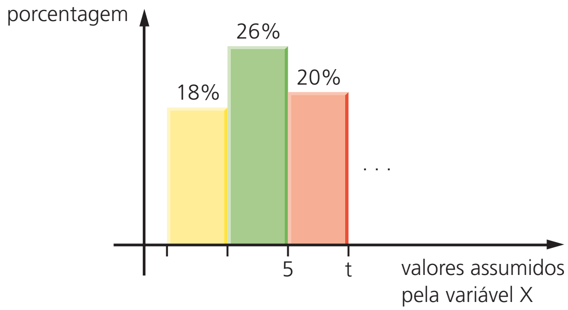
Adotou-se o seguinte critério de avaliação: caso a diferença entre as notas médias obtidas nas duas cidades não excedesse 5 (em módulo), a conclusão seria de que não havia diferenças significativas.

Com base nos dados apresentados a seguir, conclua se há divergência entre a opinião dos moradores de uma cidade e outra.

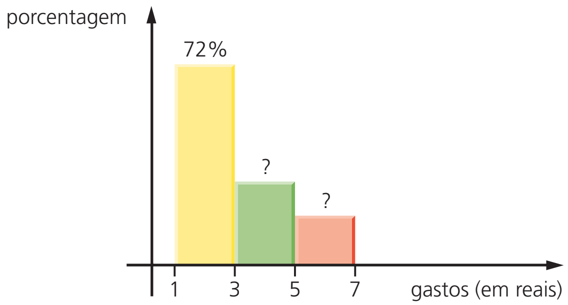




383. A figura mostra os três primeiros intervalos de um histograma que representa a distribuição de uma variável X, acompanhados das respectivas frequências. Se a mediana desses dados é 6,2, determine o valor de *t*.



384. O histograma abaixo mostra a distribuição de gastos com guloseimas registrados em uma barraca instalada na saída de uma estação de metrô.



Por falha de impressão, não aparecem no histograma as frequências relativas aos intervalos de 3 a 5 reais e de 5 a 7 reais. Sabe-se, entretanto, que a média de gastos é R$ 2,80.

a) Determine os valores relativos às frequências que não aparecem no gráfico.

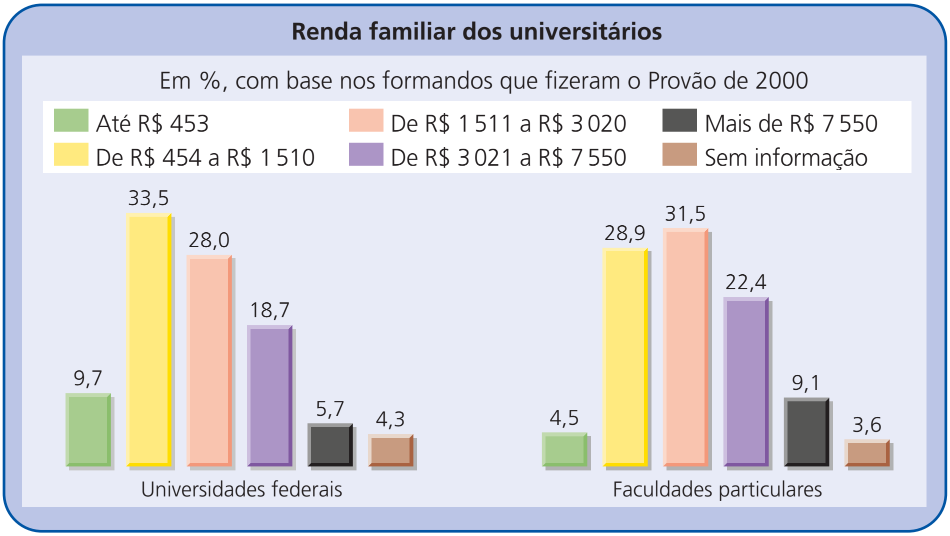
b) Qual é a variância correspondente?

385. A administradora de um condomínio residencial relacionou o atraso no pagamento das cotas condominiais relativas a certo mês, conforme mostra a tabela seguinte:

Dias de atraso	Número de apartamentos
0	48
1 a 9	30
10 a 18	24
19 a 27	18

Sabe-se que o valor do condomínio é R$ 200,00 e que há multa de 0,5% por dia de atraso. Faça uma estimativa das arrecadações mínima e máxima possíveis nesse mês. Que suposições estão envolvidas nesses cálculos?

386. Deseja-se comparar a renda familiar média dos universitários nas duas situações retratadas pelos gráficos.



Fonte: *Folha de S. Paulo*, 28 out. 2001.

a) Quais as dificuldades que você encontra para fazer esse cálculo?

b) Calcule a média em cada caso, admitindo que, no primeiro intervalo, a renda varia de R$ 200,00 a R$ 453,00 e, no último intervalo, de R$ 7 550,00 a R$ 10 000,00. Despreze o último intervalo (sem informação), isto é, faça os cálculos sobre 95,7% e 96,4%, respectivamente.

387. Um radar fotográfico, instalado em uma rodovia na qual o limite de velocidade é 100 km/h, registrou em uma semana *x* multas por excesso de velocidade, conforme a tabela:

Velocidade (em km/h)	Número de ocorrências
$101 \vdash 108$	34
$108 \vdash 115$	41
$115 \vdash 122$	35
$122 \vdash 129$	22
$129 \vdash 136$	18

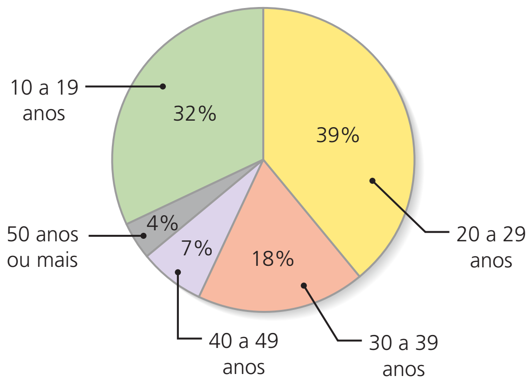
a) Determine o valor de *x*.

b) Calcule a média, a classe modal, a mediana e o desvio padrão da velocidade em que estavam os veículos quando foram multados.

c) Se o valor das multas varia de acordo com a faixa de velocidade ultrapassada, começando por R$ 180,00 e aumentando sempre 20% em relação à faixa anterior, determine o valor médio das multas aplicadas.

388. O gráfico seguinte mostra a distribuição dos espectadores de cinema, segundo faixas etárias, na Grande São Paulo.

Faixa etária



Fonte: *Veja São Paulo*, 14 maio 2003.

a) Admitindo que a classe de menor frequência tenha seus valores na faixa de 50 a 59 anos, determine a idade média dos espectadores.

b) Faça o cálculo da média supondo que os valores da classe de menor frequência pertençam ao intervalo [50, 65].

389. Um professor aplicou um teste de raciocínio lógico nas suas duas turmas do 3º ano do ensino médio. As notas obtidas pelos alunos são dadas a seguir:

Turma A

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

Turma B

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

a) Em cada turma, divida os alunos em cinco categorias de aproveitamento – péssimo, fraco, regular, bom e ótimo –, estabelecendo os limites de cada uma. A seguir, faça uma tabela de frequências.

b) Utilizando apenas os dados agrupados, responda:
- Qual turma apresentou melhor aproveitamento?
- Qual turma apresentou desempenho mais regular?

390. Um provedor da Internet desejava saber o tempo (em minutos) de acesso diário de seus assinantes à rede. Para isso, encomendou uma pesquisa com 80 pessoas, cujas informações sobre o tempo de acesso diário estão relacionadas a seguir:

39 – 52 – 99 – 125 – 81 – 87 – 175 – 71 – 77 – 41 – 20 – 63 – 89 – 72
61 – 91 – 140 – 18 – 72 – 15 – 43 – 27 – 92 – 35 – 55 – 50 – 17 – 130
62 – 115 – 32 – 24 – 161 – 96 – 192 – 80 – 54 – 50 – 20 – 86 – 51
129 – 96 – 19 – 163 – 21 – 55 – 98 – 135 – 100 – 123 – 23 – 170 – 143
128 – 84 – 71 – 37 – 232 – 64 – 15 – 158 – 105 – 103 – 76 – 42 – 110
112 – 86 – 65 – 47 – 200 – 57 – 80 – 34 – 84 – 38 – 67 – 78 – 114

a) Agrupe as informações em oito classes de amplitude igual a 30 minutos e faça um histograma correspondente.

b) Usando os dados agrupados, encontre as três medidas de centralidade correspondentes ao tempo de acesso.

c) A partir do histograma construído no item *a*, construa um novo histograma, agrupando os tempos de hora em hora. Em seguida, encontre as três medidas de centralidade.

391. A tabela seguinte mostra a evolução do índice de desenvolvimento humano (IDH) em uma década no Brasil.

***Ranking* dos estados**		
	IDH	
	1991	**2000**
Acre	0,624	0,697
Alagoas	0,548	0,649
Amapá	0,691	0,753
Amazonas	0,664	0,713
Bahia	0,590	0,688
Ceará	0,593	0,700
Distrito Federal	0,799	0,844
Espírito Santo	0,690	0,765
Goiás	0,700	0,776
Maranhão	0,543	0,636
Mato Grosso	0,685	0,773
Mato Grosso do Sul	0,716	0,778
Minas Gerais	0,697	0,773
Pará	0,650	0,723

***Ranking* dos estados**		
	IDH	
	1991	**2000**
Paraíba	0,561	0,661
Paraná	0,711	0,787
Pernambuco	0,620	0,705
Piauí	0,566	0,656
Rio de Janeiro	0,753	0,807
Rio Grande do Norte	0,604	0,705
Rio Grande do Sul	0,753	0,814
Rondônia	0,660	0,735
Roraima	0,692	0,746
Santa Catarina	0,748	0,822
São Paulo	0,778	0,820
Sergipe	0,597	0,682
Tocantins	0,611	0,710

Fonte: *O Estado de S. Paulo*, 3 out. 2003.

a) Em relação às duas datas mencionadas, agrupe os estados em classes de amplitude igual a 0,1 e faça uma tabela de frequências correspondente.

b) Utilizando os dados agrupados, compare as médias de IDH dos dois períodos.

c) Qual foi o aumento percentual registrado na média calculada no item anterior?

d) Utilizando os dados agrupados, encontre o desvio padrão do IDH em 2000.

392. Os dados seguintes referem-se às taxas de ocupação de um teatro durante os cinquenta dias em que uma peça ficou em cartaz.

30% – 43% – 66% – 57% – 72% – 78% – 38% – 61% – 59% – 53%

62% – 49% – 82% – 68% – 59% – 45% – 60% – 65% – 73% – 76%

70% – 64% – 68% – 75% – 80% – 62% – 54% – 71% – 82% – 49%

55% – 60% – 66% – 72% – 70% – 60% – 58% – 64% – 50% – 83%

82% – 56% – 79% – 80% – 71% – 88% – 84% – 80% – 70% – 55%

a) Organize os dados em cinco intervalos de amplitude igual a 10, a partir do menor valor encontrado, e faça uma tabela de frequências correspondente.

b) Utilizando os dados agrupados, encontre a média dos percentuais relacionados acima.

c) Suponha que a peça tenha ficado em cartaz por mais *n* dias, numa longa temporada de preços populares. Nessa temporada, verificou-se que a ocupação da sala, em cada dia, nunca foi inferior a 80%, mas não chegou a 90%. Qual deve ser o menor valor de *n* para que a média de ocupação de **todo o período** seja no mínimo de 80%?

393. A tabela ao lado informa a quantidade diária de reclamações recebidas por um órgão de defesa do consumidor durante um ano.

Determine o percentual de dias em que foram registradas menos de 110 reclamações.

Número de queixas	Número de dias
$0 \vdash 40$	30
$40 \vdash 80$	75
$80 \vdash 120$	120
$120 \vdash 160$	95
$160 \vdash 200$	40
Total	**360**

Solução

Os dois primeiros intervalos reúnem 30 + 75 = 105 dias, em que foram registradas até 80 queixas. Por outro lado, não sabemos de que modo as 120 ocorrências estão distribuídas no intervalo de 80 a 120 queixas.

Admitindo novamente uma distribuição uniforme dentro do intervalo, podemos separar o terceiro intervalo em subintervalos de amplitude 10: ►

80 ⊢ 90	30 dias
90 ⊢ 100	30 dias
100 ⊢ 110	30 dias
110 ⊢ 120	30 dias

Assim, em 30 + 30 + 30 = 90 dias foram registradas de 80 a 110 queixas.

Por fim, o percentual de dias procurado é:

$$\frac{105 + 90}{360} = \frac{195}{360} \cong 54{,}2\%$$

394. A tabela ao lado mostra um levantamento sobre o público pagante nas duas primeiras rodadas (40 jogos) do Campeonato Brasileiro de Futebol.

a) Qual é o número estimado de jogos que apresentaram público pagante inferior a 15 000 pessoas?

b) Qual é o número estimado de jogos que apresentaram público pagante de pelo menos 20 000 pessoas?

Público pagante	Número de jogos
1 000 ⊢ 7 000	3
7 000 ⊢ 13 000	4
13 000 ⊢ 19 000	9
19 000 ⊢ 25 000	12
25 000 ⊢ 31 000	6
31 000 ⊢ 37 000	4
37 000 ⊢ 43 000	2
Total	**40**

395. A pontuação dos 100 000 alunos que fizeram a primeira fase de um exame vestibular está mostrada na tabela seguinte.
Observação: Só são possíveis pontos inteiros.

Número de pontos	Número de alunos
0 ⊢ 10	1 400
10 ⊢ 20	6 900
20 ⊢ 30	13 000
30 ⊢ 40	14 500
40 ⊢ 50	19 300
50 ⊢ 60	16 800
60 ⊢ 70	11 400
70 ⊢ 80	10 700
80 ⊢ 90	5 100
90 ⊢ 100	900
Total	**100 000**

Estime a proporção de alunos que obtiveram:

a) pelo menos 55 pontos; b) menos de 37 pontos; c) 88 pontos ou mais.

396. Os dados seguintes, coletados durante uma semana em um hospital veterinário, referem-se aos "pesos" de 160 cachorros recém-nascidos.

"Peso" (em kg)	Número de cachorros recém-nascidos
$0 \vdash 0,5$	6
$0,5 \vdash 1,0$	9
$1,0 \vdash 1,5$	21
$1,5 \vdash 2,0$	30
$2,0 \vdash 2,5$	42
$2,5 \vdash 3,0$	27
$3,0 \vdash 3,5$	20
$3,5 \vdash 4,0$	4
$4,0 \vdash 4,5$	1
Total	**160**

Estime a proporção de cachorros recém-nascidos com "pesos" pertencentes ao intervalo [1,7; 3,3[.

XVII. Outras medidas de separação de dados

Vimos no item XII que a mediana é um valor que divide um conjunto de dados em duas partes. Vamos agora ver outras medidas de separação de dados (ou separatrizes):

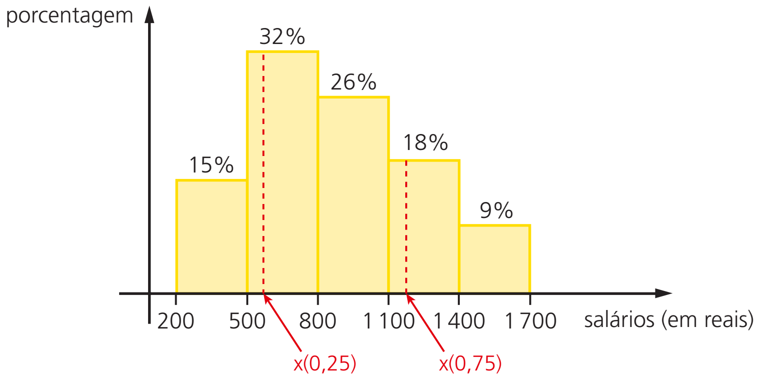
19. Quartis

Seja x uma variável quantitativa cujos valores estão agrupados em classes (intervalos).

- O **primeiro quartil**, indicado por x(0,25), é o valor que divide o conjunto de dados em duas partes, tais que:
  - 25% dos valores assumidos por x são menores ou iguais a x(0,25);
  - 75% dos valores assumidos por x são maiores ou iguais a x(0,25).
- O **segundo quartil**, indicado por x(0,50), é o valor correspondente à mediana.
- O **terceiro quartil**, indicado por x(0,75), é o valor que divide o conjunto de dados em duas partes, tais que:
  - 75% dos valores assumidos por x são menores ou iguais a x(0,75);
  - 25% dos valores assumidos por x são maiores ou iguais a x(0,75).

Exemplo:

Os salários dos funcionários de um supermercado estão mostrados no histograma seguinte:



Vejamos como determinar x(0,25) e x(0,75).

- Cálculo de x(0,25):

  As duas primeiras classes reunidas concentram 15% + 32% = 47% dos salários dos funcionários.

  Assim, o primeiro quartil, x(0,25), pertence ao segundo intervalo.

  Analogamente ao cálculo da mediana, podemos escrever:

$$\frac{x(0{,}25) - 500}{25\% - 15\%} = \frac{800 - 500}{32\%} \Rightarrow x(0{,}25) = 593{,}75$$

  Isso significa que, no conjunto dos salários de todos os funcionários do supermercado, 25% são menores que R$ 593,75 e 75% são maiores que esse valor.

- Cálculo de x(0,75):

  As três primeiras faixas salariais concentram 15% + 32% + 26% + 73% dos salários.

  Desse modo, o terceiro quartil é um valor que pertence ao intervalo $1100 \vdash 1400$, pois os quatro primeiros intervalos contêm 91% dos salários.

  Temos:

$$\frac{x(0{,}75) - 1100}{75\% - 73\%} = \frac{1400 - 1100}{18\%} \Rightarrow x(0{,}75) \cong 1133{,}33$$

  Isso significa que, no conjunto dos salários de todos os funcionários do supermercado, os 25% mais altos estão na faixa de R$ 1133,33 a R$ 1700,00.

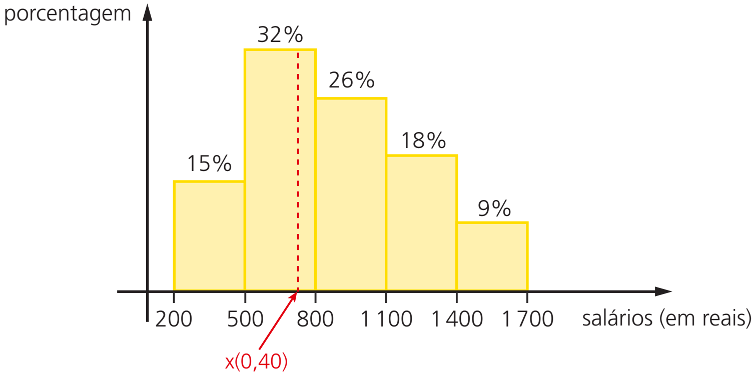
20. Decis

Com base na mesma ideia de quartis, é possível dividir um conjunto de dados agrupados em duas partes usando os **decis**.

Em geral, o **n-ésimo decil** (n = 1, 2, ..., 9) é um valor que divide o conjunto de dados em duas partes, tais que $(10 \cdot n)\%$ dos valores da distribuição são menores ou iguais a ele e $(100 - 10 \cdot n)\%$ são maiores ou iguais a ele.

Exemplo:

Considerando o histograma apresentado no exemplo do item anterior, vejamos como determinar o **quarto decil** em relação à distribuição dos salários dos funcionários do supermercado.



O quarto decil, x(0,4), é um valor pertencente à faixa $500 \vdash 800$.
Temos:

$$\frac{x(0{,}4) - 500}{40\% - 15\%} = \frac{800 - 500}{32\%} \Rightarrow x(0{,}4) \cong 734{,}38$$

O valor 734,38 divide o conjunto de salários em duas partes, tais que uma delas contém os 40% dos salários mais baixos – de 200 a 734,38 reais – e a outra parte – de 734,38 a 1 700 reais – reúne os 60% dos salários mais altos.

21. Percentis

Com base na mesma ideia de quartis e decis, é possível dividir um conjunto de dados agrupados em duas partes usando os **percentis**.

O **n-ésimo percentil** (n = 1, 2, ..., 99) é um valor que divide o conjunto de dados em duas partes, tais que n% dos valores da distribuição são menores ou iguais a ele e (100 – n)% são maiores ou iguais a ele.

Exemplos:

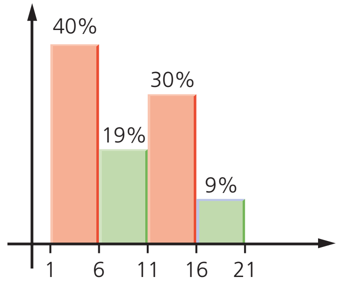
1º) O **décimo quarto percentil** é indicado por x(0,14); 14% dos dados são menores que x(0,14) e 86% são maiores que x(0,14).

2º) O **sexagésimo percentil** (ou sexto decil) é indicado por x(0,60); 60% dos dados são menores que x(0,60) e 40% são maiores que x(0,60).

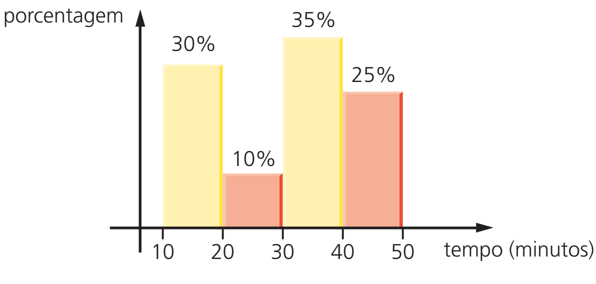
O cálculo de percentis segue exatamente as proporções apresentadas nos cálculos relativos a quartis e decis.

EXERCÍCIOS

397. Observe o gráfico ao lado, que representa a distribuição de valores de uma variável quantitativa.
Determine:
a) o primeiro quartil;
b) o terceiro decil;
c) o segundo quartil;
d) o oitavo decil.



398. Os dados seguintes, coletados em uma manhã de nevoeiro em um aeroporto, referem-se ao tempo de atraso na decolagem dos voos.



Determine:
a) o tempo médio de atraso em cada voo naquela manhã;
b) o intervalo interquartil, isto é, o intervalo [x(0,25); x(0,75)];
c) o sexto decil;
d) o valor de *n*, considerando que *n* seja o tempo de atraso em minutos em noventa por cento dos voos.

399. A tabela ao lado informa a distribuição das notas obtidas por uma turma em uma prova de Estatística.

Do conjunto de todas as notas, as 25% maiores **não** são inferiores a *x*. Qual é o valor de *x*?

Nota	Porcentagem (%)
$0 \vdash 2$	7,50
$2 \vdash 4$	33,75
$4 \vdash 6$	30,00
$6 \vdash 8$	20,00
$8 \vdash 10$	8,75

400. Levantamentos realizados com alunos e funcionários de uma faculdade revelaram que a média de tempo diário gasto com a leitura de jornais não excedia 15 minutos. Para incentivar o hábito da leitura, cada departamento disponibilizou alguns exemplares de jornais em suas bibliotecas. Uma nova pesquisa foi realizada semanas após o início da experiência, a fim de verificar se alguma mudança havia ocorrido. Os resultados são mostrados na tabela.

Tempo de leitura	Porcentagem (%)
$0 \vdash 5$	7,0
$5 \vdash 10$	19,5
$10 \vdash 15$	33,5
$15 \vdash 20$	28,5
$20 \vdash 25$	10,5
$25 \vdash 30$	1,0
Total	**100**

a) A medida tomada com o propósito de incentivar a leitura surtiu efeito, isto é, elevou a média histórica de tempo diário de leitura?

b) Para verificar a conveniência de repetir essa estratégia em outro momento, adotou-se o seguinte critério: 75% dos leitores deveriam dedicar, no mínimo, 8 minutos diários à leitura. Com base nos resultados obtidos, verifique se o procedimento adotado possibilitou atingir a meta estabelecida.

401. Peixes de uma determinada espécie são vendidos a um restaurante. A tabela ao lado informa a distribuição, em porcentagem, do "peso" de certo número de peixes dessa espécie vendidos ao restaurante em determinado dia.

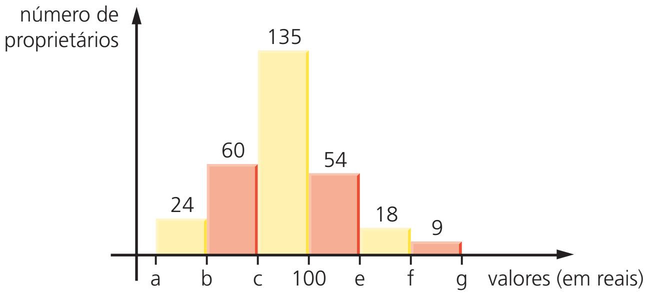
Peixes de "peso" reduzido entram no preparo de pratos com acompanhamentos. Por essa razão, o gerente do restaurante propôs a separação dos peixes em três categorias, de forma que:

"Peso" (em gramas)	Porcentagem (%)
$50 \vdash 100$	2,5
$100 \vdash 150$	30,0
$150 \vdash 200$	27,5
$200 \vdash 250$	35,0
$250 \vdash 300$	5,0

- os 20% mais leves pertençam à classe A;
- os 50% de "peso" intermediário pertençam à classe B;
- os 30% mais pesados pertençam à classe C.

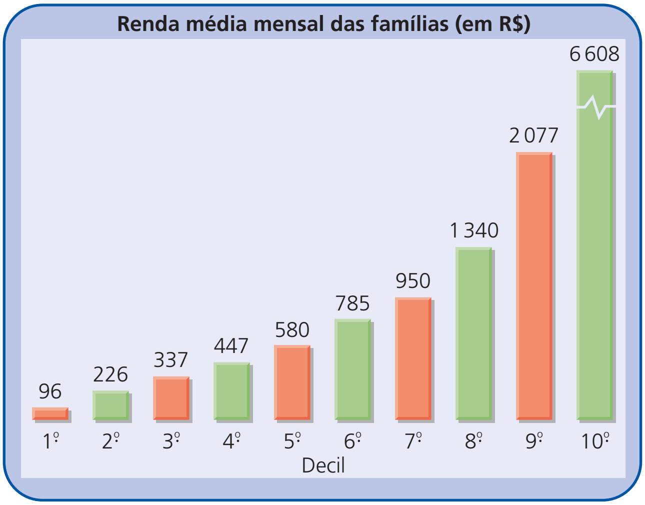
Determine os limites aproximados de "peso" que definem cada uma dessas classes.

402. No gráfico seguinte estão representados os valores das despesas mensais com combustível relacionadas por 300 proprietários de veículos.



Sabendo que 20% dos proprietários gastam até 58 reais e 30% deles gastam no mínimo 98 reais, determine os valores de *b* e *c*. Que suposição deve ser feita a fim de que seja possível encontrar os valores de *a*, *e*, *f* e *g*?

403. O gráfico a seguir mostra a renda média mensal das famílias brasileiras e a sua desigual distribuição entre a população do país.



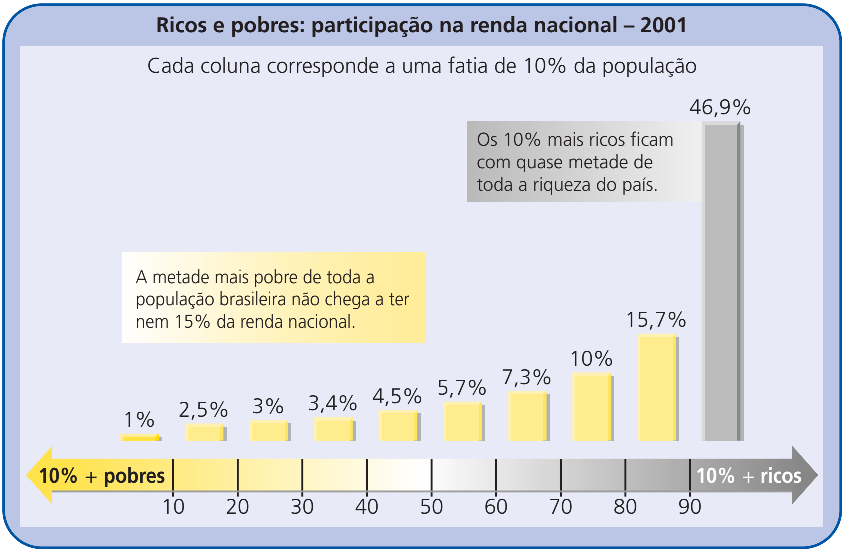
Fonte: *O Estado de S. Paulo*, 12 maio 2003.

a) Qual é a renda familiar média mensal dos 10% mais pobres?

b) Entre as famílias brasileiras, as 20% mais ricas têm renda média mensal superior a *x* reais. Qual é o valor de *x*?

c) O intervalo (a, b) contém os valores da renda mensal cuja média é superior à média dos 20% mais pobres e inferior à média dos 30% mais ricos. Determine *a* e *b*.

404. Complete corretamente as afirmações seguintes, de acordo com o gráfico.



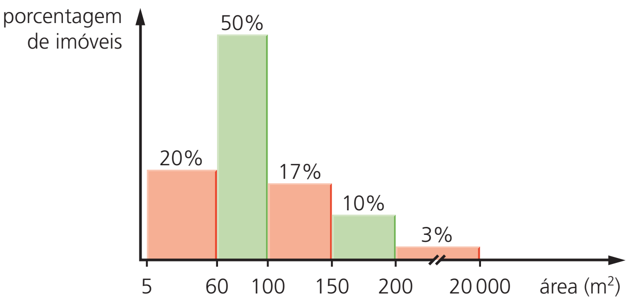
Fonte: *Almanaque Abril* – atualidades de vestibular, 2004.

a) A metade mais pobre de toda a população brasileira detém ▲% de toda a renda do país.

b) Os 20% mais ricos de toda a população concentram ▲% de toda a renda do país.

c) O intervalo compreendido entre o 3º e o 8º decil reúne ▲% da renda nacional.

405. (EEM-SP) O histograma abaixo refere-se às áreas dos imóveis de um pequeno município.



O prefeito pretende isentar do pagamento do imposto predial e territorial urbano (IPTU) os proprietários dos imóveis de menor área, até o limite de 30% dos imóveis do município. Determine a área máxima de um imóvel para que seu proprietário fique isento do pagamento do IPTU.