

# Proposta de resolução

## Página 86

- As questões estatísticas são **(B)**, **(D)** e **(E)**.
- População: conjunto dos visitantes da galeria de arte
- Amostra: visitantes da galeria nos últimos três dias  
Dimensão da amostra: 547
- Variáveis estatísticas:
  - País de origem
  - Tempo da visita à galeria
  - Número de visitas à galeria

## Página 87

- Variável estatística: número de visitantes anual da Fundação de Serralves  
A variável estatística é quantitativa discreta.
- Número de visitantes em 2021: 409 216  
Número de visitantes em 2022: 758 103  
 $758\ 103 - 409\ 216 = 348\ 887$   
 $\frac{348\ 887}{409\ 216} \times 100 \approx 85$   
O aumento do número de visitantes de 2021 para 2022 foi, aproximadamente, 85%.  
Opção **(A)**
- A:** variável quantitativa discreta  
**B:** variável quantitativa contínua  
**C:** variável quantitativa discreta  
**D:** variável qualitativa nominal  
**E:** variável quantitativa contínua

## Página 88

- Ocupação dos apartamentos

N.º de pessoas que habitam o apartamento	Freq. absoluta	Freq. relativa
2	5	$\frac{5}{25} \times 100 = 20\%$
3	9	$\frac{9}{25} \times 100 = 36\%$
4	5	$\frac{5}{25} \times 100 = 20\%$
5	4	$\frac{4}{25} \times 100 = 16\%$
6	2	$\frac{2}{25} \times 100 = 8\%$

- A variável estatística é o *peso* da bagagem de mão e é quantitativa contínua.

- Dimensão da amostra (número total de bagagens selecionadas):  $18 + 30 + 23 + 4 = 75$

- 

Peso da bagagem (em kg)	Freq. absoluta	Freq. relativa
$0 \leq p < 5$	18	$\frac{18}{75} \times 100 = 24\%$
$5 \leq p < 10$	30	$\frac{30}{75} \times 100 = 40\%$
$10 \leq p < 15$	23	$\frac{23}{75} \times 100 \approx 31\%$
$15 \leq p < 20$	4	$\frac{4}{75} \times 100 \approx 5\%$

## Página 89

- A moda é 4 porque é o valor com maior frequência absoluta.

$$\text{Média: } \bar{x} = \frac{3+7+1+4+4+6+2}{7} = \frac{27}{7} \approx 3,86$$

Para determinar a mediana e os quartis é preciso ordenar os dados:

$$1 \quad \boxed{2} \quad 3 \quad \boxed{4} \quad 4 \quad \boxed{6} \quad 7$$

$Q_1 \quad Q_2 \quad Q_3$

A mediana é 4, o 1.º quartil é 2, o 2.º quartil é 4 ( $Q_2 = M_e$ ) e o quartil é 6.

$$M_0 = 4 \quad \bar{x} \approx 3,86 \quad M_e = Q_2 = 4$$

$$Q_1 = 2 \quad Q_3 = 6$$

- A moda é 2 porque é o valor com maior frequência absoluta.

Média:

$$\bar{x} = \frac{5+6+2+4+3+8+2+6+2+7}{10} = \frac{45}{10} = 4,5$$

Para determinar a mediana e os quartis é preciso ordenar os dados:

$$2 \quad 2 \quad \boxed{2} \quad 3 \quad \boxed{4 \quad 5} \quad 6 \quad \boxed{6} \quad 7 \quad 8$$

$Q_1 \quad Q_2 = \frac{4+5}{2} = 4,5 \quad Q_3$

$$M_0 = 2 \quad \bar{x} = 4,5 \quad M_e = Q_2 = 4,5$$

$$Q_1 = 2 \quad Q_3 = 6$$

## Página 90

- O problema que motivou os estudos referidos no excerto foi a possibilidade da Grã-Bretanha mudar para o Sistema Métrico.
- A população é construída pelos habitantes da Grã-Bretanha.



3. A amostra utilizada pelo jornal *Sunday* é constituída pelos seus leitores.  
A amostra não é representativa da população porque só é formada por leitores do jornal (os habitantes que não são leitores do jornal *Sunday* não têm representação na amostra).

4.1.  $0,33 \times 5700 = 1881$

No estudo feito pela *Gallup*, 1881 pessoas responderam não conhecer o sistema métrico.

4.2.  $0,33 \times 49\,000\,000 = 16\,170\,000$

Estima-se que, em 1950, aproximadamente, 16 170 000 pessoas na Grã-Bretanha desconhecia o sistema métrico.

#### Página 91

- 1.1. Amplitude técnica diária – existe variabilidade  
1.2. Extensão, em quilómetros, da linha de metro – não existe variabilidade  
2.1. Tempo oficial do jogo – não há variabilidade  
2.2. Tempo oficial de intervalo – não há variabilidade  
2.3. Tempo útil de jogo – há variabilidade

#### Página 92

- 1.1. a) População: alojamentos familiares clássicos em Portugal  
b) Variável em estudo: número de divisões dos alojamentos familiares clássicos.  
Variável quantitativa discreta  
1.2. a) Consultando a base de dados Pordata, sabe-se que os dados em falta no gráfico referentes ao número de alojamentos com 2, 4 e 5 divisões são 109 008, 1 250 622 e 1 295 053, respetivamente.  
b) Número total de alojamentos:  
 $17\,542 + 109\,008 + 428\,644 + 1\,250\,622 + 1\,295\,053 + 1\,037\,712 = 4\,138\,581$   
Número de alojamentos com 6 ou mais divisões: 1 037 712  
 $\frac{1\,037\,712}{4\,138\,581} \times 100 \approx 25\%$  Opção (A)

- 2.1. Dimensão da amostra: 1006 entrevistas

- 2.2. Mulheres: 45%, aproximadamente  
Homens: 55%, aproximadamente  
 $0,55 \times 1006 = 553,3 \approx 553$   
Aos homens foram feitas 553 entrevistas.  
2.3. A variável associada à questão colocada é qualitativa ordinal.

#### Página 93

3. Nesse controlo é feita uma sondagem pois só uma parte dos fósforos produzidos é utilizada no teste de controlo.  
4.1. Sim, porque é uma das casas que fazem parte dos alojamentos familiares de aldeia (população em estudo).  
4.2. A amostra utilizada é constituída por alojamentos da rua do Francisco.  
Dimensão da amostra: 20  
4.3. A amostra não é representativa da população porque só contém alojamentos da mesma rua (deveria ter alojamentos das várias ruas da aldeia).

#### Página 94

- 5.1. Amostra aleatória simples  
5.2. Amostra estratificada ou proporcional  
5.3. Amostra aleatória sistemática  
6. Para estudar o número médio de páginas dos livros de uma biblioteca usaria a amostra aleatória simples atendendo à diversidade do número de páginas dos livros que podem ocorrer.

#### Página 95

- 7.1. A: Conjunto dos alunos do 2.º Ciclo da escola  
B: Conjunto dos alunos da escola  
7.2. Os valores da variável representem maior variabilidade no estudo B  
7.3. A amostra deve ter maior dimensão no estudo B porque há uma maior variabilidade dos valores da variável.

#### Página 96

1. Conjunto dos eleitores residentes em Portugal  
2.1. A dimensão da amostra é 1002.



- 2.2. Mulheres inquiridas: 46%  
Homens inquiridas: 54%  
 $0,54 \times 1002 \approx 541$   
Fizeram parte da amostra 541 homens.
3. A técnica de amostragem utilizada neste estudo foi a amostra estratificada ou proporcional.
4. A informação referida no destaque 5. está relacionada com a 3.ª fase de um estudo estatístico (Interpretação das representações obtidas).

### Página 97

- 1.1.  $a = 0,348 \times 1011 \approx 352$   
 $b = 0,163 \times 1011 \approx 165$   
 $c = 0,06 \times 1011 \approx 61$   
 $a = 352$ ,  $b = 165$  e  $c = 61$
- 1.2.  $n_1 = 0,039 \times 1011 \approx 39$   
 $n_2 = 0,045 \times 1011 \approx 45$   
 $n_3 = 0,091 \times 1011 \approx 92$   
 $n_1 = 39$ ,  $n_2 = 45$  e  $n_3 = 92$
2. A análise dos resultados permite concluir que a maioria das pessoas (57,1%) conhecem bem as ODS.

### Página 98

8. **A:** Mínimo de computadores por família  
Variável quantitativa discreta
- B:** Grau de satisfação de um serviço  
Variável qualitativa
- C:** Capacidade do depósito de combustível  
Variável quantitativa contínua
- D:** Intensidade da chuva (fraca, moderada, forte)  
Variável qualitativa
- E:** Pressão sanguínea  
Variável quantitativa contínua
- F:** Número de lugares sentados no autocarro  
Variável quantitativa discreta
9. A variável “tamanho dos livros” é qualitativa.

### Página 99

1. A variável em estudo é quantitativa.
- 2.1. Estimativas
- |   |   |   |   |   |   |   |   |  |  |
|---|---|---|---|---|---|---|---|--|--|
| 2 | 3 | 7 |   |   |   |   |   |  |  |
| 3 | 0 | 1 | 3 | 5 | 7 | 9 | 9 |  |  |
| 4 | 0 | 0 | 2 | 3 | 7 | 8 |   |  |  |
| 5 | 0 | 4 | 5 | 6 | 6 | 7 | 9 |  |  |
| 6 | 3 | 5 | 8 |   |   |   |   |  |  |
- Legenda:**  
5|6 significa 56
- 2.2. A variável assume 22 valores diferentes, logo o número de barras seria 22. Opção (C)
- 2.3. O diagrama de barras não é adequado a esta situação porque a variável assume muitos valores diferentes.
- 2.4.  $a = 23$  (valor mínimo da variável)  
 $e = 68$  (valor máximo da variável)  
Número total de dados: 25  
A mediana corresponde ao 13.º valor da amostra ordenada  
 $c = \bar{x} = 43$   
O 1.º quartil é a média aritmética entre o 6.º e o 7.º valores da amostra ordenada e o 3.º quartil é a média aritmética entre o 19.º e o 20.º valores da amostra ordenada.  
Assim,  $b = \frac{35+37}{2} = 36$  e  $d = \frac{56+56}{2} = 56$   
 $a = 23$ ,  $b = 36$ ,  $c = 43$ ,  $d = 56$  e  $e = 68$

### Página 100

#### 10.1.

Valores da variável $x_i$	Freq. absoluta $n_i$	Freq. relativa $f_i$
1	1	$\frac{1}{12} \approx 0,08 = 8\%$
2	1	$\frac{1}{12} \approx 0,08 = 8\%$
3	4	$\frac{4}{12} \approx 0,33 = 33\%$
4	2	$\frac{2}{12} \approx 0,17 = 17\%$
5	2	$\frac{2}{12} \approx 0,17 = 17\%$
6	2	$\frac{2}{12} \approx 0,17 = 17\%$
$N = 12$		<b>Soma: 1 (100%)</b>



## 10.2.

$x_i$	$n_i$	$f_i$
1	4	0,2
2	3	0,15
3	6	0,3
4	3	0,15
5	2	0,1
6	2	0,1

Cálculos auxiliares:

$$\frac{4}{N} = 0,2 \Leftrightarrow N = \frac{4}{0,2} \Leftrightarrow N = 20$$

$$0,3 \times 20 = 6 \quad 0,15 \times 20 = 3 \quad 0,1 \times 20 = 2$$

## Página 101

## 11.1.

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	3	3	$\frac{3}{8} = 0,375 = 37,5\%$	37,5%
1	2	5	$\frac{2}{8} = 0,25 = 25\%$	62,5%
2	2	7	$\frac{2}{8} = 0,25 = 25\%$	87,5%
3	1	8	$\frac{1}{8} = 0,125 = 12,5\%$	100%
$N = 8$			$S = 1 (100\%)$	

- 11.2. a) Consultando a tabela anterior, verifica-se que a equipa marcou, no máximo, dois golos em 87,5% dos jogos.
- b) Consultando a tabela anterior, verifica-se que a equipa marcou pelo menos um golo em  $100\% - 37,5\%$  dos jogos, ou seja, 62,5% dos jogos.

## 12.

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
4	3	3	0,12	0,12
7	6	9	0,24	0,36
10	12	21	0,48	0,84
15	4	25	0,16	1

Cálculos auxiliares:

$$N = 25 \quad 9 - 3 = 6$$

$$9 + 12 = 21 \quad 25 - 21 = 4$$

$$\frac{3}{25} = 0,12 \quad \frac{6}{25} = 0,24$$

$$\frac{12}{25} = 0,48 \quad \frac{4}{25} = 0,16$$

## Página 102

## 1.

N.º mín. de fotos $x_i$	Freq. absoluta $n_i$	Freq. absoluta acumulada $N_i$
0	3	3
1	4	7
2	2	9
3	1	10

N.º máx. de fotos $x_i$	Freq. absoluta $n_i$	Freq. relativa $f_i$	Freq. relativa acumulada $F_i$
1	1	10%	10%
3	1	10%	20%
4	3	30%	50%
5	2	20%	70%
6	1	10%	80%
8	1	10%	90%
10	1	10%	100%

- 2.1. Sete alunos da amostra admitiram ter publicado, no mínimo, menos de duas fotografias.
- 2.2. Pelo menos 5 fotografias, ou seja, 5 ou mais fotografias  
 $20\% + 10\% + 10\% + 10\% = 50\%$   
 50% dos alunos da amostra admitiu, no máximo, ter publicado pelo menos cinco fotografias.

## Página 103

## 1.1.

N.º de postos $x_i$	Freq. absoluta $n_i$	Freq. absoluta acumulada $N_i$	Freq. relativa $f_i$	Freq. relativa acumulada $F_i$
2	5	5	25%	25%
3	3	8	15%	40%
4	5	13	25%	65%
5	3	16	15%	80%
6	4	20	20%	100%

- 1.2. a) Há 13 ( $5 + 3 + 5$ ) concelhos que têm, pelo menos, cinco postos de combustível.
- b) Há 12 ( $5 + 3 + 4$ ) concelhos que têm mais de três postos de combustível.
- c) Há 7 ( $3 + 4$ ) concelhos que têm, pelo menos, cinco postos de combustível.



- 1.3. a) 15% dos concelhos têm três postos de combustível.  
 b) 75% dos concelhos (100% – 25%) têm mais de dois postos de combustível.

2.1.

N.º de atividades $x_i$	$n_i$	$N_i$	$f_i$	$F_i$
0	4	4	$\frac{4}{40} = 10\%$	10%
1	12	16	$\frac{12}{40} = 30\%$	40%
2	15	31	$\frac{15}{40} = 37,5\%$	77,5%
3	6	37	$\frac{6}{40} = 15\%$	92,5%
4	3	40	$\frac{3}{40} = 7,5\%$	100%

- 2.2. a) Consultando a tabela anterior, verifica-se que há 31 alunos com menos de três atividades extracurriculares.  
 b) Consultando a tabela anterior, verifica-se que 22,5% (100% – 77,5%) dos alunos têm mais de duas atividades extracurriculares.

## Página 104

13.1.

Classes $x_i$	$n_i$	$f_i$
[10, 11[	4	$\frac{4}{10} = 0,4 = 40\%$
[11, 12[	2	$\frac{2}{10} = 0,2 = 20\%$
[12, 13[	3	$\frac{3}{10} = 0,3 = 30\%$
[13, 14[	1	$\frac{1}{10} = 0,1 = 10\%$

13.2.

Classes $x_i$	$n_i$	$f_i$
[9,5; 10,5[	1	$\frac{1}{10} = 0,1 = 10\%$
[10,5; 11,5[	4	$\frac{4}{10} = 0,4 = 40\%$
[11,5; 12,5[	4	$\frac{4}{10} = 0,4 = 40\%$
[12,5; 13,5[	1	$\frac{1}{10} = 0,1 = 10\%$

14.

Classes $x_i$	Freq. abs. $n_i$	Freq. rel. $f_i$
[5, 10[	12	0,24
[10, 15[	28	0,56
[15, 20[	10	0,2

Cálculos auxiliares:

$$\frac{28}{N} = 0,56 \Leftrightarrow N = \frac{28}{0,56} \Leftrightarrow N = 50$$

$$\frac{12}{50} = 0,24$$

$$0,2 \times 50 = 10$$

## Página 105

15.1.

$x_i$	$n_i$	$N_i$	$f_i$	$F_i$
[0, 2[	6	6	$\frac{6}{40} = 0,15$	0,15
[2, 4[	11	17	$\frac{11}{40} = 0,275$	0,425
[4, 6[	9	26	$\frac{9}{40} = 0,225$	0,65
[6, 8[	14	40	$\frac{14}{40} = 0,35$	1

15.2. Há 26 dados cujo valor é inferior a 6.

15.3.  $22,5\% + 35\% = 57,5\%$ 

57,5% dos dados tem valor não inferior a 4.

16.

Classes $x_i$	$n_i$	$N_i$	$f_i$	$F_i$
[7, 8[	5	5	$\frac{5}{16} = 0,3125$	0,3125
[8, 9[	4	9	$\frac{4}{16} = 0,25$	0,5625
[9, 10[	7	16	$\frac{7}{16} = 0,4375$	1



## 3.1.

Velocidade (km/h)	Freq. absoluta	Freq. absoluta acum.	Freq. relativa	Freq. relativa acum.
[50, 60[	8	8	$\frac{8}{200} = 0,04$	0,04
[60, 70[	6	14	$\frac{6}{200} = 0,03$	0,07
[70, 80[	25	39	$\frac{25}{200} = 0,125$	0,195
[80, 90[	42	81	$\frac{42}{200} = 0,21$	0,405
[90, 100[	68	149	$\frac{68}{200} = 0,34$	0,745
[100, 110[	34	183	$\frac{34}{200} = 0,17$	0,915
[110, 120[	10	193	$\frac{10}{200} = 0,05$	0,965
[120, 130[	4	197	$\frac{4}{200} = 0,02$	0,985
[130, 140[	3	200	$\frac{3}{200} = 0,15$	1

- 3.2. a) Não podemos garantir que 51 automóveis excedam a velocidade máxima permitida (pode ter havido automóveis que passaram exatamente a 100 km/h).

Só podemos garantir que 51 automóveis passam a pelo menos 100 km/h.

- b)  $100 - 51 = 149$   
 149 automóveis passaram a menos 100 km/h, logo podemos garantir que pelo menos 149 automóveis não excederam a velocidade máxima permitida.

- 4.1. A variável quantitativa, em estudo, é discreta.

## 4.2.

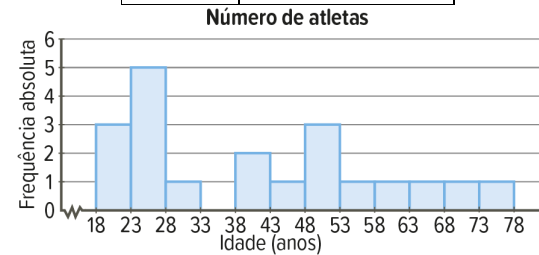
Classes $x_i$	$n_i$	$N_i$	$f_i$	$F_i$
[5, 13[	15	15	$\frac{15}{36} \approx 0,42$	0,42
[13, 21[	12	27	$\frac{12}{36} \approx 0,33$	0,75
[21, 29[	9	36	$\frac{9}{36} \approx 0,25$	1

- 4.3.  $100\% - 42\% = 58\%$

58% de linhas tem pelo menos treze palavras.

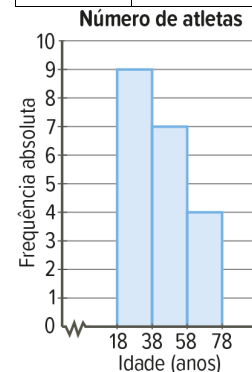
## 17.1. a)

Classes	Freq. absoluta $n_i$
[18, 23[	3
[23, 28[	5
[28, 33[	1
[33, 38[	0
[38, 43[	2
[43, 48[	1
[48, 53[	3
[53, 58[	1
[58, 63[	1
[63, 68[	1
[68, 73[	1
[73, 78[	1



## b)

Classes	Freq. absoluta, $n_i$
[18, 38[	9
[38, 58[	7
[58, 78[	4



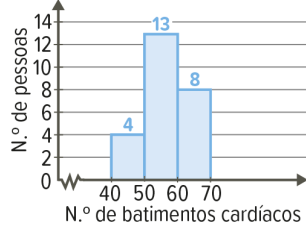
- 17.2. Comparando os histogramas obtidos em 17.1. escolheria o segundo histograma para evidenciar a informação relativas a níveis etários dos participantes (embora tenha poucas classes, evidencia o padrão existente no conjunto de dados).



1. A representação relativa à amostra A.  
Classes:  $[40, 50[$ ;  $[50, 60[$  e  $[60, 70[$

Histograma:

Batimentos cardíacos, por minuto

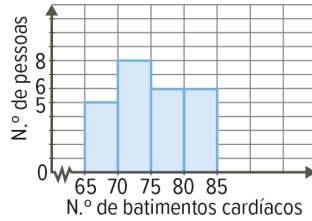


2.1.

Classes	Freq. absoluta $n_i$
$[65, 70[$	5
$[70, 75[$	8
$[75, 80[$	6
$[80, 85[$	6

2.2.

Batimentos cardíacos, por minuto



2.3.

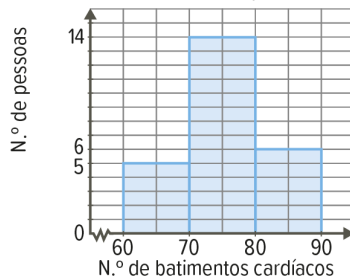
6	5	5	6	8	9									
7	0	0	1	2	2	3	3	3	5	5	6	7	7	8
8	0	1	2	3	3	4								

Legenda:

6 | 5 significa 65

2.4.

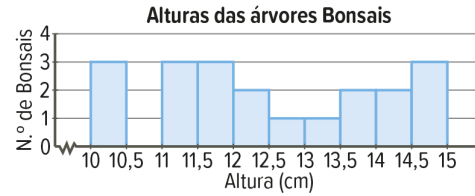
Batimentos cardíacos, por minuto



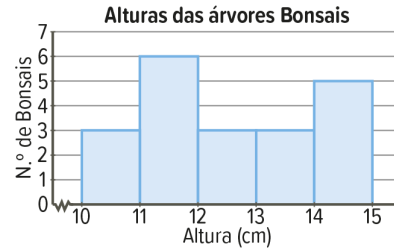
- 2.5. Esse padrão é mais evidente no histograma relativo a 2.4..

Página 109

1.1.



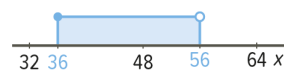
1.2.



2. O histograma construído em 1.2. parece mais adequado pois é o que melhor evidencia o padrão existente no conjunto de dados.  
3. Não é muito adequado aumentar a amplitude das classes porque diminuiria o número das classes e perder-se-ia informação sobre o padrão existente nos dados.

Página 110

1. O vento é considerado forte se a sua intensidade for não inferior a 36 km/h e inferior a 56 km/h .

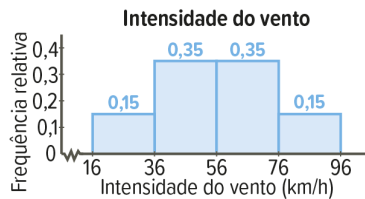
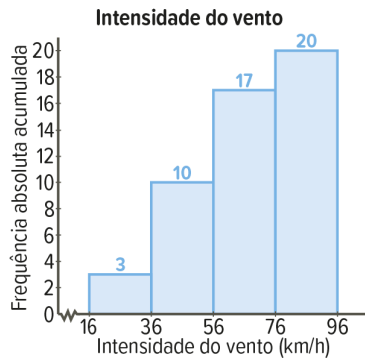


Não podemos responder à questão atendendo a que, a partir da informação dada no histograma, não é conhecida a distribuição dos dados dentro de cada classe.

2. Não, a proposta da Inês não é adequada ao que se pretende pois a escolha que foi feita para a amplitude das classes não permite saber quantas localidades registarem vento com intensidade superior a 16 km/h e inferior a 36 km/h .



3.1.



3.2. A percentagem de localidades com vento moderado foi 15%.

3.3. Por observação do histograma de frequências absolutas acumuladas, constata-se que em 10 localidades a intensidade do vento foi inferior a 56 km/h.

Assim sendo, as restantes 10 localidades (20–10) a intensidade do vento foi, pelo menos, 56 km/h.

### Página 111

5.1. A variável em estudo é a área das casas e é quantitativa contínua.

$$5.2. \quad a = \frac{3}{16} = 0,1875$$

$$b = \frac{5}{16} = 0,3125$$

$$c = d = \frac{2}{16} = 0,125$$

5.3.  $b + c + d = 0,3125 + 0,125 + 0,125 = 0,5625$  (56,25%)

A partir da informação dada, podemos concluir que 56,25% das casas (mais de 50%) têm área superior ou igual a 120 m<sup>2</sup>.

No entanto, não podemos concluir que mais de 50% das casas disponíveis têm área superior a 120 m<sup>2</sup>.

6.1. a) 1 minuto = 60 segundos

A percentagem de chamadas com tempo de espera inferiores a um minuto foi 35,5%.

b) 2 minuto = 120 segundos

$$100\% - 85\% = 15\%$$

A percentagem de chamadas com tempo de espera de pelo menos dois minutos foi 15%.

6.2.

Classes	Freq. absoluta, $n_i$
[0, 30[	23
[30, 60[	48
[60, 90[	65
[90, 120[	34
[120, 150[	30

Cálculos auxiliares:

$$N_i \quad n_i$$

$$0,115 \times 200 = 23 \quad 23$$

$$0,355 \times 200 = 71 \quad 71 - 23 = 48$$

$$0,68 \times 200 = 136 \quad 136 - 71 = 65$$

$$0,85 \times 200 = 170 \quad 170 - 136 = 34$$

$$200 - 170 = 30$$

### Página 112

18.1.

Preço (€), $x_i$	$n_i$
9	2
10	2
11	2
12	3
13	1

18.2. A moda do conjunto de dados é 12 porque é o valor da variável com maior frequência absoluta,  $M_o = 12$ .



## 19.1

$x_i$	$n_i$
2	2
3	1
4	1
5	3
7	3
$a$	1

Há uma só moda se o valor de  $a$  for 5 ou 7,

$$a = 5 \vee a = 7.$$

19.2. Por exemplo, se  $a = 4$  então há exatamente duas modas (5 e 7).

19.3. Se  $a = 2$ , então há três modas (2, 5 e 7).

## Página 113

$$20.1. \quad \bar{x} = \frac{7+6+9+9+8+7+7+3}{8} = \frac{56}{8} = 7$$

A média é 7.

## 20.2.

$$\begin{aligned} \bar{x} &= \frac{16+17+19+15+17+19+14+16+17+20}{10} = \\ &= \frac{170}{10} = 17 \end{aligned}$$

A média é 17.

$$20.3. \quad \bar{x} = \frac{91+84+86+88+90+83}{6} = \frac{522}{6} = 87$$

A média é 87.

$$21. \quad \begin{aligned} \bar{x} &= \frac{36 \times 3 + 37 \times 2 + 38 \times 5 + 39 \times 10 + 40 \times 2 + 41 \times 2}{24} = \\ &= \frac{924}{24} = 38,5 \end{aligned}$$

O tamanho médio do calçado dos alunos da turma é 38,5.

$$22. \quad \bar{x} = 0 \times 0,1 + 1 \times 0,3 + 2 \times 0,2 + 3 \times 0,3 + 4 \times 0,1 = 2$$

Em média, cada família tem dois carros.

## Página 114

## 23.1.

Classes	$n_i$	Marca da classe
$[0, 10[$	8	$\frac{0+10}{2} = 5$
$[10, 20[$	7	$\frac{10+20}{2} = 15$
$[20, 30[$	6	$\frac{20+30}{2} = 25$
$[30, 40[$	8	$\frac{30+40}{2} = 35$
$[40, 50[$	5	$\frac{40+50}{2} = 45$
$[50, 60[$	4	$\frac{50+60}{2} = 55$
$[60, 70[$	3	$\frac{60+70}{2} = 65$
$[70, 80[$	1	$\frac{70+80}{2} = 75$

$$23.2. \quad \begin{aligned} \bar{x} &= \frac{8 \times 5 + 7 \times 15 + 6 \times 25 + 8 \times 35 + 5 \times 45}{42} + \\ &+ \frac{4 \times 55 + 3 \times 65 + 1 \times 75}{42} = \frac{1290}{42} \approx 30,7 \end{aligned}$$

A idade média dos pacientes que estiveram na clínica é, aproximadamente, 30,7 anos.

## 24.1.

Tempos	$x_i$	$n_i$
$[11, 13[$	12	4
$[13, 15[$	14	4
$[15, 17[$	16	3
$[17, 19[$	18	4

$$24.2. \quad \bar{x} = \frac{4 \times 12 + 4 \times 14 + 3 \times 16 + 4 \times 18}{15} = \frac{224}{15} \approx 14,9$$

Em média, o tempo realizado por cada aluno foi 14,9 segundos

## Página 115

25.1. Recorrendo à calculadora gráfica, conclui-se que em média foram vendidos 49 gelados por dia.

	A	B	C	D
			=OneVar(a[,1]	
1	25.	Título	Estatísticas ...	
2	26.	$\bar{x}$	49.0667	
3	26.	$\Sigma x$	1472.	
4	32.	$\Sigma x^2$	77084.	
5	32.	$s_x := s_n \dots$	12.9427	



25.2. a)

N.º de gelados	$x_i$	$n_i$
[20, 30[	25	3
[30, 40[	35	5
[40, 50[	45	7
[50, 60[	55	8
[60, 70[	65	7

$$b) \bar{x} = \frac{3 \times 25 + 5 \times 35 + 7 \times 45 + 8 \times 55 + 7 \times 65}{30} =$$

$$= \frac{1460}{30} \approx 49$$

O valor aproximado da média é de 49.

### Página 116

$$7.1. \quad 4 + 6 + 3 + 1 + 1 + 9 + 3 + 5 + 2 = 34$$

No total, Portugal marcou 34 golos.

$$7.2. \quad \bar{x} = \frac{34}{9} \approx 3,8$$

Em média, Portugal marcou 3,8 golos por jogo.

$$7.3. \quad 0 + 0 + 0 + 0 + 0 + 0 + 2 + 0 + 0 = 2$$

No total, Portugal sofreu 2 golos.

$$7.4. \quad \bar{x} = \frac{2}{9} \approx 0,2$$

Em média, Portugal sofreu 0,2 golos por jogo.

$$8.1. \quad \bar{x} = \frac{9 \times 2 + 10 \times 4 + 8 \times 5 + 5 \times 9}{9 + 10 + 8 + 5} = \frac{143}{32} \approx 4,5$$

A média do conjunto de dados é, aproximadamente, 4,5.

$$8.2. \quad \bar{x} = \frac{13 + 14 + 17 + 20 + 25 + 36 + 37 + 41 + 43 + 45}{10} =$$

$$= \frac{291}{10} = 29,1$$

A média do conjunto de dados é 29,1.

9.1.

Tempo, $x_i$	$n_i$
0	5
1	6
2	4
3	4
4	2
5	4

9.2. A moda do conjunto de dados é 1 porque é o valor da variável com maior frequência absoluta.  $M_0 = 1$

$$9.3. \quad \bar{x} = \frac{5 \times 0 + 6 \times 1 + 4 \times 2 + 4 \times 3 + 2 \times 4 + 4 \times 5}{25} =$$

$$= \frac{54}{25} = 2,16$$

A média do conjunto de dados é 2,16. Significa que, em média, cada aluno chegou 2,16 minutos atrasado à aula.

$$9.4. \quad \bar{x} = \frac{6 \times 1 + 4 \times 2 + 4 \times 3 + 2 \times 4 + 4 \times 5}{20} = \frac{54}{20} = 2,7$$

A média do novo conjunto de dados é 2,7. Significa que, em média, cada um dos alunos que não estava no início da aula chegou 2,7 minutos atrasado.

### Página 117

10.1.

$$\bar{x} = \frac{301 + 211 + 320 + 379 + 241 + 329 + 322 + 301}{26} +$$

$$\frac{239 + 379 + 304 + 319 + 233 + 251 + 262 + 219 + 371}{26} +$$

$$\frac{351 + 292 + 230 + 319 + 203 + 222 + 398 + 324 + 203}{26} =$$

$$= \frac{7183}{26} \approx 276$$

A média das áreas das superfícies dos municípios da amostra é, aproximadamente, 276 km<sup>2</sup>.

10.2.

Superfície (em km <sup>2</sup> )	Marca da classe $x_i$	Freq. absoluta $n_i$
[200, 250[	225	9
[250, 300[	275	3
[300, 350[	325	9
[350, 400[	375	5

$$10.3. \quad \frac{9 \times 225 + 3 \times 275 + 9 \times 325 + 5 \times 375}{26} = \frac{7650}{26} \approx 294$$

Utilizando a informação da tabela 12.2., conclui-se que uma aproximação da média é 294 km<sup>2</sup>. O valor obtido para aproximação da média é superior ao valor da média calculado em 10.1.



## 11.1. Fila A

$$\begin{aligned}\bar{x}_A &= \frac{2 \times 14 + 17 + 23 + 25 + 26 + 29 + 31 + 36}{17} + \\ &+ \frac{40 + 2 \times 43 + 62 + 64 + 65 + 2 \times 67}{17} = \\ &= \frac{666}{17} \approx 39\end{aligned}$$

## Fila B

$$\begin{aligned}\bar{x}_B &= \frac{15 + 16 + 2 \times 18 + 20 + 24 + 27 + 32 + 2 \times 35}{19} + \\ &+ \frac{36 + 39 + 41 + 44 + 2 \times 48 + 63 + 66 + 68}{19} = \frac{693}{19} \approx 36\end{aligned}$$

Em média, na fila A foram contadas, de 5 em 5 minutos, 39 pessoas na fila A e 36 pessoas na fila B.

## 11.2.

Classes	$n_i$	Marca da classe $x_i$
[10, 20[	4	15
[20, 30[	3	25
[30, 40[	5	35
[40, 50[	4	45
[50, 60[	0	55
[60, 70[	3	65

$$\frac{4 \times 15 + 3 \times 25 + 5 \times 35 + 4 \times 45 + 0 \times 55 + 3 \times 65}{19} = \frac{685}{19} \approx 36$$

Partindo dos dados agrupados, conclui-se que um valor aproximado da média de pessoas na fila B é 36.

12. Sequência 1:  $N_1 = 12$  e  $\bar{x}_1 = 4$ 

Sequência 2: Suas pontuações: 26 e

$$\bar{x}_2 = 3,25$$

$$\bar{x}_2 = \frac{26}{N_2} \Leftrightarrow N_2 = \frac{26}{3,25} \Leftrightarrow N_2 = 8$$

$$\text{Sequência 1+2: } \bar{x} = \frac{4 \times 12 + 26}{12 + 8} = \frac{74}{20} = 3,7$$

No conjunto das duas sequências, a pontuação média por lance foi de 3,7.

## Página 118

$$1. \quad \bar{x} = \frac{3 + 7 + 4 + 6 + 4}{5} = \frac{24}{5} = 4,8$$

A média dos dados é 4,8.

2. Adicionando 2 a cada um dos dados, obtém-se o seguinte conjunto de dados:  $\overline{5 \quad 9 \quad 6 \quad 8 \quad 6}$ 

$$\bar{x}_1 = \frac{5 + 9 + 6 + 8 + 6}{5} = \frac{34}{5} = 6,8$$

A nova média é igual à média inicial adicionada de 2 unidades.

3. Se, por exemplo, adicionarmos 3 a cada um dos dados, obtém-se o seguinte conjunto de dados:  $\overline{6 \quad 10 \quad 7 \quad 9 \quad 7}$ 

Conjetura: A média será  $4,8 + 3$ , ou seja, 7,8.

Calculando a média, tem-se:

$$\bar{x}_2 = \frac{6 + 10 + 7 + 9 + 7}{5} = \frac{39}{5} = 7,8$$

## 4. O novo conjunto de dados é:

$$x_1 + k \quad x_2 + k \quad x_3 + k \quad \dots \quad x_N + k$$

A média é:

$$\bar{x}_2 = \frac{x_1 + k + x_2 + k + x_3 + k + \dots + x_N + k}{N} =$$

$$= \frac{(x_1 + x_2 + x_3 + \dots + x_N) + k \times N}{N} =$$

$$= \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} + \frac{k \times N}{N} = \bar{x} + k$$

## 5. Por exemplo, multiplicando cada dado por 2 obtém-se o seguinte conjunto de dados:

$$\overline{6 \quad 14 \quad 8 \quad 12 \quad 8}$$

A média deste novo conjunto de dados é:

$$\bar{x}_3 = \frac{6 + 14 + 8 + 12 + 8}{5} = \frac{48}{5} = 9,6$$

A média deste novo conjunto de dados é o dobro da média inicial ( $\bar{x}_3 = 2 \times \bar{x}$ ).

Se, por exemplo, multiplicarmos cada um dos dados por 1,5, então iremos obter uma média igual a 7,2, pois  $1,5 \times 4,8 = 7,2$ .

## 6. O novo conjunto de dados é:

$$kx_1 \quad kx_2 \quad kx_3 \quad \dots \quad kx_N$$

A média deste novo conjunto de dados é dada por:

$$\frac{kx_1 + kx_2 + kx_3 + \dots + kx_N}{N} = \frac{k(x_1 + x_2 + x_3 + \dots + x_N)}{N} =$$

$$= k \times \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = k \times \bar{x}$$

## Página 119



- 1.1. a) Como a média dos salários mensais brutos, no setor dos Recursos Humanos, é 1400 €, então, se cada salário bruto tiver um aumento de 30 €, a média dos salários brutos mensais também terá um aumento de 30 €. Assim, será de 1430 €.
- b) Como a média dos salários mensais brutos, no setor dos Recursos Humanos, é 1400 €, então, se cada salário bruto tiver um aumento de 4%, a média passará a ser 1456 € (pois sofrerá um aumento de 4%).
- c) Como a média dos salários mensais brutos, no setor dos recursos humanos, é 1400 €, então, se cada salário bruto tiver uma redução de 2%, a média passará a ser 1372 €, pois também sofrerá uma redução de 2%.

$$1.2. \quad \bar{x} = \frac{10 \times 1400 - 1500 + 2 \times 1300}{11} \approx 1373$$

Após a alteração, a média dos salários brutos mensais será de 1373 €.

- 2.1. a)  $\bar{x} = 1200$   
Se os salários líquidos sofrerem um aumento mensal de 60 €, então a média passa a ser de 1260 € (pois também sofre um aumento de 60 €).
- b) Se os salários líquidos sofrerem um aumento mensal de 5%, então a média também sofre um aumento de 5%, passando a ser 1260 €.
- 2.2. a) Face ao comentário da Ana, conclui-se que o vencimento atual da Ana é igual ao valor da média dos salários líquidos mensais, ou seja, é de 1200 €.
- b) A proposta que beneficia o Gil é a B porque 5% de 1300 € são 65 € (valor superior aos 60 € da proposta A).  
A proposta que beneficia o Bernardo é a A porque 5% de 1040 € são 52 € (valor inferior aos 60 € da proposta A).

## Página 120

- 26.1. Vamos começar por ordenar os dados por ordem crescente:

$$6 \ 12 \ 13 \ 14 \ 14 \ \boxed{18} \ 21 \ 25 \ 27 \ 36 \ 40$$

$$M_e = 18$$

26.2. 
$$M_e = \frac{9 + 12}{2} = 10,5$$

- 27.1.

$x_i$	$n_i$	$N_i$
5	13	13
6	12	25
7	15	40

- 27.2. O número total de dados é 40 (par).  
O valor da mediana é a média dos dois dados "centrais", ou seja,  $M_e = \frac{x_{20} + x_{21}}{2} = \frac{6 + 6}{2} = 6$
28. O número total de dados é 125 (ímpar).  
O valor da mediana é igual ao dado que ocupa a posição central, ou seja,  $M_e = x_{63} = 3$

## Página 121

- 29.1.

Idade	$n_i$	$N_i$
1	2	2
3	6	8
4	8	16
5	5	21
6	3	24

- 29.2. O número total de dados é 24 (par).  
O valor da mediana é a média dos dois dados

$$\text{"centrais"}, \text{ ou seja, } M_e = \frac{x_{12} + x_{13}}{2} = \frac{4 + 4}{2} = 4$$

- 30.1.  $M_0 = 16$  porque é o dado com maior frequência absoluta.

$$\bar{x} = \frac{6 + 7 + 2 \times 8 + 9 + 2 \times 10 + 2 \times 11 + 2 \times 12 + 14 + 2 \times 15 + 3 \times 16 + 19 + 2 \times 20}{20} = \frac{255}{20} = 12,75$$

Como o número total de dados é 20 (par), o valor da mediana é a média dos dois "centrais",

$$\text{ou seja, } M_e = \frac{x_{10} + x_{11}}{2} = \frac{12 + 12}{2} = 12$$

$$M_0 = 16, \quad \bar{x} = 12,75 \text{ e } M_e = 12$$



30.2.  $x$ : pontuação do aluno em falta

$$\frac{255 + x}{21} = 13 \Leftrightarrow 255 + x = 273 \Leftrightarrow x = 18$$

A pontuação desse aluno é 18.

### Página 122

#### 31.1.

Alturas	$n_i$	$x_i$	$N_i$	$f_i$	$F_i$
[160, 165[	2	162,5	2	0,08	0,08
[165, 170[	4	167,5	6	0,16	0,24
[170, 175[	6	172,5	12	0,24	0,48
[175, 180[	7	177,5	19	0,28	0,76
[180, 185[	4	182,5	23	0,16	0,92
[185, 190[	2	187,5	25	0,08	1

31.2. Como o número total de dados é 25 (ímpar), então a classe a que pertence a mediana é [175, 180[, pois é a classe a que pertence o 13.º dado.

32. A mediana pertence à classe [40, 45[.

### Página 123

1.1. O valor errado nas classificações apresentadas é o 81.

1.2. Eliminando o *outlier* (valor 81) do conjunto de dados, tem-se:

- A distribuição tem moda 9 (classificação com maior frequência absoluta)

$$\bar{x} = \frac{2+3+5+2 \times 6+2 \times 7+1 \times 8+3 \times 9+2 \times 10+2 \times 11+2 \times 12+2 \times 13+14+15+18+20}{23} = \frac{230}{23} = 10$$

- $M_e = x_{12} = 10$

$$M_0 = 9; \bar{x} = 10; M_e = 10$$

#### 1.3.

Resultados dos alunos da turma A no concurso



2.  $M_e = 10 \rightarrow$  Gráfico 1;  $\bar{x} = M_e$

$M_e = 7 \rightarrow$  Gráfico 2;  $\bar{x} > M_e$

$M_e = 14,5 \rightarrow$  Gráfico 3;  $\bar{x} < M_e$

### Página 124

$$13. \quad \bar{x} = \frac{12+6+11+8+12+9+8+5+10}{9} = \frac{81}{9} = 9$$

$M_e$ :

1.º Ordenar os dados por ordem crescente

5 6 8 8  $\boxed{9}$  10 11 12 12

$$\bar{x} = M_e = 9$$

14.1. 4 5 6 7  $\boxed{7}$  7 8 9 10

$$\bar{x} = \frac{4+5+6+3 \times 7+8+9+10}{9} = \frac{63}{9} = 7$$

$$M_e = 9 \text{ e } \bar{x} = 7$$

14.2. Se substituirmos o dado com valor máximo por um com maior valor, então a mediana não se altera e a média aumenta.

14.3. Se substituirmos o dado com valor mínimo por um com menor valor, então a mediana não se altera e a média diminui.

$$15.1. \quad \bar{x} = 5,5 \Leftrightarrow \frac{6+3+7+1+a+9}{6} = 5,5 \Leftrightarrow$$

$$\Leftrightarrow \frac{26+a}{6} = 5,5 \Leftrightarrow 26+a = 33 \Leftrightarrow a = 7$$

$M_0 = 7$  (valor com maior frequência absoluta)

15.2. 1 3  $\boxed{6 \ 7}$  7 9

$$M_e = \frac{6+7}{2} = 6,5$$

16.1. a) Daniela:

$$M_0 = 1$$

$$\bar{x} = \frac{8 \times 1 + 6 \times 2 + 2 \times 3 + 3 \times 4 + 1 \times 5}{20} = \frac{43}{20} = 2,15$$

b) A mediana é inferior ao centro da distribuição de dados (3).

c) Como o número total de dados é 20 (par), o valor da mediana é a média dos dois dados "centrais", ou seja,

$$M_e = \frac{x_{10} + x_{11}}{2} = \frac{2+2}{2} = 2 \text{ e } M_e < 3$$

(confirma-se a resposta dada em b))



16.2. a)

N.º de ocupantes $x_i$	Freq. absoluta $n_i$	Freq. absoluta acum. $N_i$
1	9	9
2	12	21
3	7	28
4	5	33
5	2	35

b)  $M_0 = 2$  (valor com maior frequência absoluta)

$$c) \bar{x} = \frac{9 \times 1 + 12 \times 2 + 7 \times 3 + 5 \times 4 + 2 \times 5}{35} = \frac{84}{35} = 2,4$$

d) Como o número de dados é 35 (ímpar), o valor da mediana é igual ao dado que ocupa a posição central, ou seja,  $M_e = x_{18} = 2$ .

A percentagem de municípios que tem, pelo menos, oito escolas com ensino superior é 52%.

$$c) 100\% - 84\% = 16\%$$

16% de municípios tem, no mínimo, 12 escolas com ensino secundário

$$d) M_e - 50\% \quad Q_1 - 25\% \quad Q_3 - 75\%$$

Consultando a coluna de frequência relativa acumulada, conclui-se que a mediana e o 3.º quartil pertencem à classe  $[8, 12[$  e o 1.º quartil pertence à classe  $[4, 8[$ .

19. Dados ordenados por ordem crescente (sendo números positivos):  $a \ b \ c \ d \ e$

$$a = 2; \ e = 8; \ M_0 = 6; \ \bar{x} = 54$$

$$\bar{x} = 54 \Leftrightarrow \frac{a+b+c+d+e}{5} = 5,4 \Leftrightarrow$$

$$\Leftrightarrow 2+b+c+d+8 = 27 \Leftrightarrow b+c+d = 17$$

Como a moda é 6, sabe-se que dois dos valores  $b, c$  e  $d$  têm de ser iguais a 6 (não podem ser os três pois nesse caso a soma dos três seria 18).

$$6+6 = 12 \text{ e } 17-12 = 5$$

Então,  $b = 5$  e  $c = d = 6$ .

A mediana é o valor de  $c$ , ou seja, 6.

$$20.1. \bar{x} = \frac{7+9+5+6+7+23+8}{7} = \frac{65}{7} \approx 9,3$$

Ordenando os dados por ordem crescente, tem-se:

$$5 \ 6 \ 7 \ \boxed{7} \ 8 \ 9 \ 23$$

$$M_e = 7$$

$$\bar{x} \approx 9,3 \text{ e } M_e = 7$$

20.2. a) O outlier é 23. Retirando-o do conjunto de dados, temos: 5 6 7 7 8 9

$$\bar{x} = \frac{5+6+7+7+8+9}{6} = \frac{42}{6} = 7$$

$$M_e = \frac{7+7}{2} = 7$$

A média e a mediana do novo conjunto de dados são iguais a 7.

## Página 125

17.

dados: [9, 9, 11, 12, 12, 15, 16, 17]  
 média: 12.625  
 mediana: 12.0

$$18.1. \bar{x} = \frac{2 \times 1 + 2 \times 2 + 3 \times 4 + 3 \times 5 + 2 \times 6 + 7 + 2 \times 8 + 2 \times 9 + 3 \times 10 + 2 \times 11 + 13 + 14 + 2 \times 15}{25} = \frac{191}{25} = 7,64 \approx 8$$

Em média, por município, existem oito escolas com ensino secundário.

18.2.

Classes	Marca da classe $x_i$	Freq. absoluta $n_i$	Freq. absoluta a acum. $N_i$	Freq. relativa $f_i$	Freq. relativa acum. $F_i$
$[0, 4[$	2	5	5	0,2	0,2
$[4, 8[$	6	7	12	0,28	0,48
$[8, 12[$	10	9	21	0,36	0,84
$[12, 16[$	14	4	25	0,16	1

a) Classe modal:  $[8, 12[$  (classe com maior frequência absoluta);

$$\frac{5 \times 2 + 7 \times 6 + 9 \times 10 + 4 \times 14}{25} = \frac{198}{25} = 7,92$$

Um valor aproximado da média é 7,92.

b)  $100\% - 48\% = 52\%$



- b) O *outlier*, sendo neste caso o valor máximo do conjunto inicial de dados, faz aumentar a média, mas não altera a mediana (pois esta só depende do(s) dado(s) central(ais)).

### Página 126

33.1. a)  $P_{100} = 4,50$

O preço máximo de cada quilograma de castanhas nas 20 lojas dessa localidade, é 4,50 €.

b)  $P_{75} = 4,23$

Em pelo menos 75% das 20 lojas dessa localidade, ou seja, 15 lojas, o preço de cada quilograma de castanhas é inferior ou igual a 4,23 € e, no máximo, em 5 lojas (25%) o preço é superior a 4,23 €.

33.2.  $P_{80} = 4,30$

Significa que, em pelo menos 80% das vinte lojas dessa localidade, ou seja, dezasseis lojas, o preço de cada quilograma de castanhas é inferior ou igual a 4,30 € e, no máximo, em quatro lojas o preço é superior a 4,30 €.

Opção (C)

### Página 127

- 1.1. Consultando a tabela do comprimento, conclui-se que a afirmação é verdadeira.
- 1.2. Consultando a tabela do peso, conclui-se que a afirmação é verdadeira.
- 1.3. Consultando a tabela do comprimento, conclui-se que a afirmação é falsa.  
Como a Sofia mede 80 cm e tem menos de 18 meses, o seu comprimento está acima do percentil 50.
- 2.1. Atendendo a que a Carolina mede 75 cm e está no percentil 25, segundo a tabela, estima-se que tenha 15 meses.
- 2.2. Atendendo a que a Carolina tem 9 kg e está no percentil 10, segundo a tabela, estima-se que tenha 15 meses.
3. A Joana aos 5 meses estava, em termos de comprimento, no percentil 25. Consultando a tabela, conclui-se que nessa altura media 62 cm.

Passado meio ano, ou seja, aos 11 meses a Joana estava no percentil 10. Consultando a tabela, conclui-se que nessa altura media 69 cm. Então, estima-se que, nesse período a Joana cresceu 7 cm.

4. Como a Catarina mede 70 cm e está no percentil 50, segundo a tabela terá 9 meses. Relativamente ao peso encontra-se no percentil 5, logo, segundo a tabela, estima-se que tem 7 kg.

### Página 128

- 34.1. Como o percentil 100 corresponde ao valor máximo dos dados, na amostra A é 13,5 e na amostra B é 52.

A:  $P_{100} = 13,5$  ; B:  $P_{100} = 52$

#### 34.2. Conjunto de dados A:

Começa-se por ordenar os dados, por ordem crescente:

8,7 8,9 9 9,1 9,2 9,6 9,9 10,1  
10,6 10,6 10,8 11 12,3 13,1 13,5

A amostra tem dimensão 15.

50% de 15:  $0,5 \times 15 = 7,5$  (não é número inteiro, sendo 8 o inteiro seguinte)

Então,  $P_{50} = \overline{x_8} = 10,1$

#### Conjunto de dados B:

Ordenando os dados, por ordem crescente, tem-se:

48 49 49 50 50 50 50 50 51 51 52 52

Neste caso a amostra tem dimensão 12.

50% de 12:  $0,5 \times 12 = 6$  (é número inteiro)

Então,  $P_{50} = \frac{x_6 + x_7}{2} = \frac{50 + 50}{2} = 50$

A:  $P_{50} = 10,1$  ; B:  $P_{50} = 50$

#### 34.3. Conjunto de dados B

No máximo, 20% das caixas têm um número de pionés superior ao valor de  $P_k$  é o mesmo que dizer que, pelo menos, 80% das caixas têm um número de pionés menor ou igual a  $P_k$ .

Assim sendo,  $k = 80$ .



### Página 129

1. Dimensão do conjunto de dados: 30  
Pelo menos em 25% dos dias do mês, o número de reclamações diárias apresentadas *online* foi inferior ou igual a 4 e, no máximo, em 75% dos dias do mês o número de reclamações diárias apresentadas *online* foi superior a 4.
2. O 3.º quartil corresponde ao percentil de ordem 75.  
 $Q_3 = P_{75} = 11$   
Significa que em pelo menos 75% dos dias do mês o número de reclamações diárias apresentadas *online* foi inferior ou igual a 11 e, no máximo, em 25% dos dias do mês o número de reclamações diárias apresentadas *online* foi superior a 11.
3.  $P_{30} = 4,5$  e  $P_{65} = 9$   
Entre  $P_{30}$  e  $P_{65}$  estão 10 dados (5, 5, 5, 6, 6, 7, 7, 7, 8, 8)

### Página 130

- 1.1. a) Recorrendo à tecnologia, conclui-se que  $P_{10} = 4,5$ .  
b)  $\tilde{x} = P_{50}$   
Recorrendo à tecnologia, conclui-se que  $\tilde{x} = 7$   
c) Recorrendo à tecnologia, conclui-se que  $P_{75} = 8$
- 1.2. Recorrendo à tecnologia, conclui-se que  $P_{90} = 9$ .  
x: número de automóveis da fotografia  
 $x > P_{90} \Leftrightarrow x > 9$   
Só há uma fotografia em que o número de automóveis é superior a  $P_{90}$  (a que tem 10 automóveis).
- 1.3. a) Em pelo menos 70% dos registos, o número de automóveis contabilizados foi menor ou igual a 8.  
**Cálculo auxiliar:**  
 $P_{70} = 8$   
b) No máximo, em 40% dos registos, o

número de automóveis contabilizados foi superior a 7,5.

#### Cálculos auxiliares:

$$100\% - 40\% = 60\% \text{ e } P_{60} = 7,5$$

- 2.1. Recorrendo à tecnologia, conclui-se que  $P_{50} = 74,5$ .  
Significa que pelo menos 50% dos veículos observados seguiam a uma velocidade inferior ou igual a 74,5 km/h.
- 2.2. a) Como  $P_{80} = 88$ , a afirmação dada é falsa.  
Pode afirmar-se que, no máximo, em 20% dos veículos observados, a velocidade a que seguiam é superior a 88 km/h.  
b) Como  $P_{70} = 83,5$ , a afirmação dada é falsa.  
Pode afirmar-se que, em pelo menos 70% dos veículos observados, a velocidade a que seguiam é inferior ou igual a 83,5 km/h.  
c) Como  $P_{90} = 95$ , a afirmação dada é verdadeira.

### Página 131

#### 35.1.

Classes	Freq. abs.	Marca da classe
[0, 5[	12	2,5
[5, 10[	9	7,5
[10, 15[	8	12,5
[15, 20[	11	17,5

$$\bar{x} = \frac{12 \times 2,5 + 9 \times 7,5 + 8 \times 12,5 + 11 \times 17,5}{40} =$$

$$= \frac{390}{40} = 9,75$$

Um valor aproximado da média é 9,75.

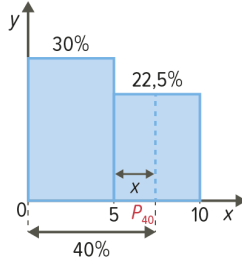
#### 35.2.

Classes	Freq. abs.	Freq. rel. (%)	Freq. rel. acum. (%)
[0, 5[	12	30	30
[5, 10[	9	22,5	52,5
[10, 15[	8	20	72,5
[15, 20[	11	27,5	100



- 35.3. Por observação da coluna das frequências relativas acumuladas da tabela, verifica-se que o percentil 40 pertence à classe  $[5, 10[$ .

35.4.



Admitindo que os dados têm uma distribuição uniforme nas classes.

$$P_{40} = 5 + x$$

Amplitude de classe: 5

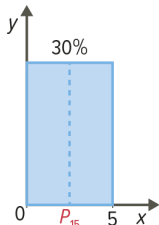
O valor de  $x$  é tal que:

$$\frac{5}{0,225} = \frac{x}{0,1}$$

$$\Leftrightarrow x = \frac{0,5}{0,225} \Leftrightarrow x = \frac{20}{9}$$

$$\text{Então, } P_{40} = 5 + \frac{20}{9} \approx 7,22.$$

35.5.  $P_{15} = ?$



Admitindo que os dados têm uma distribuição uniforme nas classes, e atendendo ao esquema ao lado, conclui-se que  $P_{15} = 2,5$ .

### Página 132

21.1.  $M_0 = 4$  (valor com maior frequência absoluta)

$$\bar{x} = \frac{7 \times 0 + 9 \times 1 + 6 \times 2 + 3 \times 3 + 12 \times 4 + 1 \times 5 + 1 \times 6 + 1 \times 20}{40} = \frac{109}{40} = 2,725$$

Dimensão da amostra: 40 (par)

$$M_e = \frac{x_{20} + x_{21}}{2} = \frac{2 + 2}{2} = 2$$

$$M_o = 4; \bar{x} = 2,725; M_e = 2$$

21.2. O *outlier* é 20 e não tinha qualquer efeito na moda e na mediana. Como o *outlier* é o valor máximo da amostra, provoca um aumento na média.

21.3.  $P_{20} = ?$

$$0,2 \times 40 = 8 \text{ (é número inteiro)}$$

$$\text{Então, } P_{20} = \frac{x_8 + x_9}{2} = \frac{1 + 1}{2} = 1.$$

No contexto, indica-nos que pelo menos 20% dos estudantes assistem, na última semana, a um ou nenhum dos episódios da série.

$P_{75} = ?$

$$0,75 \times 40 = 30 \text{ (é número inteiro)}$$

$$\text{Então, } P_{75} = \frac{x_{30} + x_{31}}{2} = \frac{4 + 4}{2} = 4.$$

No contexto, indica-nos que pelo menos 75% dos estudantes assistem, na última semana, a quatro ou menos episódios da série.

21.4. a)  $k = 80$

b) Afirmar que, no máximo, 40% dos alunos assistiram a mais do que  $P_k$  episódios da série é o mesmo que dizer que pelo menos 60% dos alunos assistiu a  $P_k$  episódios da série ou menos. Assim,  
 $k = 60$ .

### Página 133

22.1.  $P_{75} = 635\,000$

40 imóveis

$$0,75 \times 40 = 30$$

Como  $P_{75} = 635\,000$ , conclui-se que pelo menos 30 imóveis custam, no máximo, 635 000.

Opção (C)

22.2.  $P_{20} = 341\,000$  e  $M_e = 494\,000$

a) Como  $P_{20} = 341\,000$  e o valor mínimo é 204 000, a afirmação é falsa. Pode afirmar-se que o preço de pelo menos 20% das casas varia entre 204 000 e 341 000.



b) A afirmação é falsa.

$$P_{20} = 341 \text{ e } P_{50} = 494$$

$$50\% - 20\% = 30\% \text{ e } 0,3 \times 40 = 12$$

Sabe-se apenas que o número de imóveis cujo preço varia entre 341 000 e 494 000 é inferior ou igual a 12, não sabendo o número exato.

### 23.1. Ano 2000

Classes	Freq. rel.	Freq. rel. acum.
[0, 5[	38%	38%
[5, 10[	36%	74%
[10, 15[	20%	94%
[15, 20[	5%	99%
[20, 25[	1%	100%

### Ano 2022

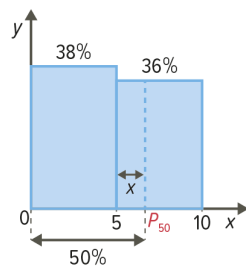
Classes	Freq. rel.	Freq. rel. acum.
[0, 5[	18%	18%
[5, 10[	20%	38%
[10, 15[	18%	56%
[15, 20[	19%	75%
[20, 25[	25%	100%

23.2. a) Ano 2000:  $M_e \in [5, 10[$

Ano 2022:  $M_e \in [10, 15[$

b) **Ano 2000**

Admitindo que os dados têm uma distribuição uniforme nas classes e atendendo ao esquema seguinte, tem-se:



$$M_e = P_{50} = 5 + x$$

O valor de  $x$  é tal que:

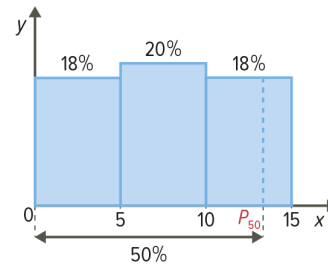
$$\frac{5}{0,36} = \frac{x}{0,12}$$

$$\Leftrightarrow x = \frac{5 \times 0,12}{0,36} \Leftrightarrow x = \frac{5}{3}$$

Então,  $P_{50} = 5 + \frac{5}{3} \approx 6,67$ ;

**Ano 2022**

Admitindo que os dados têm uma distribuição uniforme nas classes e atendendo ao esquema seguinte, tem-se:



$$M_e = P_{50} = 10 + x$$

O valor de  $x$  é tal que:

$$\frac{5}{0,18} = \frac{x}{0,12}$$

$$\Leftrightarrow x = \frac{5 \times 0,12}{0,18} \Leftrightarrow x = \frac{10}{3}$$

Então,  $P_{50} = 10 + \frac{10}{3} = \frac{40}{3} \approx 13,33$ .

Ano 2000:  $M_e \approx 6,67$

Ano 2022:  $M_e \approx 13,33$

c) Ano 2000: A mediana (6,67) é inferior à média (7).

Ano 2022: A mediana (13,33) é superior à média (13).

23.3. Ano 2022:  $P_{75} = 20$  e  $P_{90} = 23$

a)  $P_{75} = 20$  – pelo menos 75% dos carros em circulação em 2022 têm 20 anos ou menos.

$$0,75 \times 5,6 \times 10^6 = 4,2 \times 10^6$$

Então, em 2022, no mínimo, há 4,2 milhões de carros em circulação com, no máximo, 20 anos.

b)  $P_{90} = 23$  – no máximo 10% dos carros em circulação em 2022, têm mais de 23 anos.

$$0,1 \times 5,6 \times 10^6 = 0,56 \times 10^6 = 560 \times 10^3$$

Então, em 2022, o número máximo de carros em circulação com pelo menos 23 anos é 560 milhares.



## Página 134

### 36.1. Situação A

#### Equipa 1

Valor mínimo: 148

Valor máximo: 182

$$\bar{x} = \frac{148 + 160 + 165 + 170 + 182}{5} = \frac{825}{5} = 165$$

$$M_e = 165$$

#### Equipa 2

Valor mínimo: 161

Valor máximo: 170

$$\bar{x} = \frac{162 + 165 + 167 + 170 + 161}{5} = \frac{825}{5} = 165$$

$$M_e = 165$$

### Situação B

#### Loja 1

Valor mínimo: 1,10

Valor máximo: 8,00

$$\bar{x} = \frac{1,10 + 2,20 + 1,40 + 3,10 + 5,00 + 4,20 + 3,40 +$$

$$+ 3,80 + 4,10 + 3,50 + 4,00 + 8,00}{12} = \frac{43,8}{12} = 3,65$$

Ordenando os valores por ordem crescente, tem-se:

1,10 1,40 2,20 3,10 3,40 3,50 3,80 4,00

4,10 4,20 5,00 8,00

$$M_e = \frac{3,50 + 3,80}{2} = 3,65$$

#### Loja 2

Valor mínimo: 2,60

Valor máximo: 5,30

$$\bar{x} = \frac{2,60 + 3,60 + 3,20 + 4,50 + 2,80 + 2,80 + 3,10 + 3,70 +$$

$$+ 3,80 + 4,60 + 3,80 + 5,30}{12} = \frac{43,8}{12} = 3,65$$

Ordenando os valores por ordem crescente, tem-se:

2,60 2,80 2,80 3,10 3,20 3,60 3,70 3,80

4,50 4,60 5,30

$$M_e = \frac{3,60 + 3,70}{2} = 3,65$$

Conclusão: Analisando os valores obtidos, constata-se que em cada conjunto de dados a média é igual à mediana e iguais entre si.

36.2. O conjunto de dados que transmite a ideia de maior dispersão é o referente à equipa 1, da situação A.

## Página 135

37.1. - Valor mínimo: 8

- Valor máximo: 21

- Amplitude:  $21 - 8 = 13$

- Quartis: 8 11 11 15 17 20 21

- Amplitude interquartil:

$$Q_3 - Q_1 = 20 - 11 = 9$$

37.2. - Valor mínimo: 37

- Valor máximo: 42

- Amplitude:  $42 - 37 = 5$

- Quartis:

$$37 \quad \left[ \frac{37+37}{2} = 37 \right] \quad 38 \quad \left[ \frac{38}{2} = 38 \right] \quad 38 \quad \left[ \frac{38+38}{2} = 38 \right] \quad 42$$

- Amplitude interquartil:

$$Q_3 - Q_1 = 38 - 37 = 1$$

37.3. - Valor mínimo: 1

- Valor máximo: 6

- Amplitude:  $6 - 1 = 5$

- Quartis:

$$1 \quad 1 \quad 1 \quad \left[ \frac{2}{2} = 1 \right] \quad 3 \quad 3 \quad \left[ \frac{3+3}{2} = 3 \right] \quad 4 \quad 4 \quad \left[ \frac{5}{2} = 2,5 \right] \quad 5 \quad 5 \quad 6$$

- Amplitude interquartil:  $Q_3 - Q_1 = 5 - 2 = 3$

37.4. - Valor mínimo: 21

- Valor máximo: 34

- Amplitude:  $34 - 21 = 13$

- Quartis:

$$21 \quad 24 \quad 24 \quad \left[ \frac{25}{2} = 12,5 \right] \quad 26 \quad 26 \quad \left[ \frac{27+27}{2} = 27 \right] \quad 27 \quad 28 \quad \left[ \frac{28}{2} = 28 \right] \quad 30 \quad 31 \quad 34$$

- Amplitude interquartil:

$$Q_3 - Q_1 = 28 - 25 = 3$$

### 38.1. Empresa A

- Amplitude:  $80 - 24 = 56$

- Amplitude interquartil:  $58 - 30 = 28$

**Empresa B**

- Amplitude:  $60 - 38 = 22$
- Amplitude interquartil:  $56 - 44 = 12$

**38.2.** A empresa onde há menos variabilidade salarial é a B.

**Página 136**

- 1.1. a)** Se a amplitude, das pontuações obtidas nos quatro lançamentos é zero, então significa que saiu a mesma pontuação nos quatro lançamentos. Assim, a amplitude interquartil das pontuações obtidas é zero.
- b)** Não é possível que a média das pontuações obtidas seja 3,5 pois, como saiu a mesma pontuação nos quatro lançamentos, a média é igual ao valor da pontuação de cada lançamento (1, 2, 3, 4, 5 ou 6).
- c)** Sabendo que a média das pontuações obtidas é 3, então o Vasco obteve as pontuações 3, 3, 3 e 3.
- 1.2. a)** Sabendo que a média interquartil é zero, então  $Q_3 = Q_1$ . Logo, a mediana também tem de ser igual ao 1.º quartil e ao 3.º quartil. Não é possível a mediana seja diferente dos restantes quartis.  $Q_3 = Q_1$  e como  $Q_1 \leq M_e \leq Q_3$ , então  $M_e = Q_1 = Q_3$
- b)** Não se pode concluir que a amplitude é zero.  
Exemplo:

$$1 \quad \boxed{2} \quad \boxed{2} \quad \boxed{2} \quad 5$$

$Q_1 \quad M_e = \frac{2+2}{2} = 2 \quad Q_3$

Amplitude interquartil:  $Q_3 - Q_1 = 2 - 2 = 0$

Amplitude:  $5 - 1 = 4$

A amplitude interquartil é zero e a amplitude é diferente de zero.

- 1.3. Conclusão 1:**  
Se a amplitude de um conjunto de dados é zero então não existe variabilidade.
- Conclusão 2:**

Se a amplitude interquartil de um conjunto de dados é zero nada se pode concluir quanto à amplitude desse conjunto de dados.

- 2.1.** A dimensão da amostra é 21.
- 2.2.** Recorrendo à tecnologia, conclui-se que:  
Amplitude:  $8:00 - 1:00 = 7:00$   
Amplitude interquartil:  
 $Q_3 - Q_1 = 4:00 - 1:00 = 3:00$
- 2.3.** Como o 1.º quartil é igual ao valor mínimo (1), então os dados cujo valor é inferior ao 1.º quartil são todas iguais a 1. Portanto a afirmação é verdadeira pois os dados do conjunto A não apresentam variabilidade (são todos iguais a 1).

**Página 137**

**24.1.**  $\bar{x} = \frac{2 \times 12 + 13 + 14 + 16 + 21 + 22 + 3 \times 30 + 33 +$

$$\frac{+44 + 47 + 48 + 2 \times 51 + 75 + 77 + 78 + 81}{20} =$$

$$= \frac{785}{20} = 39,25$$

$$N = 20$$

$$M_e = \frac{x_{10} + x_{11}}{2} = \frac{30 + 33}{2} = 31,5$$

A média do conjunto de dados é 39,25 e a mediana é 31,5.

$$\bar{x} = 39,25 \text{ e } M_e = 31,5$$

**24.2.** Amplitude:  $81 - 12 = 69$

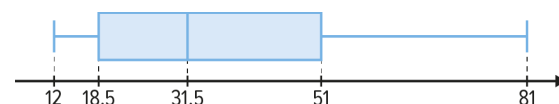
$$\text{Quartis: } Q_1 = \frac{x_5 + x_6}{2} = \frac{16 + 21}{2} = 18,5$$

$$Q_3 = \frac{x_{15} + x_{16}}{2} = \frac{51 + 51}{2} = 51$$

Amplitude interquartil:

$$Q_3 - Q_1 = 51 - 18,5 = 32,5$$

**24.3.**



**25.1. Cidade A:**  $M_e = 22$

**Cidade B:**  $M_e = 20$

**25.2. Cidade A**Amplitude:  $24 - 18 = 6$ Amplitude interquartil:  $23 - 19 = 4$ 

Cidade B

Amplitude:  $26 - 13 = 13$ Amplitude interquartil:  $24 - 16 = 8$ 

- 25.3. a)** As temperaturas mínimas e máximas são muito diferentes. A amplitude e a amplitude interquartil é muito maior na cidade B do que na cidade A
- b)** As temperaturas foram mais regulares ao longo do ano na cidade A.

$$26.1. \quad \bar{x} = \frac{53 + 51 + 42 + 32 + 30 + 50 + 26 + 18 + 56}{9} = \frac{315}{9} = 35$$

Ordenando os dados, por ordem crescente, tem-se:

18 26 30 32  $\boxed{42}$  50 51 53 58  
 $M_e$

A média dos salários é de 35 milhares de euros e a mediana é de 42 milhares de euros.

- 26.2.**
- Amplitude:
- $58 - 18 = 40$
- ;

$$\text{Quartis: } Q_1 = \frac{26 + 30}{2} = 28 ; Q_3 = \frac{51 + 53}{2} = 52$$

Amplitude interquartil:  $52 - 28 = 24$ 

- 26.3.**
- Só há dois elementos deste gabinete que auferem anualmente, no máximo, 28 000 euros.

$$\frac{2}{9} \times 100 \approx 22,2, \text{ logo a percentagem pedida é,}$$

aproximadamente, 22,2%.

- 26.4. a)** O motivo para excluir o salário do presidente no cálculo das medidas das alíneas anteriores é ele ser um *outlier*.
- b)** A medida de dispersão mais afetada seria a amplitude porque depende dos valores extremos do conjunto de dados.
- c)** No caso de o salário do presidente ser incluído, a medida que seria mais adequada para localizar o centro da distribuição dos dados era a mediana pois a média é muito sensível a valores extremos.

**Página 138**

- 39.1.**
- Amplitude:
- $102 - 96 = 6$
- ;

$$\bar{x} = \frac{96 + 101 + 101 + 98 + 102 + 99}{6} = \frac{597}{6} = 99,5$$

- 39.2. a)**
- O dado em que o valor absoluto do desvio em relação à média é maior é 96.

$$|96 - 99,5| = |-3,5| = 3,5$$

- b)**
- O dado em que o valor absoluto do desvio em relação à média é menor é 99.

$$|99 - 99,5| = |-0,5| = 0,5$$

40.

$x_i$	$x_i - \bar{x}$
a	-4
b	-2
4	c
d	4

Como a soma dos desvios em relação à média é igual a zero, conclui-se que  $c = 2$ .

$$c = 2 \Leftrightarrow 4 - \bar{x} = 2 \Leftrightarrow \bar{x} = 2$$

$$a - 2 = -4 \Leftrightarrow a = -2$$

$$b - 2 = -2 \Leftrightarrow b = 0$$

$$d - 2 = 4 \Leftrightarrow d = 6$$

Então, tem-se:

$x_i$	$x_i - \bar{x}$
-2	-4
0	-2
4	2
6	4

**Página 139**

$$41.1. \quad \bar{x} = \frac{3 + 4 + 4 + 5 + 8}{5} = \frac{24}{5} = 4,8$$

$$s = \sqrt{\frac{(3 - 4,8)^2 + (4 - 4,8)^2 + (4 - 4,8)^2 + (5 - 4,8)^2 + (8 - 4,8)^2}{5 - 1}}$$

$$\approx 1,92$$

- 41.2. a)**
- Para aumentar o desvio padrão, convém substituir um dos dados de forma a que a dispersão de dados seja maior, ou seja, alterar um dos extremos.

Por exemplo, substituindo o dado maior (8) por 14, tem-se:

$$\bar{x} = \frac{3 + 4 + 4 + 5 + 14}{5} = \frac{30}{5} = 6$$



Neste caso,

$$s = \sqrt{\frac{(3-6)^2 + (4-6)^2 + (4-6)^2 + (5-6)^2 + (14-6)^2}{5-1}}$$

$$\approx 4,53$$

Confirma-se que o desvio-padrão aumentou.

- b) Para diminuir o desvio padrão, convém substituir um dos extremos de modo a que a dispersão de dados seja menor. Por exemplo, substituindo o dado menor (3) por 4, tem-se:

$$\bar{x} = \frac{4+4+4+5+8}{5} = \frac{25}{5} = 5$$

Neste caso,

$$s = \sqrt{\frac{3 \times (4-5)^2 + (5-5)^2 + (8-5)^2}{5-1}} \approx 1,73$$

Confirma-se que o desvio-padrão diminuiu.

#### 42.1. a) Amostra A

$$\text{Amplitude: } 9 - 4 = 5$$

Amostra B

$$\text{Amplitude: } 15 - 5 = 10$$

A amplitude da amostra B é maior que a amplitude da amostra A.

#### b) Amostra A

$$Q_1 = \frac{4+7}{2} = 5,5$$

$$Q_3 = \frac{8+9}{2} = 8,5$$

$$\text{Amplitude interquartil: } 8,5 - 5,5 = 3$$

Amostra B

$$\text{Amplitude interquartil: } 10 - 6 = 4$$

A amplitude interquartil é maior na amostra B.

#### c) Amostra A: $M_e = 8$

Amostra B:  $M_e = 8$

As medianas das amostras A e B são iguais.

#### d) Amostra A: $\bar{x} = \frac{4+7+8+8+9}{5} = \frac{36}{5} = 7,2$

Amostra B:  $\bar{x} = 8,6$

A média da amostra B é superior à média da amostra A.

#### 42.2. Parece haver maior dispersão na amostra B.

#### 42.3. Amostra A:

$$s_A = \sqrt{\frac{(4-7,2)^2 + (7-7,2)^2 + 2 \times (8-7,2)^2 + (9-7,2)^2}{5-1}}$$

$$\approx 1,92$$

Amostra B:

$$s_A = 3,06$$

$s_B > s_A$ , o que confirma que há maior dispersão na amostra B.

### Página 140

#### 1.1. a) $\bar{x}_s$ : média das tensões sistólica

$\bar{x}_d$ : média das tensões diastólica

Recorrendo à calculadora gráfica, conclui-se que  $\bar{x}_s \approx 130,1$  e  $\bar{x}_d \approx 82,5$

Atendendo ao quadro inicial, conclui-se que, neste caso, a classificação da tensão arterial é Pré-HTA.

#### b) $M_e(s)$ : mediana das tensões sistólica

$M_e(d)$ : mediana das tensões diastólica

Recorrendo à calculadora gráfica, conclui-se que  $M_e(s) = 129$  e  $M_e(d) = 83,5$

Atendendo ao quadro inicial, conclui-se que, neste caso, a classificação da tensão arterial é Pré-HTA.

#### 1.2. O registo de tensão sistólica que apresenta um maior desvio, em valor absoluto, em relação à média é 100.

Desvio (em valor absoluto) em relação à média:

$$|100 - 130,01| = |-30,01| = 30,1$$

#### 2. Recorrendo à calculadora gráfica, obtemos os seguintes valores:

$$s_s \approx 13,43 \text{ e } s_d \approx 13,69$$

Como a meta é diminuir o desvio-padrão em, pelo menos, 5 unidades em cada conjunto de registo, então os desvios-padrão máximos para atingir essa meta são:

$$s_s \approx 8,43 \text{ e } s_d \approx 8,69$$



## Página 141

$$1. \quad \bar{x} = \frac{60 + 66 + 2 \times 73 + 65 + 88 + 84 + 90 + 77}{12} + \frac{79 + 64 + 87}{12} = \frac{900}{12} = 75$$

Cálculo auxiliar:

$$(60 - 75)^2 + (66 - 75)^2 + 2 \times (73 - 75)^2 + (65 - 75)^2 + (88 - 75)^2 + (84 - 75)^2 + (90 - 75)^2 + (77 - 75)^2 + (79 - 75)^2 + (64 - 75)^2 + (81 - 75)^2 = 1066$$

$$s = \sqrt{\frac{1066}{12-1}} \approx 9,84$$

$$\bar{x} = 75 \text{ e } s = 9,84$$

- Se adicionarmos um número real  $b$  a cada dado da amostra, então a média do "novo" conjunto de dados é dado pela expressão  $75 + b$ .
- Utilizando a tecnologia para preencher a tabela, tem-se:

Conjunto de dados						
Amostra inicial	$x_i$	$x_i + 1$	$x_i + 2$	$x_i + 3$	$x_i + 4$	$x_i + b$
<b>Desvio-padrão (com 2 c.d.)</b>						
	9,84	9,84	9,84	9,84	9,84	9,84

- Se multiplicarmos por um número real  $b$ , não nulo, cada dado da amostra, então a média deste "novo" conjunto de dados é dado pela expressão  $75b$ .
- Utilizando a tecnologia para preencher a tabela, tem-se:

Conjunto de dados						
Amostra inicial	$x_i$	$2x_i$	$3x_i$	$4x_i$	$5x_i$	$bx_i$
<b>Desvio-padrão (com 2 c.d.)</b>						
	9,84	19,68	29,52	39,36	49,2	9,84b

- Analisando as tabelas construídas em 3. e em 5., conclui-se que:
  - se adicionarmos um número real  $b$  a todos os dados de um conjunto de dados cujo desvio-padrão é  $s$ , então o desvio-padrão do novo conjunto de dados mantém-se (é igual a  $s$ );
  - se multiplicarmos todos os dados de um conjunto de dados cujo desvio-padrão é  $s$  por um número real  $b$ , diferente de zero, então o desvio-padrão do novo conjunto de dados é igual a  $|b|s$ .

## Página 142

- Se  $s = 5,4$  e cada elemento do conjunto de dados foi multiplicado por 2,5, então o desvio-padrão do novo conjunto de dados é  $2,5 \times 5,4$ , ou seja, 13,5.  
Opção (B)
- Seja  $k$  o número pelo qual se multiplicou cada elemento de  $A$ .  
Então,  $k = \frac{3}{2} = 1,5$
- Como  $\bar{x}_A = 2,125$ , então  
 $\bar{x}_B = 1,5\bar{x}_A = 1,5 \times 2,125 = 3,1875$   
 $\bar{x}_B = 3,1875$
- Como  $s_B = 1,275$ , então  
 $s_A = s_B : 1,5 = 1,275 : 1,5 = 0,85$   
 $s_A = 0,85$

## Página 143

- O gráfico que permite conhecer a amplitude e a amplitude interquartil das precipitações máximas diárias é o gráfico 2.  
Amplitude:  $89,6 - 8,5 = 81,1$   
Amplitude interquartil:  $45,7 - 29,7 = 16$
- A afirmação é verdadeira.  
O gráfico que permite justificar a resposta é o gráfico 2.
- Como  $M_e = P_{50} = 36,3$  e  $Q_3 = P_{75} = 45,7$ , então sabe-se que  $36,3 < P_{65} < 45,7$ .  
Conclui-se que, dos quatro valores apresentados, o único que pode representar o percentil de ordem 65 é o 38.  
Opção (D)  
O gráfico que permite justificar a opção escolhida é o 2.
- A afirmação é verdadeira.  
O gráfico que permite justificar a resposta é o gráfico 1.
- Excluindo o valor extremo 89,6 do conjunto de dados, sabe-se que a amplitude, a média, a mediana, a amplitude interquartil e o desvio-padrão vão diminuir.  
Efeito: diminuição.



- 5.2. Recorrendo à tecnologia, conclui-se que:  
 Amplitude:  $59,9 - 8,5 = 51,4$   
 Média: 34,7  
 Mediana: 36,2  
 Amplitude interquartil:  $43,05 - 26,6 = 16,45$   
 Desvio-padrão: 13,4

#### Página 144

- 1.1. a) Variável qualitativa: tipo de garagem  
 b) Variável quantitativa discreta: número de quartos  
 c) Variáveis quantitativas contínuas: área útil e preço
- 1.2. No caso de se utilizarem as categorias T1 a T4 para a indicação do número de quartos, os dados recolhidos seriam do tipo qualitativa.

- 1.3. a)

$x_i$	$n_i$	$N_i$	$f_i$ (%)	$F_i$ (%)
1	1	1	10%	10%
2	4	5	40%	50%
3	1	6	10%	60%
4	4	10	40%	100%

- b) Há 9 apartamentos ( $10 - 1$ ) que têm, pelo menos, dois quartos.  
 c) 50% dos apartamentos têm, no máximo, dois quartos.
- 1.4. Há 4 apartamentos do tipo T2, dos quais apenas 1 tem garagem individual.  
 $\frac{1}{4} \times 100\% = 25\%$ , assim sendo, 25% de apartamentos T2 têm garagem individual.

- 2.1. a)

Classe	$n_i$	$N_i$	$f_i$ (%)	$F_i$ (%)
[90, 130[	1	1	10%	10%
[130, 170[	5	6	50%	60%
[170, 210[	1	7	10%	70%
[210, 250[	3	10	30%	100%

- b) Pela leitura da tabela anterior, conclui-se que 60% dos apartamentos têm, pelo menos, 170 m<sup>2</sup> de área útil.  
 c) A afirmação é falsa. Analisando os dados recolhidos constata-se que há 6 apartamentos com área útil inferior a 170 m<sup>2</sup>, o que corresponde a uma percentagem de 60%.

- 2.2. a) A percentagem de apartamentos que custam menos de 200 000 é 50% ( $30\% + 20\%$ ), logo são 5 apartamentos.  
 b)  $0,3 + 0,2 + 0,1 = 0,6$   
 Não, 60% dos apartamentos à venda custam menos do que 250 000 e apenas 40% dos apartamentos à venda custam, pelo menos, 250 000 euros.

#### Página 145

3. Como a altura das plantas, nesta fase de crescimento se deve situar entre 13 e 14 cm, o histograma mais adequado para obter o número de plantas nessa situação é o histograma II (visto que não há nenhuma planta com 14 cm - ver dados iniciais).  
 O histograma I só permite saber o número de plantas que tem pelo menos 12 cm e menos do que 14 cm.  
 O histograma III só permite saber o número de plantas que tem pelo menos 12,5 cm e menos do que 14,5 cm.
- 4.1. O histograma que poderá representar a distribuição das classificações dos alunos na realização de um teste muito fácil é o III, porque as barras "crescem" à medida que as classificações sobem e o número de alunos com classificações baixas é muito reduzido.
- 4.2. O histograma que poderá representar a distribuição das classificações dos alunos na realização de um teste numa turma em que a maioria dos alunos tem muitas dificuldades é o I porque as primeiras classes têm uma elevada frequência absoluta.
- 4.3. O histograma que poderá representar a distribuição das classificações dos alunos na realização de um teste com grau de dificuldade equilibrado é o II porque a distribuição dos dados no histograma apresenta uma simetria em relação à classe central.
- 4.4. O histograma que poderá representar a distribuição das classificações dos alunos na realização de um teste numa turma muito heterogénea é o IV porque nessa turma se verificam resultados quer muito baixos, quer muito altos.



## Página 146

## 5.1. a) Conjunto de dados A

Média

$$\bar{x} = \frac{2 \times 15 + 4 \times 16 + 3 \times 14 + 2 \times 13 + 2 \times 12 + 11}{14} =$$

$$= \frac{197}{14} \approx 14,1$$

Moda

$M_o = 16$  (valor com maior frequência absoluta)

Mediana

Começamos por ordenar os dados (por ordem crescente)

$$11 \ 12 \ 12 \ 13 \ 13 \ 14 \ \boxed{14 \ 14} \ 15 \ 15 \ 16 \ 16 \ 16 \ 16$$

$$M_e = \frac{14+14}{2} = 14$$

$$\bar{x} \approx 14,1 \quad M_o = 16 \quad M_e = 14$$

- b) Se todas as temperaturas máximas fossem 2 °C mais baixas, então as medidas calculadas na alínea anterior diminuiriam todas 2 °C.

## 5.2. Conjunto de dados B

$$\bar{x} = 7, (3)$$

Vamos determinar a mediana deste conjunto de dados e, para tal, começamos por ordená-los por ordem crescente.

$$1 \ 3 \ 4 \ 5 \ 5 \ 6 \ 6 \ 6 \ 7 \ 7 \ 7 \ 7 \ 8 \ \boxed{8 \ 8} \ 8 \ 8 \ 8$$

$$8 \ 9 \ 9 \ 9 \ 9 \ 9 \ 9 \ 10 \ 10 \ 10$$

$$M_e = \frac{8+8}{2} = 8$$

Assim, a mediana é maior do que a média pois  $\bar{x} = 7,13$  e  $M_e = 8$

## 5.3. Conjunto de dados C

a) Média

$$\bar{x} = \frac{800 + 1880 + 1035 + 920 + 5400 + 1250}{8} +$$

$$+ \frac{1500 + 2100}{8} = \frac{14 \ 885}{8} = 1860,625 \approx 1861$$

Mediana

Ordenados os dados por ordem crescente, tem-se:

$$800 \ 920 \ 1035 \ \boxed{1250 \ 1500} \ 1880 \ 2100 \ 5400$$

$$M_e = \frac{1250+1500}{2} = 1375$$

$$\bar{x} \approx 1861 \text{ € e } M_e = 1375 \text{ €}$$

- b) O outlier deste conjunto de dados é o 5400 que corresponde ao vencimento do diretor deste setor.

Vencimento do diretor: 5400 €

A medida que melhor descreve o centro da distribuição dos dados é a mediana pois não é afetada pelo(s) extremo(s).

- c) Se cada um dos colaboradores deste setor receber um aumento de 5%, então a média salarial também sofrerá um aumento de 5%.

6.1. Média

$$\bar{x} = \frac{10 \times 48 + 45 \times 49 + 52 \times 50 + 39 \times 51 + 9 \times 52 + 5 \times 53}{160}$$

$$= \frac{8007}{160} \approx 50$$

Como  $\bar{x} \approx 50$ , a informação presente na embalagem está concordante com os dados apurados na amostra.

## 6.2.

N.º de bombons	Freq. Absoluta	Freq. Rel. (%)	Freq. Rel. Acum. (%)
48	10	6,25%	6,25%
48	45	28,125%	34,375%
50	52	32,5%	66,875%
51	39	24,375%	91,25%
52	9	5,625%	96,875%
53	5	3,125%	100%

$N = 160$  (par)

$$M_e = \frac{x_{80} + x_{81}}{2} = \frac{50 + 50}{2} = 50$$

## Página 147

- 7.1. Número total de colaboradores da cadeia de lojas: 50 colaboradores.

## 7.2.

Classe	$n_i$	$N_i$	$f_i$ (%)	$F_i$ (%)
[15, 20[	16	16	32%	32%
[20, 25[	36 - 16 = 20	36	40%	72%
[25, 30[	40 - 36 = 4	40	8%	80%
[30, 35[	45 - 40 = 5	45	10%	90%
[35, 40[	46 - 45 = 1	46	2%	92%
[40, 45[	48 - 46 = 2	48	4%	96%
[45, 50[	50 - 48 = 2	50	30%	100%

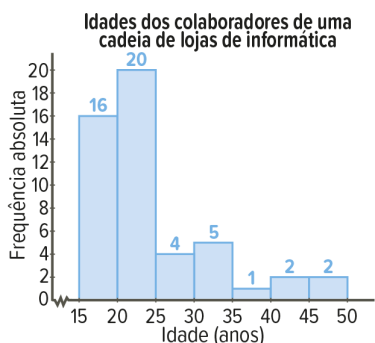


- 7.3. Marcas das classes: 17,5; 22,5; 27,5; 32,5; 37,5; 42,5; 47,5.

$$\bar{x} = \frac{16 \times 17,5 + 20 \times 22,5 + 4 \times 27,5 + 5 \times 32,5}{50} + \frac{1 \times 37,5 + 2 \times 42,5 + 2 \times 47,5}{50} = \frac{1220}{50} = 24,4$$

- 7.4. A mediana pertence à classe  $[20, 25[$ .

7.5.



Por observação da distribuição, verifica-se que a mediana é inferior à média do conjunto de dados.

- 8.1. Como o conjunto é constituído por 6 elementos, a mediana é a média aritmética dos dois valores centrais (depois dos dados estarem ordenados). Como  $M_e = 4,5$ , então a soma desses dois valores será  $4,5 \times 2$ , ou seja, 9. Então, o menor valor possível de  $a$  é 1 ( $a$  é um número inteiro positivo).

8.2.  $\bar{x} = 5 \Leftrightarrow \frac{1+9+a+7+4+5}{6} = 5 \Leftrightarrow$

$$\Leftrightarrow 26 + a = 30 \Leftrightarrow a = 4$$

Ordenando os dados por ordem crescente:

1 4 4 5 7 9

Então,

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

- 8.3. Sendo  $M_o = 5$ , então  $a = 5$  (pois 5 terá de ser o valor com mais frequência absoluta).

$$\bar{x} = \frac{1+9+5+7+4+5}{6} = \frac{31}{6} \approx 5,2$$

Ordenando os dados por ordem crescente:

1 4 5 5 7 9

$$\text{Então, } M_e = \frac{5+5}{2} = 5$$

9.1. a)  $\bar{x} = \frac{1 \times 0 + 3 \times 1 + 8 \times 2 + 4 \times 3 + 2 \times 4 + 1 \times 5 + 1 \times 8}{20} = \frac{52}{20} = 2,6$

$\bar{x} = 2,6$  significa que em média, cada família desse contexto habitacional tem 2,6 crianças.

- b)  $M_o = 2$ , valor com maior frequência absoluta, significa que a maioria das famílias desse complexo habitacional tem dois filhos.

c)  $M_e = \frac{x_{10} + x_{11}}{2} = \frac{2+2}{2} = 2$

$M_e = 2$  significa que, pelo menos, 50% das famílias desse complexo habitacional tem dois ou menos filhos e, no máximo, 50% têm dois ou mais filhos.

9.2. Outlier: 8

- a) A média diminuirá porque o *outlier* é o dado maior do conjunto inicial.  
 b) A moda não será afetada pois o *outlier* é um valor extremo (não é o mais frequente).  
 c) A mediana não será afetada pois só depende dos valores centrais e o *outlier* é um extremo.

- 10.1. A média faz intervir todos os dados de um conjunto e, por isso, é muito sensível à alteração de um valor.  
 10.2. A mediana é igual a um dos dados se a amplitude do conjunto for um número ímpar.  
 10.3. Pelo menos 50% dos dados são inferiores ou iguais à mediana.  
 10.4. Um conjunto de dados pode ter mais do que uma moda.  
 10.5. A média é influenciada pela existência de outliers.  
 10.6. A moda é necessariamente um dos dados do conjunto.

#### Página 148

- 11.1. a) Um rapaz com 4 anos que mede 104 cm está no percentil 75.  
 b) Um rapaz com 15 anos que mede 180 cm está no percentil 90.



- 11.2. a) Estima-se que um rapaz que mede 170 cm e está no percentil 25 tem, aproximadamente, 17 anos.
- b) Estima-se que um rapaz que mede 110 cm e está no percentil 10 tenha, aproximadamente, 6 anos.
- 11.3. Afirmar que relativamente à sua estatura, o Eduardo está no percentil 75 significa que em 100 rapazes saudáveis com a idade do Eduardo, pelo menos, 75 têm estatura menor ou igual à do Eduardo, e, no máximo, 25 têm estatura superior à do Eduardo.
- 11.4. Se, em 100 rapazes saudáveis da idade do Tomás, se estima que, no máximo, 5 sejam mais altos do que ele é porque o Tomás se encontra no percentil 95.

12.1. Média:

$$\bar{x} = \frac{28 \times 25 + 45 \times 35 + 74 \times 45 + 63 \times 55 + 24 \times 65}{234}$$

$$= \frac{10\,630}{234} \approx 45$$

O valor aproximado da média das alturas dos pinheiros que foram medidos é 45 cm.

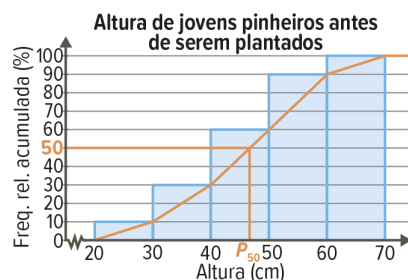
12.2. a)

Altura	N.º de pinheiros	$f_i$ (%)	$F_i$ (%)
[20, 30[	28	12%	12%
[30, 40[	45	19%	31%
[40, 50[	74	32%	63%
[50, 60[	63	27%	90%
[60, 70[	24	10%	92%

A mediana da altura dos pinheiros pertence à classe [40, 50[.

b) O percentil 90 pertence à classe [50, 60[.

12.3.



- 12.4. Construindo o polígono de frequências relativas acumuladas, estima-se que o percentil de ordem 50 seja aproximadamente igual a 46.

No contexto apresentado, significa que, pelo menos 50% dos pinheiros da amostra tem uma altura inferior ou igual a 46 cm e, no máximo, 50% dos pinheiros têm uma altura superior a 46 cm.

- 13.1. a)  $P_{100} = 780$  significa que o preço máximo do telemóvel, nas 10 lojas, é 780 euros.
- b)  $P_{40} = 730$  significa que, em pelo menos, 40% das 10 lojas o preço do telemóvel é inferior ou igual a 730 euros e, no máximo, em 60% das 10 lojas o preço do telemóvel é superior a 730 euros.
- 13.2.  $P_{20} = 718$ , então no máximo, em 80% das 10 lojas, ou seja, em 8 lojas, o preço de telemóveis é superior a 718 euros.
- Opção (D)

Página 149

14.1.  $18,2 = Q_1 = P_{25}$

Como a Rita tem IMC igual a 18,2 e 18,2 corresponde ao 1.º quartil da distribuição, então sabe-se que pelo menos 25% das raparigas de 15 anos tem IMC abaixo ou igual da Rita.

- 14.2. Sabe-se que, no máximo, 10% das raparigas de 15 anos têm IMC maior do que a Mariana, logo, pelo menos 90% das raparigas de 15 anos têm IMC inferior ou igual ao da Mariana. Como  $P_{90} = 25,4$ , conclui-se que o IMC da Mariana é 25,4 kg/m<sup>2</sup>.

- 14.3. Como a Eva tem IMC igual a 25,4 e  $P_{10} = 25,4$ , sabe-se que, no máximo, a percentagem de raparigas de 15 anos com IMC mais alto do que o da Eva é 10%.

- 14.4. Ao IMC igual a 19,8 corresponde ao percentil 50 pois  $M_e = 19,8$ .

15. Opção (C)

- 16.1. Se, pelo menos, 10% dos dados são menores ou iguais a  $P_k$ , então  $k = 10$ .
- 16.2. Se, no máximo, 60% dos dados são superiores a  $P_k$ , então  $k = 100 - 60$ , ou seja,  $k = 40$ .
- 16.3. Se, no máximo, um quarto dos dados (25%) são superiores a  $P_k$ , então  $k = 100 - 25$ , ou seja,  $k = 75$ .



- 17.1. Como  $M_e = P_{50}$ , então a mediana dos pesos dos bebés meninos com 6 meses é 7,9.
- 17.2. a) Se o Francisco tem 1 ano e pesa 9,6 kg, de acordo com a tabela de crescimento da DGS, o Francisco encontra-se no percentil 25. Logo, conclui-se que a percentagem de bebés meninos da mesma idade que pesa menos ou o mesmo que o Francisco é 25%.
- b) De acordo com a tabela de crescimento da DGS, se o Francisco se mantiver no mesmo percentil do peso ( $P_{25}$ ), daqui a 6 meses, ou seja, aos 18 meses, espera-se que ele pese 10,9 kg.
- 17.3. O Manuel tem 18 meses e sabe-se que, no máximo, 5% dos bebés meninos da idade do Manuel pesam mais do que ele, então sabe-se que o Manuel se encontra no percentil 95. Consultando a tabela de crescimento da DGS, conclui-se que o Manuel pesa 14 Kg.
- 17.4. De acordo com a tabela de crescimento da DGS, se o David tem 2 anos e pesa 11,8 Kg, então encontra-se no percentil 25. Assim sendo, sabe-se que, no máximo, 75% de bebés meninos pesam mais do que o David.
- 17.5. De acordo com a tabela de crescimento da DGS, se o Santiago se mantiver no mesmo percentil de peso ( $P_{75}$ , pois  $Q_3 = P_{75}$ ), então daqui a 6 meses, ou seja, aos 24 meses, estima-se que ele pese 13,6 kg.

### Página 150

#### 18.1. Amplitude

$$34 - 18 = 16$$

#### Quartis

$$18 \quad \boxed{19 \quad 20} \quad 20 \quad 20 \quad \boxed{25 \quad 31} \quad 34$$

$$Q_1 = \frac{19+20}{2} = 19,5 \quad Q_3 = \frac{25+31}{2} = 28$$

#### Amplitude interquartil

$$Q_3 - Q_1 = 28 - 19,5 = 8,5$$

#### 18.2. Amplitude

$$30 - 10 = 20$$

#### Quartis

$$10 \quad 10 \quad 10 \quad \boxed{10} \quad 10 \quad 10 \quad 10 \quad 20 \quad 20 \quad 20 \quad \boxed{20} \quad 30 \quad 30 \quad 30$$

$$Q_1 \quad Q_2$$

$$\text{Amplitude interquartil: } Q_3 - Q_1 = 9 - 9 = 0$$

$$18.3. \text{ Amplitude } 10 - 9 = 1$$

$$\text{Quartis } 9 \quad \boxed{9} \quad 9 \quad 9 \quad \boxed{9} \quad 10$$

$$Q_1 \quad Q_2$$

$$\text{Amplitude interquartil } Q_3 - Q_1 = 9 - 9 = 0$$

- 19.1. A mediana das semanadas, de acordo com o diagrama de extremos e quartil, é 20 euros.
- 19.2. Amplitude:  $35 - 5 = 30$   
Amplitude interquartil:  $Q_3 - Q_1 = 25 - 10 = 15$
- 19.3. A maior concentração dos dados relativos às semanadas situa-se entre os 20 e os 25 euros e, também, entre os 5 e os 10 euros.
- 19.4. 25% das crianças recebem entre 20 e 25 euros de semanada, ou seja, 15 crianças.
- 19.5. Recebem mais de 25 euros de semanada, no máximo 15 crianças.
- 19.6. A medida que é afetada é a amplitude e será aumentada (pois o máximo também aumentou).
20. A turma que demonstrou ser mais homogénea no desempenho foi a B, porque é aquela onde a amplitude das classificações é menor.
- Turma A: Amplitude =  $19,2 - 9,7 = 9,5$   
Turma B: Amplitude =  $18,5 - 10,6 = 7,9$
- 21.1.  $\bar{x} = \frac{3+0+2+4+2}{5} = \frac{11}{5} = 2,2$

A média do tempo de atraso desse autocarro foi 2,2 minutos.

$$21.2. (3-2,2)^2 + (0-2,2)^2 + (2-2,2)^2 + (4-2,2)^2 + (2-2,2)^2 = 8,8$$

$$s = \sqrt{\frac{8,8}{5-1}} = \sqrt{2,2} \approx 1,48$$

- 21.3. Se o desvio-padrão aumentou, então o dado 4 foi substituído por um dado cujo valor absoluto do desvio em reação à média é maior. Como  $|4 - 2,2| = 1,8$ , conclui-se que o dado terá de ser o 5.

#### Opção (D)

- 21.4. a) Se duplicarmos todos os dados do enunciado, então a média e o desvio-padrão também serão duplicados. Assim, a média e o desvio-padrão do novo conjunto de dados são  $\bar{x} = 2 \times 2,2 = 4,4$  e  $s = 2 \times 1,48 = 2,96$



- b) Se adicionarmos 1 minuto a todos os dados do enunciado, então a média também será adicionada de 1 minuto e o desvio-padrão não se altera.

Logo,  $\bar{x} = 1 + 2,2 = 3,2$  e  $s \approx 1,48$ .

## Página 151

### 22.1. Mão esquerda

Amplitude:  $0,90 - 0,26 = 0,64$

Média:

$$\bar{x} = \frac{3 \times 0,30 + 2 \times 0,28 + 0,40 + 4 \times 0,38}{15} + \frac{0,35 + 0,39 + 0,26 + 0,45 + 0,90}{15} = \frac{5,73}{15} = 0,382$$

Mão direita

Amplitude:  $0,52 - 0,28 = 0,24$

Média:

$$\bar{x} = \frac{0,32 + 2 \times 0,31 + 0,42 + 0,37 + 0,43 + 0,38 + 0,50}{15} + \frac{0,52 + 0,45 + 0,29 + 0,46 + 0,36 + 0,33 + 0,28}{15} = \frac{5,73}{15} = 0,382$$

### 22.2.

Mão esquerda		Mão direita	
Tempo (s) $x_i$	$x_i - \bar{x}$	Tempo (s) $x_i$	$x_i - \bar{x}$
0,30	-0,082	0,32	-0,062
0,28	-0,102	0,31	-0,072
0,40	0,018	0,42	0,038
0,30	-0,082	0,31	-0,072
0,38	-0,002	0,37	-0,012
0,38	0,002	0,43	0,048
0,35	-0,032	0,38	-0,002
0,39	0,008	0,50	0,118
0,38	-0,002	0,52	0,138
0,38	-0,002	0,45	0,068
0,26	-0,122	0,29	-0,092
0,45	0,068	0,46	0,078
0,28	-0,102	0,36	-0,022
0,30	-0,082	0,33	-0,052
0,90	0,518	0,28	-0,102

- a) O dado resultante da mão direita cujo valor absoluto do desvio em relação à média é o maior é 0,52.

- b) O dado resultante da mão esquerda cujo valor absoluto do desvio em relação à média é o menor é 0,38.

22.3.  $s_d$  : desvio-padrão relativo à mão direita

$s_e$  : desvio-padrão relativo à mão esquerda

Recorrendo à calculadora gráfica, conclui-se que

$$s_e \approx 0,154 \text{ e } s_d \approx 0,077$$

22.4.  $\bar{x}_e = \bar{x}_d$  e  $s_e > s_d$

Os dados relativos à mão esquerda são mais dispersos.

22.5. Se o outlier, 0,90, fosse corrigido para 0,29, o valor do desvio-padrão não se mantém (diminui) porque  $(0,29 - 0,38)^2 < (0,9 - 0,382)^2$ .

23.1. Praia A: Inserindo os dados numa lista da calculadora, conclui-se que:

Média:  $\bar{x} \approx 72$

Mediana:  $M_e \approx 72$

Amplitude:  $94 - 52 = 42$

Amplitude interquartil:  $Q_3 - Q_1 = 84 - 60 = 24$

23.2. Praia A: Amplitude: 42

Praia B: Amplitude:  $117 - 10 = 107$

O número de bolas de Berlim vendidas apresentou maior variabilidade na praia B porque a amplitude do conjunto de dados é muito maior. Verifica-se, ainda, que pelo menos 25% dos dados observados estão entre 10 e 59 e pelo menos 25% dos dados estão entre 91 e 117.

## Página 152

1. A variável quantitativa discreta é a V2.

Opção (C)

2.1.  $9 + a + 10 + 7 = 50 \Leftrightarrow a = 24$

$$b = \frac{9 + 24}{50} \times 100 = 66$$

Opção (D)

2.2. Como a mediana corresponde a 50% dos dados, conclui-se que a mediana pertence à classe  $[2, 4[$ .

Opção (B)



3.1.  $\bar{x} = 41,6 \Leftrightarrow$   
 $\Leftrightarrow \frac{7 + 13 + 18 + 21 + 29 + a + 64 + 71 + 75 + 83}{10} = 41,6 \Leftrightarrow$

$$\Leftrightarrow 381 + a = 416 \Leftrightarrow a = 35$$

Opção (B)

3.2.  $P_{70} = 65,9$

$$P_{70} - P_{30} = ?$$

Sabe-se que  $P_{70} - P_{30} \leq P_{70} - P_{25}$  e

$$P_{70} - P_{25} = 65,9 - 18 = 47,9$$

Assim conclui-se que  $P_{70} - P_{30} \leq 47,9$

Assim sendo, o único valor que pode representar  $P_{70} - P_{30}$  é 46.

Opção (A)

4.1. Quando o desvio-padrão é nulo, não há variabilidade entre os dados, logo estes são iguais. Como  $s_r = 0$ , então é no conjunto I que não há variabilidade entre os dados.

Opção (A)

4.2. Quanto maior é o desvio-padrão, maior é a dispersão dos dados. Assim sendo, a maior dispersão dos dados verifica-se no conjunto III (pois é o que tem o desvio-padrão maior).

Opção (C)

### Página 153

1.1. Utilizando a calculadora e inserindo os valores dos salários numa lista e os correspondentes valores da frequência absoluta na outra lista (2, 2, 7, 6, 1), obtém-se:

$$\text{Média: } \bar{x} = 1257$$

$$\text{Mediana: } M_e = 1100$$

Amplitude interquartil:

$$Q_3 - Q_1 = 1200 - 950 = 250$$

1.2. Ao ser considerado o dado 4800, que foi um erro de transcrição, a medida que melhor representa o centro da distribuição dos salários da empresa A é a mediana porque esta medida não é sensível aos valores extremos (depende apenas dos valores centrais).

1.3. Como na empresa B haverá um aumento salarial de 10%, a média e o desvio-padrão também sofrerão um aumento de 10%.

Assim, temos:

$$\bar{x}_{\text{final}} = 1,1 \times \bar{x} = 1,1 \times 1140 = 1254$$

$$s_{\text{final}} = 1,1 \times s = 1,1 \times 267,7 = 294,47$$

2. Consultando a tabela de crescimento do perímetro cefálico, verifica-se que um bebé com 1 ano de idade que tenha perímetro cefálico entre 45,5 cm e 48,5 cm encontra-se entre o percentil 25 e o percentil 95.

$$0,25 \times 100 = 25$$

$$0,95 \times 100 = 95$$

$$95 - 25 = 70$$

Então, numa amostra de 100 bebés meninos com 1 ano de idade, é de esperar que 70 tenham perímetro cefálico entre 45,5 cm e 48,5 cm.

3.1. Na calculadora, começamos por inserir os valores da idade numa lista e o correspondente número de alunos na outra lista. Seguidamente, determinam-se os valores da média e do desvio-padrão das idades dos alunos, a saber:

$$\bar{x} = 13,40 \text{ e } s \approx 0,83$$

3.2. Se forem consideradas as idades desses alunos no dia 1 de setembro de 2024, ou seja, daqui a 1 ano, então a medida é adicionada de 1 ano e o desvio-padrão mantém-se (pois a todos os dados é adicionado 1 unidade).

$$\bar{x} = 13,4 + 1 = 14,4 \text{ e } s \approx 0,83$$

### Página 154

1.1. A população é o conjunto dos alunos da turma.

1.2. Unidades estatísticas – elementos de população.

Exemplos de dois exemplos de unidades estatísticas: Ana e Vasco.

1.3. Há duas variáveis a estudo: o número de horas diárias que cada aluno está *online* e o número de horas que cada aluno passa a dormir.



2. Como o Rafael marcou o ponto de coordenadas  $(4, 7)$ , então o Rafael está 4 horas diárias *online* e passa 7 horas diárias a dormir.
3. Se, diariamente, a Sofia está *online* 7 horas e dorme 7 horas, então vai marcar no referencial o ponto de coordenadas  $(7, 6)$ .
4. A informação disponibilizada sugere que, de um modo geral, os alunos da turma não estão alinhados com o título do artigo pois o número de horas de como é, na maioria das situações conhecidas, superior ao número diário de horas *online*.

#### Página 155

- 45.1. a)  $(5, 3) \rightarrow$  um automobilista que efetuou uma viagem de 5 horas, fez 3 paragens.  
b)  $(3, 2) \rightarrow$  um automobilista que efetuou uma viagem de 3 horas, fez 2 paragens.
- 45.2. O par  $(1, 2)$  não faz parte do conjunto de dados e o par  $(2, 1)$  faz parte.
- 45.3. A dimensão da amostra é 6.

#### Página 156

- 46.1. Variável independente: número de acidentes  
Variável dependente: número de vítimas mortais
- 46.2. Da informação registada na tabela, sabe-se que os pares que formam o conjunto de dados que permitem construir a nuvem de pontos apresentada são:  
 $(2668, 42)$ ,  $(2315, 28)$ ,  $(2682, 31)$ ,  
 $(2847, 53)$ ,  $(3015, 34)$ ,  $(3050, 45)$

#### Página 157

- 47.1. Variável exploratória: número de alunos na turma  
Variável resposta: número de casos de indisciplina
- 47.2. Pela análise do diagrama de dispersão, pode-se concluir que o número de casos de indisciplina não está relacionado com o número de alunos na turma.

#### Página 158

- 48.1. À medida que a variável independente aumenta, a tendência é que a variável dependente também aumente.  
Assim, conclui-se que o coeficiente de correlação linear é positivo.  
Como os pontos do diagrama de dispersão não estão sobre uma reta,  $r \neq 1$ .  
Então, ter-se-á  $r = 0,97$ .  
Opção (C)
- 48.2. À medida que a variável independente aumenta, a variável dependente diminui, logo  $r < 0$ .  
Como os pontos do diagrama de dispersão estão sobre uma reta, conclui-se que  $r = -1$ .  
Opção (B)

#### Página 159

- 49.1. A correlação linear existente entre as duas variáveis é positiva pois, à medida que o tempo de experiência aumenta, a tendência é que a temperatura da substância também aumente.
- 49.2. Recorrendo à calculadora gráfica, determina-se o coeficiente de correlação linear:  $r \approx 0,991$ .

#### Página 160

- 27.1. As variáveis que devem ser estudadas são: número de jogadores estrangeiros no plantel e média das idades dos jogadores do plantel.  
Variável independente: número de jogadores estrangeiros no plantel.  
Variável dependente: média das idades dos jogadores do plantel.
- 27.2. a) Sim, a dimensão da amostra é 9 (corresponde ao número de pontos do diagrama de dispersão).  
b) Pela análise do diagrama de dispersão, conclui-se que o número de jogadores estrangeiros não influencia a média das idades dos jogadores do plantel pois não há uma tendência para que os pontos estejam próximos de uma reta.



- c) Como  $B(19; 23,3)$  corresponde ao F.C.

Famalicão, conclui-se que na equipa do F.C.

Famalicão existem 19 jogadores

estrangeiros.

A média das idades dos jogadores da equipa é 23,3 anos.

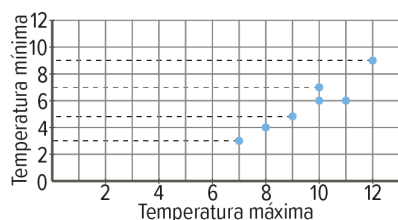
- d) No conjunto das equipas que fazem parte da amostra, o máximo de jogadores estrangeiros é 21.

- 28.1. a) Variável A: Temperatura mínima diária (nessa localidade).

Variável B: Temperatura máxima diária (nessa localidade).

- b) Recorrendo à calculadora gráfica, determina-se o coeficiente de correlação linear:  $r \approx 0,94$ .

28.2.



Recorrendo à calculadora gráfica, conclui-se que  $r \approx 0,94$ .

Página 161

### 1. Casais com filho

Altura (cm)		Altura média dos pais (cm)	Altura do filho (cm)
Pai	Mãe		
177	155	166	174
188	170	179	184
168	150	159	169
177	165	171	175
190	166	178	185
179	157	168	174
182	166	174	183
169	159	164	172
172	152	162	168
185	175	180	189

### Casais com filha

Altura (cm)		Altura média dos pais (cm)	Altura da filha (cm)
Pai	Mãe		
179	153	166	160
176	170	173	166
191	169	180	175
172	160	166	158
178	162	170	164
179	155	167	163
182	166	174	169
174	168	171	163
176	158	167	162
187	179	183	178

- 2.1. As variáveis em estudo são a altura dos filhos e a média das alturas dos pais.

Variável independente: média das alturas dos pais.

Variável dependente: altura dos filhos.

- 2.2. Observando os diagramas de dispersão apresentados, o que parece ter maior coeficiente de correlação linear é o diagrama de dispersão referente ao Grupo 2.

Utilizando a calculadora gráfica, conclui-se que  $r_1 \approx 0,966$  e  $r_2 \approx 0,979$ .

Então, confirma-se que o diagrama de dispersão referente ao Grupo 2 apresenta maior coeficiente de correlação linear.

3.  $6,5 - 2 = 4,5$        $6,5 + 2 = 8,5$

$[4,5; 8,5]$

$$174 - 166 = 8 \quad \checkmark \qquad 184 - 179 = 5 \quad \checkmark$$

$$169 - 159 = 10 \quad \checkmark \qquad 175 - 170 = 4 \quad \times$$

$$185 - 178 = 7 \quad \checkmark \qquad 174 - 168 = 6 \quad \checkmark$$

$$183 - 174 = 9 \quad \times \qquad 172 - 164 = 8 \quad \checkmark$$

$$168 - 162 = 6 \quad \checkmark \qquad 189 - 180 = 9 \quad \times$$

$$\frac{7}{10} \times 100\% = 70\%$$

No grupo 1, a percentagem das alturas dos filhos que satisfaz a regra com uma margem de erro de 2 cm é 70%.



## Página 162

50.1. A amostra é constituída pelos seguintes pares ordenados:

$(1, 2)$ ,  $(2, 2)$ ,  $(2, 3)$ ,  $(3, 3)$  e  $(4, 5)$

$$\bar{x} = \frac{1+2+2+3+4}{5} = \frac{12}{5} = 2,4$$

$$\bar{y} = \frac{2+2+3+3+5}{5} = \frac{15}{5} = 3$$

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

$$\sqrt{(2-3)^2 + (2-3)^2 + (3-3)^2 + (3-3)^2 + (5-3)^2} = \sqrt{6}$$

$$r = \frac{(1-2,4)(2-3) + (2-2,4)(2-3) + (2-2,4)(3-3) + (3-2,4)(3-3) + (4-2,4)(5-3)}{\sqrt{5,2} \times \sqrt{6}}$$

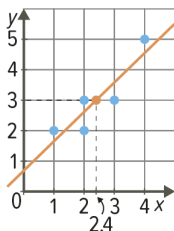
$$+ \frac{(3-2,4)(3-3) + (4-2,4)(5-3)}{\sqrt{5,2} \times \sqrt{6}} = \frac{5}{\sqrt{5,2} \times \sqrt{6}}$$

$$= \frac{5}{\sqrt{31,2}} \approx 0,895$$

50.2. Centro de gravidade da distribuição:

$(\bar{x}, \bar{y}) = (2,4; 3)$

50.3.



Nota: a reta de regressão passa no centro de gravidade da distribuição.

51.1. Centro de gravidade:  $(\bar{x}, \bar{y})$

$$\bar{x} = \frac{3+5+8+10+15}{5} = \frac{41}{5} = 8,2$$

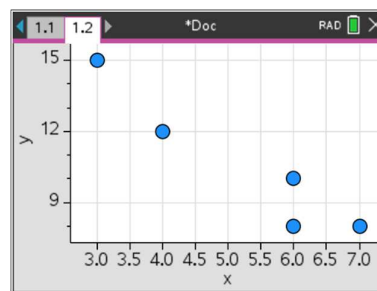
$$\bar{y} = \frac{20+32+45+48+55}{5} = \frac{200}{5} = 40$$

As coordenadas do centro de gravidade da distribuição são  $(8,2; 40)$ .

51.2. À medida que a idade aumenta, há tendência para o peso aumentar. Logo, podemos concluir que entre variáveis há uma correlação positiva. Assim sendo, o sinal do declive da reta de regressão é positivo.

## Página 163

52.1.



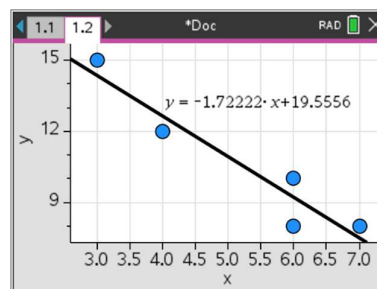
52.2. Coeficiente de correlação linear:  $r \approx -0,95$ .

x	y	C	D
2	4	12. RegEqn	m*x+b
3	6	10. m	-1.72222
4	6	8. b	19.5556
5	7	8. r²	0.910038
6		r	-0.9539...
D6			=-0.95395905508983

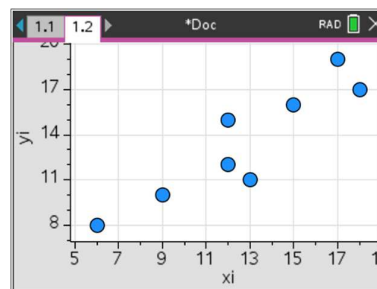
52.3. Equação da reta de regressão:  $y = ax + b$

$a \approx -1,722$  e  $b \approx 19,556$

$y = -1,722x + 19,556$



53.1.

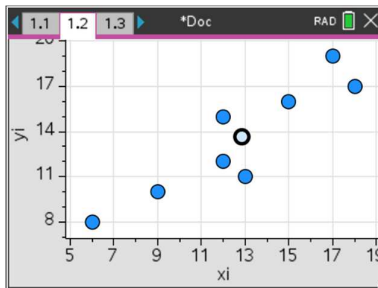




### 53.2. Centro de gravidade: $(\bar{x}, \bar{y})$

Recorrendo à calculadora gráfica, tem-se:

$$\bar{x} = 12,75 \text{ e } \bar{y} = 13,5$$



Logo, as coordenadas do centro de gravidade são  $(12,75; 13,5)$ .

### 53.3. Retas de regressão

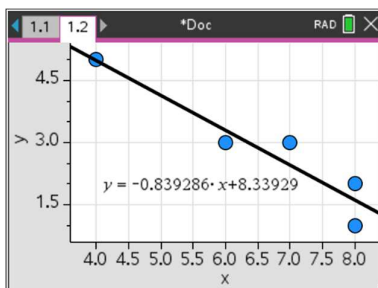
Equação da reta de regressão:  $y = ax + b$

$$a \approx 0,870 \text{ e } b \approx 2,408$$

Então,  $y = 0,870x + 2,408$

## Página 164

### 54.1.



### 54.2. Equação da reta de regressão: $y = ax + b$

$$a \approx -0,839 \text{ e } b \approx 8,339$$

$$y = -0,839x + 8,339$$

a) Se  $x = 5$ , então:

$$y = -0,839 \times 5 + 8,339 = 4,144$$

O número aproximado de defeitos foi 4.

b) 7 horas e 30 minutos  $\rightarrow$  7,5 horas

Se  $x = 7,5$ , então:

$$y = -0,839 \times 7,5 + 8,339 = 2,0465$$

O número aproximado de defeitos foi 2.

## Página 165

### 1.1. Conjunto de dados A

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

$$\bar{y} = \frac{4 + 1 + 0 + 1 + 4}{5} = \frac{10}{5} = 2$$

Centro de gravidade:  $(0, 2)$

### Conjunto de dados B

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

$$\bar{y} = \frac{0 + 1 + 2 + 3 + 4}{5} = \frac{10}{5} = 2$$

Centro de gravidade:  $(0, 2)$

### Conjunto de dados C

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

$$\bar{y} = \frac{3,5 + 3 + 2 + 1 + 0,5}{5} = \frac{10}{5} = 2$$

Centro de gravidade:  $(0, 2)$

Os três conjuntos de dados (A, B e C) têm o mesmo centro de gravidade  $((0, 2))$ .

### 1.2. Conjunto de dados A: $y = x^2$

Conjunto de dados B:  $y = x + 2$

### 1.3. Recorrendo à calculadora gráfica, conclui-se que $r_A = 0$ e $r_B = 1$ .

$r_A = 0 \rightarrow$  os dados do conjunto A não estão relacionados através de uma regressão linear.

$r_B = 1 \rightarrow$  os dados do conjunto B estão contidos numa reta de declive positivo.

### 1.4. Sim porque em cada um dos conjuntos, A e B, há uma relação entre as variáveis $x$ e $y$ .

A: Se  $x = 1,5$ , então  $y = 1,5^2 = 2,25$

B: Se  $x = 1,5$ , então  $y = 1,5 + 2 = 3,5$

### 1.5. Equação da reta de regressão: $y = ax + b$

Recorrendo à calculadora gráfica, obtém-se:

$$a = -0,8 \text{ e } b = 2$$

Então,  $y = -0,8x + 2$

Se  $x = 1,5$ , então:  $y = -0,8 \times 1,5 + 2 \Leftrightarrow y = 0,8$

Se  $x = 1,5$ , então o valor estimado para  $y$  é 0,8.



1.6. 1.ª Afirmação: O coeficiente de correlação do conjunto de dados A é nulo e, no entanto, as variáveis estão relacionadas ( $y = x^2$ ), só que não é por uma relação linear.

2.ª Afirmação: A disposição dos pontos na nuvem permite saber se as variáveis estão em relação linear.

3.ª Afirmação: Numa relação funcional, conhecido o valor de uma das variáveis, é possível conhecer o valor exato da outra. A correlação entre variáveis apenas permite fazer estimativas.

### Página 166

- Por observação dos diagramas de dispersão, o conjunto de dados que parece apresentar maior coeficiente de correlação linear é o A.
- Recorrendo à calculadora gráfica, completa-se a tabela:

	Conjunto A	Conjunto B
Coeficiente de correlação linear $r$	-0,245	0,772
Reta de regressão $y = ax + b$	$y = -0,074x + 2,593$	$y = -0,448x + 1,228$

Conclui-se que, afinal, o conjunto B apresenta maior coeficiente de correlação linear.

3.

A	x	1	2	3	4	5
	y	1,5	2	2,5	3,2	3,5
B	x	5	5,4	6	6,5	7
	y	3	4,5	5	3,2	4

Recorrendo à calculadora gráfica, tem-se:

	Conjunto A'	Conjunto B'
Coeficiente de correlação linear $r$	0,995	0,107
Reta de regressão $y = ax + b$	$y = -0,52x + 0,98$	$y = -0,113x + 3,266$

- Ao substituir o conjunto A pelo conjunto A' tem interesse calcular o coeficiente de correlação linear e determinar a equação da reta de regressão porque passa a haver uma correlação mais forte. Ao substituir o conjunto B pelo conjunto B' não tem interesse calcular o coeficiente de correlação linear e determinar a equação da reta de regressão porque a correlação passa a ser, praticamente, inexistente.

### Página 167

#### 1.1. Estratégia A

Considerando o conjunto dos 13 apartamentos e recorrendo à calculadora gráfica, conclui-se que a equação da reta de regressão é

$$y = ax + b, \text{ sendo } a \approx 3,996 \text{ e } b \approx -217,1$$

$$y = 3,996x - 217,1$$

Se  $x = 118$ , então:

$$y = 3,996 \times 118 - 217,1 = 254,428$$

Assim, conclui-se que uma estimativa do preço de um apartamento com 118 m<sup>2</sup> de área é 254 milhares de euros.

#### 1.2. Estratégia B → Recorrendo à calculadora gráfica.

Conjunto dos apartamentos usados

Equação da reta de regressão:  $y = ax + b$

$$a \approx 0,761 \text{ e } b \approx 63,226, \text{ logo}$$

$$y = 0,761x + 63,226$$

Se  $x = 118$ , então:

$$y = 0,761 \times 118 + 63,226 = 153,024$$

Estima-se que o preço de um apartamento usado com 118 m<sup>2</sup> seja 153 milhares de euros.

Conjunto dos apartamentos novos

Equação da reta de regressão:  $y = ax + b$

$$a \approx 4,035 \text{ e } b \approx -194,3, \text{ logo}$$

$$y = 4,035x - 194,3$$

Se  $x = 118$ , então:

$$y = 4,035 \times 118 - 194,3 \Leftrightarrow y = 281,83$$

Estima-se que o preço de um apartamento novo com 118 m<sup>2</sup> seja 282 milhares de euros.

Conclusão

Se o apartamento for usado: 153 milhares de euros.

Se o apartamento for novo: 282 milhares de euros.

- A resposta mais adequada é a que resulta da estratégia B. Nesta estratégia são tratados dados apenas de apartamentos do mesmo tipo (usados ou novos).



### Página 168

55.1. A exceção ocorreu no ano de 2022 pois a prova deveria ter-se realizado no ano de 2021.

55.2. Pontuação média em 2018: 492

Pontuação média em 2022: 472

Decréscimo:  $|472 - 492| = |-20| = 20$

$$\frac{20}{492} \times 100 \approx 4$$

A percentagem do decréscimo ocorrido na pontuação média de 2018 para 2022 foi, aproximadamente, 4%.

55.3. Número alunos participantes: 6793

Número rapazes: 3705

Número raparigas:  $6793 - 3705 = 3088$

$x$ : pontuação média a Matemática, obtida pelas raparigas em 2022.

$x + 11$ : pontuação média a Matemática, obtida pelos rapazes em 2022

$$3705(x + 11) + 3088x = 6793 \times \frac{472}{\text{Pontuação média, a Matemática, obtida pelos alunos portugueses em 2022}} \Leftrightarrow$$

$$\Leftrightarrow 3705x + 40\,755 + 3088x = 3\,206\,296 \Leftrightarrow$$

$$\Leftrightarrow 6793x = 3\,165\,541 \Leftrightarrow x = \frac{3\,165\,541}{6793}$$

$$\Leftrightarrow x \approx 466$$

Raparigas: 466

Rapazes: 477

### Página 169

1. Ao longo dos meses a que se refere o estudo, o número de desempregados teve maior variabilidade nas mulheres.

2. Por observação do gráfico, constata-se que no mês de abril o número de mulheres desempregadas diminuiu em relação ao mês anterior e o número de homens desempregados aumentou.

3. Verifica-se que a variável  $h$  é a que tem menos variabilidade. Assim sendo, o conjunto de dados associado à variável  $h$  será o que tem menos desvio-padrão.

Conclui-se que o valor correspondente ao desvio-padrão da variável  $h$  é 8,63.

4.  $(t_i, h_i)$ , coeficiente de correlação linear:  $r_1$

$(t_i, m_i)$ , coeficiente de correlação linear:  $r_2$

$$r_2 > r_1$$

As representações gráficas relativas ao total e às mulheres evidencia comportamentos semelhantes, o que não acontece de forma tão acentuada entre as representações gráficas do total e dos homens.

### Página 170

1.1. Amplitude =  $1,675 - 1,616 = 0,059$  €/L

1.2. Amplitude nos primeiros 5 dias:

$$1,665 = 0,010 \text{ €/L}$$

Amplitude nos últimos cinco dias:

$$1,641 - 1,622 = 0,019 \text{ €/L}$$

A amplitude dos dados foi maior nos últimos cinco dias.

1.3. A variação máxima, em valor absoluto, que houve em dois dias consecutivos foi 0,019 €/L e ocorreu entre os dias 24 e 25 de dezembro de 2023.

2.1. A dispersão dos dados é maior no gasóleo porque a amplitude do conjunto de dados referentes ao gasóleo é maior do que a amplitude dos dados referentes à gasolina 95.

2.2. A afirmação é falsa. Por observação do diagrama de extremos e quartis, conclui-se que a média é maior do que a mediana ( $\bar{x} > M_e$ ), uma vez que há uma maior concentração de dados à esquerda da mediana.

### Página 171

1. A esperança de vida à nascença, na Região Norte de Portugal, aumentou entre 2011 e 2019, diminuiu entre 2019 e 2020 e voltou a aumentar entre 2020 e 2021.

A diminuição verificada em 2020 pode estar relacionada com a pandemia (COVID-19).

A diferença entre a esperança de vida à nascença para rapaz e para raparigas manteve-se sensivelmente igual.



2.  $\bar{x} = 80,9$  (sem distinção raparigas e rapazes)

Designemos por  $\bar{x}_F$  a média do conjunto de dados relativos à esperança de vida para raparigas e  $\bar{x}_M$  a média do conjunto de dados relativos à esperança de vida para rapazes.

$$\bar{x}_F = 80,9 + 3 = 83,9 \text{ e } \bar{x}_M = 80,9 - 3 = 77,9$$

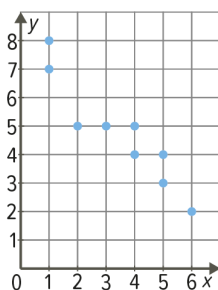
Observando o gráfico, verifica-se que os valores da esperança de vida relativos às raparigas são, aproximadamente, 3 anos superiores à esperança de vida à nascença (sem distinção de raparigas e rapazes) e os relativos aos rapazes são, aproximadamente, 3 anos inferiores.

3. Não haverá grandes diferenças entre os desvios-padrão de cada um dos três conjuntos de dados porque, por uma das propriedades do desvio-padrão, sabe-se que se adicionarmos o mesmo número real, não nulo, a cada um dos elementos de uma amostra, o desvio-padrão não se altera.

O conjunto de dados relativo às raparigas é o que aparenta ter um menor desvio-padrão porque é o que regista menor variabilidade dos dados.

## Página 172

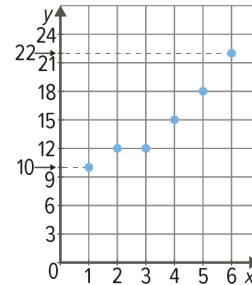
### 1.1.



Existe correlação linear entre as duas variáveis (negativa).

O coeficiente de correlação linear tem sinal negativo.

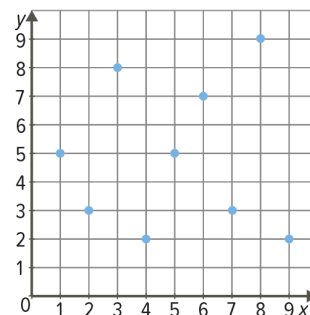
### 1.2.



Existe correlação linear entre as duas variáveis (positiva).

O coeficiente de correlação linear tem sinal positivo.

### 1.3.



Não existe correlação linear entre as duas variáveis.

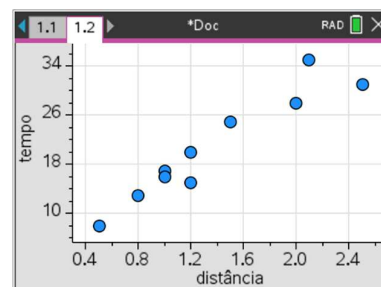
- 2.1. Variável exploratória: distância casa-escola (em quilómetros)

Variável resposta: tempo gasto no percurso casa-escola (em minutos)

- 2.2. Pode-se prever que haja relação entre as duas variáveis no conjunto B porque é construído por uma amostra de alunos que se deslocam a pé para a escola.

- 2.3. Não há evidência de existir correlação (por exemplo, não se está a considerar o meio de transporte).

- 2.4. a)



Existe correlação linear entre as duas variáveis e é positiva.



- b) Coeficiente de correlação linear calculado recorrendo à calculadora gráfica.

$$r \approx 0,95$$

- c) Centro de gravidade da distribuição:  $(\bar{x}, \bar{y})$

Recorrendo à calculadora gráfica, tem-se:

$$\bar{x} = 1,38 \text{ e } \bar{y} = 20,8$$

$$(\bar{x}, \bar{y}) = (1,38; 20,8)$$

- d) Reta de regressão:  $y = ax + b$

$$a \approx 12,89 \text{ e } b \approx 3,01$$

- e) Se  $x = 1,8$ , então:

$$y = 12,89 \times 1,8 + 3,01 = 26,212$$

Um aluno cuja casa fica a 1,8 km da escola e vá a pé para a escola demorará, aproximadamente, 26 minutos no percurso casa-escola.

- f) Se  $y = 10$ , então:

$$10 = 12,89x + 3,01 \Leftrightarrow x = \frac{10 - 3,01}{12,89}$$

$$\Leftrightarrow x \approx 0,542$$

Se um aluno se desloca para a escola a pé e demora 10 minutos, então estima-se que a sua casa fique a, aproximadamente, 0,54 km da escola.

3. Se o coeficiente de correlação linear é  $-1$  então os pontos situam-se todos sobre uma reta.

$$(2; 7,5) \text{ e } (3; 6,25) \quad y = ax + b$$

$$a = \frac{6,25 - 7,5}{3 - 2} = -1,25$$

$$y = -1,25x + b \quad (2; 7,5)$$

$$7,5 = -1,25 \times 2 + b \Leftrightarrow 7,5 = -2,5 + b \Leftrightarrow 10 = b$$

$$y = -1,25x + 10$$

$$(4, y) \quad y = -1,25 \times 4 + 10 = 5$$

$$(1, y) \quad y = -1,25 \times 1 + 10 = 8,75$$

$$(x; 2,5) \quad 2,5 = -1,25x + 10 \Leftrightarrow$$

$$\Leftrightarrow 1,25x = 7,5 \Leftrightarrow x = 6$$

<b>x</b>	4	2	6	1	3
<b>y</b>	5	7,5	2,5	8,75	6,25

## Página 173

- 4.1. Por observação da nuvem de pontos, conclui-se que existe correlação linear positiva entre as duas variáveis.

- 4.2. Centro de gravidade:  $(\bar{x}, \bar{y})$

Recorrendo à calculadora gráfica, sabe-se que:

$$\bar{x} = 163 \text{ e } \bar{y} = 207,9$$

Assim, as coordenadas do centro de gravidade da distribuição são  $(163; 207,9)$ .

- 4.3. a)  $r$ : coeficiente da correlação linear

Recorrendo à calculadora gráfica, conclui-se que  $r \approx 0,985$ .

- b) Equação da reta de regressão:  $y = ax + b$

Recorrendo à calculadora gráfica, tem-se:

$$a \approx 1,10 \text{ e } b \approx 23,62$$

$$\text{Então, } y = 1,10x + 23,62$$

- c) Se  $x = 120$ , então:

$$y = 1,10 \times 120 + 23,62 \Leftrightarrow y = 155,62$$

Estima-se que o preço de um apartamento com  $120 \text{ m}^2$  de área seja, aproximadamente, 155,62 milhares de euros.

- d) 200 000 euros = 200 milhares de euros

Se  $y = 200$ , então:

$$200 = 1,10x + 23,62 \Leftrightarrow \frac{200 - 23,62}{1,10} = x$$

$$x \approx 160$$

Estima-se que a área de um apartamento cujo preço seja 200 000 euros seja, aproximadamente,  $160 \text{ m}^2$ .

- 5.1. A nuvem de pontos que corresponde à distribuição poderá ser a que está representada na opção **(D)**.

- 5.2. **(A)**: rejeitaria porque, se traçarmos a reta de regressão que se ajusta à nuvem de pontos, verificamos que ela tem ordenada na origem positiva (e na distribuição dada, tem-se que  $b = -8,42$ ).

- (B)**: rejeitaria porque, neste caso, a correlação linear não é positiva forte (e na distribuição dada a correlação é positiva forte dado que  $a = 0,93$ ).



(C): rejeitaria porque, neste caso, a correlação é negativa.

### Página 174

6.1. A correlação linear entre as variáveis é positiva, mas não é muito forte. Assim, o possível coeficiente de correlação linear é 0,791.

Opção (A)

6.2. Equação da reta de regressão:

$$y = 2,621x - 3,484$$

Se  $x = 6,3$ , então:

$$y = 2,621 \times 6,3 - 3,484 \approx 13$$

Assim, estima-se que o número de premiados de um país cujo consumo de chocolate fosse 6,3 kg por ano seja 13.

7.1.  $x$ : número de horas de sono

$y$ : classificação obtida no teste de Matemática

A distribuição apresentada no diagrama de dispersão pode ser representada na tabela seguinte:

$x$	5	6	6	6	7	8	9	10
$y$	15	15	16	17	15	18	19	20

Centro de gravidade:  $(\bar{x}, \bar{y})$

Recorrendo à calculadora gráfica, tem-se:

$$\bar{x} = 7,125 \text{ e } \bar{y} = 16,875$$

As coordenadas do centro de gravidade são  $(7,125; 16,875)$ .

7.2. a)  $r$ : coeficiente de correlação linear.

Recorrendo à calculadora gráfica, obtém-se  $r \approx 0,89$ .

b) Equação da reta de regressão:  $y = ax + b$

Recorrendo à calculadora gráfica, obtém-se  $a \approx 1,01$  e  $b \approx 9,66$ .

$$\text{Então } y = 1,01x + 9,66$$

7.3. *Outlier*:  $(10, 9)$

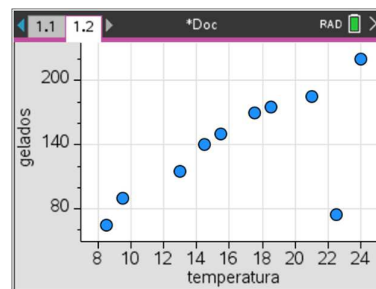
O ponto de coordenadas  $(10, 9)$  foi considerado um *outlier* porque sobressai dos restantes. Neste caso, teríamos um aluno que dormiu 10 h e só obteve 9 valores no teste de Matemática.

8. Espera-se que exista correlação negativa no par de variáveis representado na opção (C) pois à medida que a velocidade de um automóvel aumenta, a tendência é que diminua o tempo gasto num percurso fixo.

Opção (C)

### Página 175

9.1.



As coordenadas do ponto que pode ser considerado um outlier são  $(22,5; 75)$ .

9.2.

	Coeficiente de correlação ( $r$ )	Equação da reta de regressão $y = ax + b$
<i>Outlier</i> incluído	0,637	$y = 6,214x + 36,287$
<i>Outlier</i> excluído	0,988	$y = 9,496x - 4,263$

Excluindo o *outlier*, o valor do coeficiente de correlação aumentou significativamente (aproximou-se de 1).

Relativamente à reta de regressão, aumentou o valor de  $a$  e diminuiu o valor de  $b$ .

Quando se exclui o *outlier*, ao fazer previsões obtemos valores mais ajustados aos dados que são considerados.

10.1. A temperatura que apresentou maior

variabilidade foi a máxima, visto que o intervalo de variação é maior.

10.2. O dia da semana em que a temperatura máxima diminuiu e a mínima aumentou foi a 5.ª feira.

10.3. a) A temperatura mínima foi a que apresentou menor variabilidade, logo o desvio-padrão relativo à variável  $t_{\min}$  é mais pequeno do que os desvios-padrão relativos às outras duas situações.

$$\text{Assim, } s_{\min} = 0,447$$



- b) O conjunto que tem o coeficiente de correlação maior é o  $(t_{\text{média}}, t_{\text{máx.}})$ .
- As representações gráficas relativas à temperatura média e à temperatura máxima são semelhantes, o que não acontece de forma tão acentuada entre as representações gráficas relativas à temperatura média e à temperatura mínima.

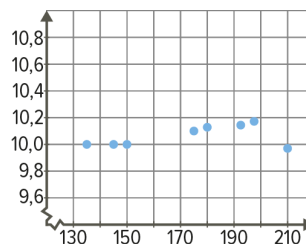
### Página 176

1. Variável exploratória: número de anos do automóvel  
Variável resposta: valor comercial do automóvel  
É esperado que exista uma correlação negativa entre as variáveis porque o aumento do número de anos do automóvel, a tendência é que diminua o seu valor comercial.  
Opção (B)
2. I: correlação linear positiva (muito forte)  
II: correlação linear nula  
III: correlação linear negativa (forte)  
IV: correlação linear positiva  
Opção (B)
3. Como  $r < 0$ , a correlação é negativa.  
Equação da reta de regressão:  $y = ax + b$   
 $a \approx -0,03348$  e  $b \approx 300,79298$   
 $y = -0,03348x + 300,79298$   
Se  $x = 1500$ , então:  
 $y = -0,03348 \times 1500 + 300,79298 \approx 250,6$   
Opção (B)
4. Analisando o diagrama de dispersão, conclui-se que existe uma correlação linear forte entre as duas variáveis. Assim sendo, sabe-se que o coeficiente de correlação é positivo (e próximo de 1) e que o declive da reta de regressão também é positivo.  
Opção (D)

### Página 177

- 1.1. Variável exploratória: tempo de reação (em milésimos de segundo)  
Variável resposta: tempo final (em segundos)

1.2.

1.3. Centro de gravidade:  $(\bar{x}, \bar{y})$ 

Recorrendo à calculadora gráfica, obtemos:  
 $\bar{x} = 173,375$  e  $\bar{y} = 10,05125$

As coordenadas do centro de gravidade da distribuição são  $(173,375; 10,05125)$ .

1.4.  $r$ : coeficiente de correlação linear

Recorrendo à calculadora gráfica, obtém-se:  
 $r \approx 0,48$ .

Como  $r > 0$ , então o sinal do declive da reta de regressão é positivo.

1.5. a) Excluindo o ponto de coordenadas

$(210; 9,98)$  e recorrendo à calculadora gráfica, conclui-se que o coeficiente de correlação linear do novo conjunto de dados é, aproximadamente, 0,90.  
 $r \approx 0,90$

A equação da reta de regressão é:  
 $y = 0,00284x + 9,58373$

b) O valor absoluto do coeficiente de correlação, quando não se considera o ponto  $(210; 9,98)$ , é 0,90, que é maior do que era inicialmente, 0,48, o que significa que a reta de regressão se ajusta melhor aos dados.

2.1. Observando o gráfico, conclui-se que ao longo dos anos se registou uma tendência de diminuição do número de atividades desenvolvidas.

2.2. A média do conjunto de dados inicial é igual a  $53,1(3) - 5$ , ou seja,  $48,1(3)$ .

O desvio-padrão do conjunto de dados inicial não sofre nenhuma alteração, ou seja, é igual a 25,06.

$$\bar{x}_{\text{inicial}} = 48,1(3) \text{ e } s_{\text{inicial}} = 25,06.$$