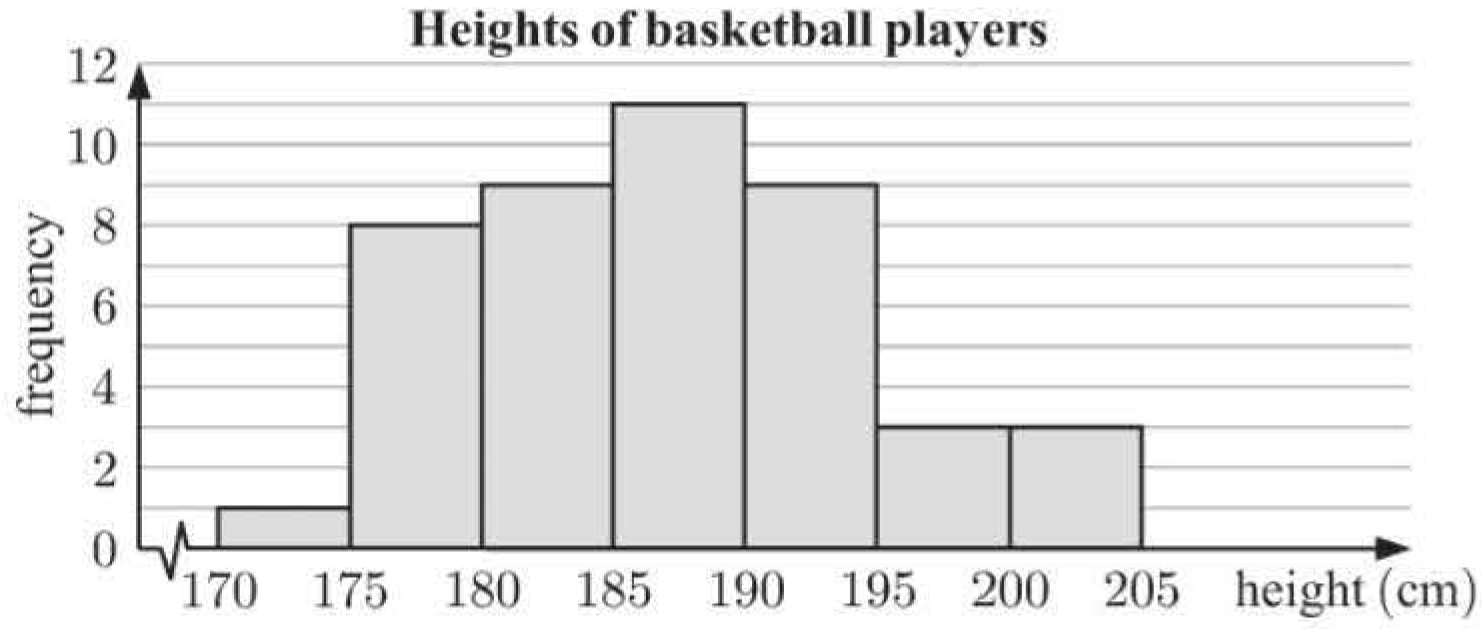
Chapter 23
DESCRIPTIVE STATISTICS

EXERCISE 23A

1 **a** Heights can take any value from 170 cm to 205 cm, including decimal values such as 181.37 cm. The 'height' variable can take any real number between 170 and 205.

b



c The modal class is the class occurring most often. This is $185 \leqslant H < 190$ cm.

d The distribution is slightly positively skewed, as there is more of a 'tail' to the right.

2 **a** The data is continuous numerical. Actual time is continuous and could be measured to the nearest second, millisecond, and so on. After it has been rounded to the nearest minute, it becomes discrete numerical data.

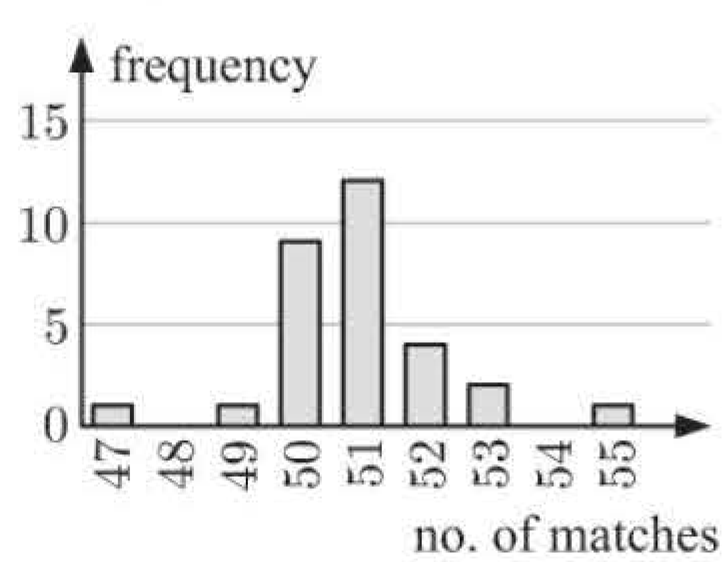
b

Time (min)	*Tally*	*Frequency*
0 - 9	~~||||~~ |	6
10 - 19	~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ ~~||||~~ |	26
20 - 29	~~||||~~ ~~||||~~ |||	13
30 - 39	~~||||~~ ||||	9
40 - 49	~~||||~~ |	6

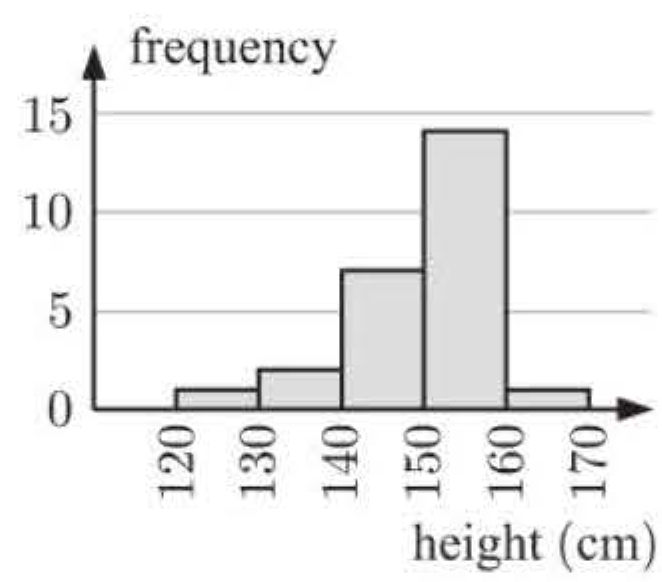
c The distribution is positively skewed, or skewed to the high end.

d The travelling time modal class is 10 - 19 min, if considering classes. (The mode is actually 10, as 10 occurs most frequently.)

3 **a** The data is discrete numerical, so a column graph should be used.



b The data is continuous, so a frequency histogram should be used.



4 **a** Number which are $\geqslant 400$ mm is $14 + 6 = 20$ seedlings.

b $12 + 18 + 42 + 28 + 14 + 6 = 120$ seedlings have been sampled.

$$\therefore \text{ \% between 349 and 400} = \frac{42+28}{120} \times 100\%$$
$$= \frac{70}{120} \times 100\%$$
$$\approx 58.3\%$$

c **i** Number less than 400 mm

$$= \frac{12+18+42+28}{120} \times 1462$$
$$= \frac{100}{120} \times 1462$$
$$\approx 1218 \text{ seedlings}$$

ii Number between 374 mm and 425 mm

$$= \frac{28+14}{120} \times 1462$$
$$= \frac{42}{120} \times 1462$$
$$\approx 512 \text{ seedlings}$$

EXERCISE 23B.1

1 **a** **i** $\text{mean} = \dfrac{2+3+3+3+4+....+9+9}{23} = \frac{129}{23} \approx 5.61$

ii median = 12th score (when in order) $= 6$

iii mode $= 6$ (6 occurs most often)

b **i** $\text{mean} = \dfrac{10+12+12+15+....+20+21}{15} = \frac{245}{15} \approx 16.3$

ii median = 8th score (when in order) $= 17$

iii mode $= 18$

c **i** $\text{mean} = \dfrac{22.4+24.6+21.8+....+23.5}{11} = \frac{273}{11} \approx 24.8$

ii median = 6th score (when in order) $= 24.9$

iii mode $= 23.5$

2 **a** $\text{mean of set A} = \dfrac{3+4+4+5+....+10}{13} \approx 6.46$ $\quad$ $\text{mean of set B} = \dfrac{3+4+4+5+....+15}{13} \approx 6.85$

b median of set A = 7th score = 7 $\quad$ median of set B = 7th score = 7

c The data sets are the same except for the last value, and the last value of set A is less than that of set B. So, the mean of set A is less than that of set B.

d The middle value of both data sets is the same, so the median is the same.

3 **a** $\text{mean} = \dfrac{23\,000+46\,000+23\,000+....+32\,000}{10} = \$29\,300$

$\text{median} = \text{middle score when in order of size} = \dfrac{\$23\,000+\$24\,000}{2} = \$23\,500$

mode $= \$23\,000$

b The mode is unsatisfactory because it is the lowest salary. It does not take the higher values into account.

c The median is too close to the lower end of the distribution since the data is positively skewed. So the median is not a satisfactory measure of the middle.

4 **a** $\text{mean} = \dfrac{3+1+0+0+....+1+0+0}{31} = \frac{99}{31} \approx 3.19$

median = 16th score (when in order) $= 0$

mode $= 0$ (most frequently occurring score)

b The median is not in the centre, as the data is very positively skewed.

c The mode is the lowest value. It does not take the higher values into account.

d Yes, 21 and 42. $\quad$ **e** No, as this would ignore actual valid data.

5 **a** $\text{mean} = \dfrac{43+55+41+37}{4} = \frac{176}{4} = 44$ points $\quad$ **b** another 44 points

c **i** $\text{new mean} = \dfrac{176+25}{5} = 40.2$ points

ii It will increase the new mean to 40.3 points as 41 points is greater than the old mean of 40.2 points.

$$\left\{\frac{5\times 40.2+41}{6} \approx 40.3\right\}$$

6 $\text{mean} = \dfrac{\text{total}}{12}$ $\quad \therefore \$15\,467 = \dfrac{\text{total}}{12}$ $\quad \therefore \text{total} = \$15\,467 \times 12 = \$185\,604$

7 mean $= \dfrac{\text{total}}{12}$ $\therefore$ $262 = \dfrac{\text{total}}{12}$ $\therefore$ total $= 262 \times 12 = 3144$ km

8 **a** mean birth mass $= \dfrac{75+70+80+....+83}{8} = \dfrac{567}{8} \approx 70.9$ grams

b mean after 2 weeks $= \dfrac{210+200+200+....+230}{8} = \dfrac{1681}{8} \approx 210$ grams

c mean increase $\approx (210.13 - 70.88)$ grams ≈ 139 grams

9 $\mu = \dfrac{\sum_{i=1}^{n} x_i}{n}$ $\therefore$ $11.6 = \dfrac{\sum_{i=1}^{10} x_i}{10}$ $\therefore$ $\sum_{i=1}^{10} x_i = 11.6 \times 10 = 116$

10 Total for first 14 matches $= 14 \times 16.5$ goals $= 231$ goals

$\therefore$ new average $= \dfrac{231+21+24}{16} = \dfrac{276}{16} = 17.25$ goals per game

11 **a** mean selling price $= \dfrac{146\,400 + 127\,600 + 211\,000 + ... + 162\,500}{10} = \$163\,770$

median selling price $= \dfrac{\text{5th} + \text{6th}}{2} = \dfrac{146\,400 + 148\,000}{2} = \$147\,200$

These figures differ by $16 570. There are more selling prices at the lower end of the market.

b **i** Use the mean as it tends to inflate the average house value of that district.

ii Use the median as you want to buy at the lowest price possible.

12 $\dfrac{5+9+11+12+13+14+17+x}{8} = 12$

$$\therefore \frac{81+x}{8} = 12$$
$$\therefore 81 + x = 96$$
$$\therefore x = 15$$

13 $\dfrac{3+0+a+a+4+a+6+a+3}{9} = 4$

$$\therefore \frac{4a+16}{9} = 4$$
$$\therefore 4a + 16 = 36$$
$$\therefore 4a = 20$$
$$\therefore a = 5$$

14 $\dfrac{29+36+32+38+35+34+39+x}{8} = 35$

$$\therefore \frac{243+x}{8} = 35$$
$$\therefore 243 + x = 280$$
$$\therefore x = 37$$

So, her 8th result was 37.

15 Total for first 10 measurements $= 10 \times 15.7$
$= 157$

Total for next 20 measurements $= 20 \times 14.3$
$= 286$

$\therefore$ mean $= \dfrac{157+286}{30} \approx 14.8$

16 If there are 9 measurements with a median of 12, then 12 must be one of the unknown measurements.
So, the measurements are 7, 9, 11, 12, 13, 14, 17, 19, and a.

mean $= \dfrac{7+9+11+12+13+14+17+19+a}{9} = \dfrac{102+a}{9}$

$$\therefore \frac{102+a}{9} = 12$$
$$\therefore 102 + a = 108$$
$$\therefore a = 6$$

So, the other measurements are 6 and 12.

17 Scores were 5 7 9 9 10 a b where $a \leqslant b$ say.

$$\text{mean} = \frac{5+7+9+9+10+a+b}{7} = 8$$

$$\therefore \quad \frac{40+a+b}{7} = 8$$

$$\therefore \quad 40+a+b = 56$$

$$\therefore \quad a+b = 16 \qquad \{a \leqslant 12,\ b \leqslant 12\}$$

Possibilities are:

a	5	6	7	8
b	11	10	9	8
	✗	✗	✓	✗

✗ ($a=8$, $b=8$): reject as modes are 8 and 9
✗ ($a=6$, $b=10$): reject as modes are 9 and 10
✗ ($a=5$, $b=11$): reject as modes are 5 and 9

So, the missing results are 7 and 9.

EXERCISE 23B.2

1 **a** The mode is 1 head, as this is the result which occurs most often.

b The median is the average of the 15th and 16th scores

$$= \frac{1+1}{2} = 1 \text{ head}$$

c

x	f	fx
0	4	0
1	12	12
2	11	22
3	3	9
$\sum$	30	43

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{43}{30} \approx 1.43 \text{ heads}$$

2 **a** **i**

x	f	fx
0	5	0
1	8	8
2	13	26
3	8	24
4	6	24
5	3	15
6	3	18
7	2	14
8	1	8
9	0	0
10	0	0
11	1	11
$\sum$	50	148

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{148}{50} = 2.96 \text{ phone calls}$$

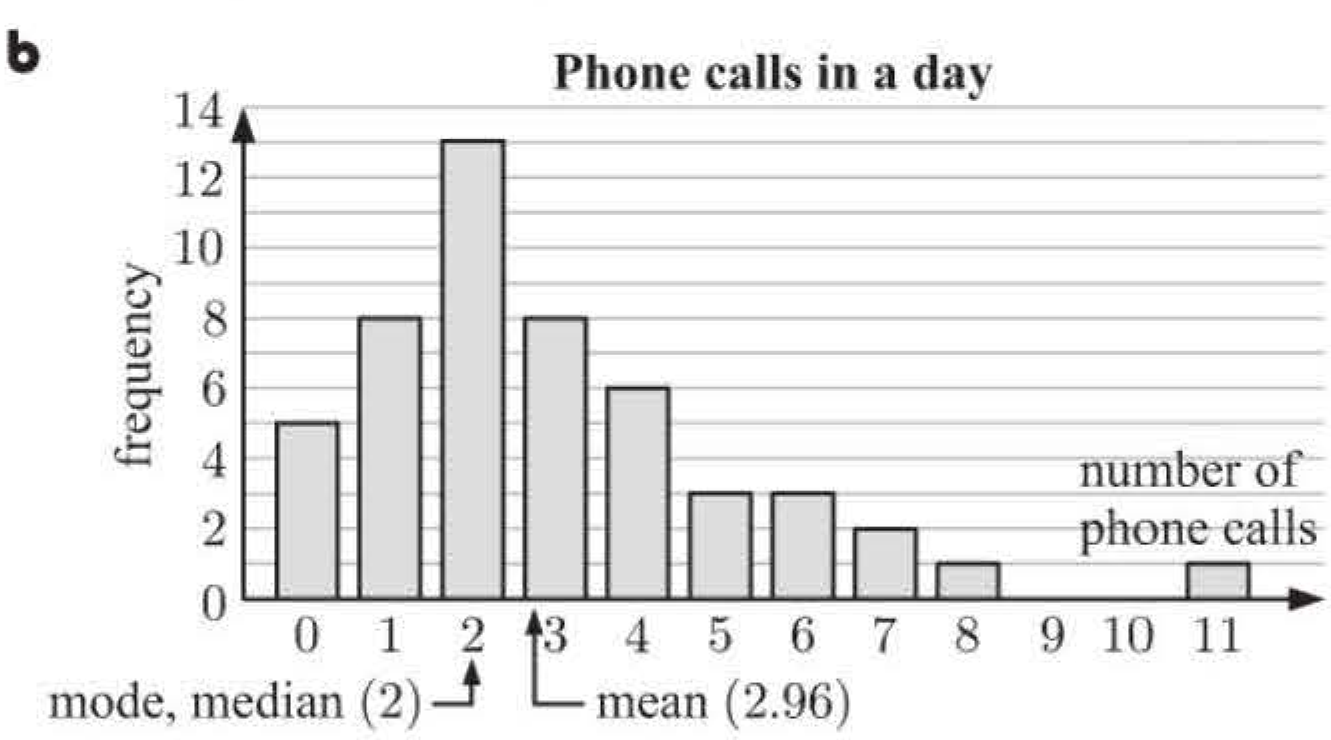
ii median
= average of 25th and 26th scores (when in order)

$$= \frac{2+2}{2} \quad \left\{\begin{array}{l}13 \text{ scores are 1 or 0} \\ 26 \text{ scores are 2, 1, or 0}\end{array}\right\}$$

= 2 phone calls

iii mode = 2 phone calls {occurs most often}

b



c The distribution is positively skewed. 11 is an outlier.

d The mean takes into account the larger numbers of phone calls.

e The mean, as it best represents all the data.

3 **a** **i** mode = 49 matches {occurs most often}

ii median = average of 15th and 16th values (when in order)

$$= \frac{49+49}{2} = 49 \text{ matches} \quad \{9 \text{ are 47 or 48 and the next 11 are 49}\}$$

iii

x	f	fx
47	5	235
48	4	192
49	11	539
50	6	300
51	3	153
52	1	52
$\sum$	30	1471

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{1471}{30} \approx 49.0 \text{ matches}$$

b No, as they claim the average is 50 matches per box.

c The sample of only 30 is not large enough. The company could have won its case by arguing that a larger sample would have found an average of 50 matches per box.

4 **a** **i**

x	f	fx
1	5	5
2	28	56
3	15	45
4	8	32
5	2	10
6	1	6
$\sum$	59	154

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{154}{59} \approx 2.61 \text{ children}$$

ii mode = 2 children {occurs most often}

iii median = 30th score = 2 children

b This school has more children per family (2.61) than the average Australian family (2.2).

c Positive as the higher values are more spread out.

d The mean is higher than the mode and median.

5 **a**

Donation (x)	*Frequency* (f)	fx
1	7	7
2	9	18
3	2	6
4	4	16
5	8	40
Total	30	87

b Total number of donations $= 7+9+2+4+8$
$= 30$

c **i** $\text{mean} = \dfrac{\sum fx}{\sum f}$
$= \dfrac{87}{30}$
$= 2.9$
$\therefore$ the mean donation is \$2.90.

ii mode = \$2 {occurs most often}

iii median
= average of the 15th and 16th values (when in order)
$= \dfrac{2+2}{2}$
$= 2$
$\therefore$ the median donation is \$2.

d The mode can be found easily using the graph only, as it is the value with the tallest column.

6 **a**

$$\text{mean} = \frac{\sum fx}{\sum f}$$

$$\therefore\ 4.45 = \frac{1\times 0 + 2\times 2 + 3\times 3 + 4\times 5 + 5\times x + 6\times 4 + 7\times 1}{0+2+3+5+x+4+1}$$

$$\therefore\ 4.45 = \frac{64+5x}{15+x}$$

$$\therefore\ 4.45(15+x) = 64+5x$$

$$\therefore\ 2.75 = 0.55x$$

$$\therefore\ x = 5$$

b From a total of $2+3+5+5+4+1 = 20$ students, $5+5+4+1 = 15$ students scored 4 or more.

$\therefore\ \frac{15}{20} = 75\%$ of the students passed.

7 **a** **Without fertiliser**

2	||
3	𝍸 𝍸 |
4	𝍸 𝍸 𝍸 ||||
5	𝍸 𝍸 𝍸 𝍸 𝍸 ||||
6	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 |
7	𝍸 𝍸 𝍸 𝍸 𝍸
8	𝍸 𝍸 ||
9	|

x	f	fx	cf
2	2	4	2
3	11	33	13
4	19	76	32
5	29	145	61
6	51	306	112
7	25	175	137
8	12	96	149
9	1	9	150

i mean $= \dfrac{\sum fx}{\sum f} = \dfrac{844}{150} \approx 5.63$ peas/pod **ii** mode $= 6$ peas/pod {occurs most often}

iii median = average of 75th and 76th scores $= \dfrac{6+6}{2} = 6$ peas/pod

b **With fertiliser**

3	||||
4	𝍸 𝍸 |||
5	𝍸 𝍸 |
6	𝍸 𝍸 𝍸 𝍸 𝍸 |||
7	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ||
8	𝍸 𝍸 𝍸 𝍸 𝍸 ||
9	𝍸 𝍸 ||||
10	||||
11	|
13	|

x	f	fx	cf
3	4	12	4
4	13	52	17
5	11	55	28
6	28	168	56
7	47	329	103
8	27	216	130
9	14	126	144
10	4	40	148
11	1	11	149
13	1	13	150

i mean $= \dfrac{\sum fx}{\sum f} = \dfrac{1022}{150} \approx 6.81$ peas/pod **ii** mode $= 7$ peas/pod {occurs most often}

iii median = average of 75th and 76th scores $= \dfrac{7+7}{2} = 7$ peas/pod

c The mean best represents the centre for this data.

d Yes, as a mean of 6.81 peas per pod is significantly greater than a mean of 5.63 peas per pod.

Note: The total yield of the crop may not have improved as, for example, the number of pods per plant may have decreased when using the fertiliser.

8 The 31 scores in order are: {15 scores below 10}, 10.1, 10.4, 10.7, 10.9, {12 scores above 11}

Median = 16th score (when in order) = 10.1 cm

9 **a** **Brand A**

x	f	fx
46	1	46
47	2	94
48	3	144
49	7	343
50	10	500
51	20	1020
52	15	780
53	3	159
55	1	55
$\sum$	62	3141

mean $= \dfrac{\sum fx}{\sum f} = \dfrac{3141}{62} \approx 50.7$

Brand B

x	f	fx
48	3	144
49	17	833
50	30	1500
51	7	357
52	2	104
53	1	53
54	1	54
$\sum$	61	3045

mean $= \dfrac{\sum fx}{\sum f} = \dfrac{3045}{61} \approx 49.9$

b Based on average contents, the C.P.S. should not prosecute either manufacturer. To the nearest toothpick, the average contents for A is 51 and for B is 50.

10 **a** **i** median salary

$= \dfrac{10\text{th} + 11\text{th}}{2}$ (when in order)

$= \dfrac{35\,000 + 28\,000}{2}$

$= €31\,500$

ii modal salary $= €28\,000$ {occurs most often}

iii

x	f	fx
50 000	1	50 000
42 000	3	126 000
35 000	6	210 000
28 000	10	280 000
$\sum$	20	666 000

mean $= \dfrac{\sum fx}{\sum f}$

$= \dfrac{666\,000}{20}$

$= €33\,300$

b The mode, as it is the most commonly occurring value.

EXERCISE 23B.3

1

midpoint (x)	f	fx
4.5	2	9
14.5	5	72.5
24.5	7	171.5
34.5	27	931.5
44.5	9	400.5
$\sum$	50	1585

$\therefore$ mean result $\approx \dfrac{1585}{50}$

≈ 31.7

2

midpoint (x)	f	fx
2500	4	10 000
3500	4	14 000
4500	9	40 500
5500	14	77 000
6500	23	149 500
7500	16	120 000
$\sum$	70	411 000

a 70

b $\approx 411\,000$ litres

≈ 411 kL

c mean

$\approx \dfrac{\sum fx}{\sum f}$

$\approx \dfrac{411\,000}{70}$

≈ 5870 litres

3 **a** $5 + 10 + 25 + 40 + 10 + 15 + 10 + 10 = 125$ people

b

midpoint (x)	*frequency* (f)	fx
85	5	425
95	10	950
105	25	2625
115	40	4600
125	10	1250
135	15	2025
145	10	1450
155	10	1550
$\sum$	125	14 875

mean

$\approx \dfrac{\sum fx}{\sum f}$

$\approx \dfrac{14\,875}{125}$

≈ 119 marks

c $\frac{15}{125} = \frac{3}{25}$ scored < 100

d There are $15 + 10 + 10 = 35$ people who scored more than 130 for the test.

$\therefore$ % who scored more than $130 = \frac{35}{125} \times 100\% = 28\%$

EXERCISE 23C.1

1 **a** Looking at the graphs, Sample A appears to have the wider spread.

b *Sample A*:

x	f	fx
4	1	4
5	2	10
6	3	18
7	4	28
8	5	40
9	4	36
10	3	30
11	2	22
12	1	12
$\sum$	25	200

$\therefore$ mean $= \frac{200}{25} = 8$

Sample B:

x	f	fx
6	2	12
7	6	42
8	9	72
9	6	54
10	2	20
$\sum$	25	200

$\therefore$ mean $= \frac{200}{25} = 8$

c *Sample A*:

x	$x-\mu$	$(x-\mu)^2$	f	$f(x-\mu)^2$
4	-4	16	1	16
5	-3	9	2	18
6	-2	4	3	12
7	-1	1	4	4
8	0	0	5	0
9	1	1	4	4
10	2	4	3	12
11	3	9	2	18
12	4	16	1	16
			$\sum$	100

$\therefore \quad \sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{100}{25}} = 2$

Sample B:

x	$x-\mu$	$(x-\mu)^2$	f	$f(x-\mu)^2$
6	-2	4	2	8
7	-1	1	6	6
8	0	0	9	0
9	1	1	6	6
10	2	4	2	8
			$\sum$	28

$\therefore \quad \sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{28}{25}} \approx 1.06$

The standard deviation is higher for Sample A.

$\therefore$ Sample A has a greater spread.

d The standard deviation is calculated using all data values, not just two. (Range only uses maximum and minimum; IQR only uses the upper and lower quartiles.)

2 **a** *Andrew*: $\mu = \dfrac{23+17+....+28+32}{8} = 25$

x	$x-\mu$	$(x-\mu)^2$
23	-2	4
17	-8	64
31	6	36
25	0	0
25	0	0
19	-6	36
28	3	9
32	7	49
	$\sum$	198

$\therefore \quad \sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{198}{8}} \approx 4.97$

Brad: $\mu = \dfrac{9+29+....+38+43}{8} = 30.5$

x	$x-\mu$	$(x-\mu)^2$
9	-21.5	462.25
29	-1.5	2.25
41	10.5	110.25
26	-4.5	20.25
14	-16.5	272.25
44	13.5	182.25
38	7.5	56.25
43	12.5	156.25
	$\sum$	1262

$\therefore \quad \sigma = \sqrt{\dfrac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{1262}{8}} \approx 12.6$

b Andrew, as he has the smaller standard deviation.

3 **a** Rockets have mean $= \dfrac{0+10+1+9+11+0+8+5+6+7}{10} = \dfrac{57}{10} = 5.7$ runs

Bullets have mean $= \dfrac{4+3+4+1+4+11+7+6+12+5}{10} = \dfrac{57}{10} = 5.7$ runs

Rockets' range $= 11 - 0 = 11$ runs Bullets' range $= 12 - 1 = 11$ runs

b We suspect the Rockets, as they have two zeros.

c *Rockets*

x	$(x-\mu)^2$
0	$(-5.7)^2$
10	$(4.3)^2$
1	$(-4.7)^2$
9	$(3.3)^2$
11	$(5.3)^2$
0	$(-5.7)^2$
8	$(2.3)^2$
5	$(-0.7)^2$
6	$(0.3)^2$
7	$(1.3)^2$
	152.1

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{152.1}{10}} = 3.9 \text{ runs}$$

greater variability

Bullets

x	$(x-\mu)^2$
4	$(-1.7)^2$
3	$(-2.7)^2$
4	$(-1.7)^2$
1	$(-4.7)^2$
4	$(-1.7)^2$
11	$(5.3)^2$
7	$(1.3)^2$
6	$(0.3)^2$
12	$(6.3)^2$
5	$(-0.7)^2$
	108.1

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{108.1}{10}} \approx 3.29 \text{ runs}$$

d The standard deviation, as it takes all values into account, not just the lowest and highest.

4 **a** We suspect variability in standard deviation since the factors may change every day.

b **i** mean **ii** standard deviation

c A low standard deviation would mean less variability in the volume of soft drink per can.

5 **a**

x	$(x-\mu)^2$
79	10^2
64	$(-5)^2$
59	$(-10)^2$
71	2^2
68	$(-1)^2$
68	$(-1)^2$
74	5^2
483	256

$$\mu = \frac{\sum x}{n} = \frac{483}{7} = 69 \text{ kg}$$

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{256}{7}} \approx 6.05 \text{ kg}$$

b

x	$(x-\mu)^2$
89	10^2
74	$(-5)^2$
69	$(-10)^2$
81	2^2
78	$(-1)^2$
78	$(-1)^2$
84	5^2
553	256

$$\mu = \frac{\sum x}{n} = \frac{553}{7} = 79 \text{ kg}$$

$$\sigma \approx 6.05 \text{ kg}$$

c The distribution has simply shifted by 10 kg. The mean increases by 10 kg and the standard deviation remains the same. In general, changing the values by a constant does not affect standard deviation.

6 **a**

x	$(x-\mu)^2$
0.8	$(-0.21)^2$
1.1	$(0.09)^2$
1.2	$(0.19)^2$
0.9	$(-0.11)^2$
1.2	$(0.19)^2$
1.2	$(0.19)^2$
0.9	$(-0.11)^2$
0.7	$(-0.31)^2$
1.0	$(-0.01)^2$
1.1	$(0.09)^2$
10.1	0.289

$$\mu = \frac{\sum x}{n} = \frac{10.1}{10} = 1.01 \text{ kg}$$

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{0.289}{10}} = 0.17 \text{ kg}$$

b

x	$(x-\mu)^2$
1.6	$(-0.42)^2$
2.2	$(0.18)^2$
2.4	$(0.38)^2$
1.8	$(-0.22)^2$
2.4	$(0.38)^2$
2.4	$(0.38)^2$
1.8	$(-0.22)^2$
1.4	$(-0.62)^2$
2.0	$(-0.02)^2$
2.2	$(0.18)^2$
20.2	1.156

$$\mu = \frac{\sum x}{n} = \frac{20.2}{10} = 2.02 \text{ kg}$$

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{1.156}{10}} = 0.34 \text{ kg}$$

c Doubling the values doubles the mean and the standard deviation.

7 **a** $\mu = \dfrac{0.8+0.6+0.7+0.8+0.4+2.8}{6}$

≈ 1.017

x	$(x-\mu)^2$
0.8	$(-0.217)^2$
0.6	$(-0.417)^2$
0.7	$(-0.317)^2$
0.8	$(-0.217)^2$
0.4	$(-0.617)^2$
2.8	$(1.783)^2$
$\sum$	3.928

$$\therefore\ \sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} \approx \sqrt{\frac{3.928}{6}} \approx 0.809$$

b $\mu = \dfrac{0.8+0.6+0.7+0.8+0.4}{5}$

$= 0.66$

x	$(x-\mu)^2$
0.8	$(0.14)^2$
0.6	$(-0.06)^2$
0.7	$(0.04)^2$
0.8	$(0.14)^2$
0.4	$(-0.26)^2$
$\sum$	0.112

$$\therefore\ \sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{0.112}{5}} \approx 0.150$$

c The extreme value greatly increases the standard deviation.

8 $\mu = \dfrac{1+3+5+7+4+5+p+q}{8} = 5$

$$\begin{aligned}\therefore\ 25+p+q &= 40\\ \therefore\ p+q &= 15\\ \therefore\ q &= 15-p\end{aligned}$$

and $\sigma = \sqrt{\dfrac{(-4)^2+(-2)^2+0^2+2^2+(-1)^2+0^2+(p-5)^2+(q-5)^2}{8}} = \sqrt{5.25}$

$$\begin{aligned}\therefore\ \frac{16+4+4+1+(p-5)^2+(15-p-5)^2}{8} &= 5.25\\ \therefore\ 25+p^2-10p+25+100-20p+p^2 &= 42\\ \therefore\ 2p^2-30p+108 &= 0\\ \therefore\ p^2-15p+54 &= 0\\ \therefore\ (p-6)(p-9) &= 0\end{aligned}$$

$\therefore\ p = 6$ or 9 and $q = 9$ or 6

But $p < q$ $\quad\therefore\ p = 6,\ q = 9$

9 $\mu = \dfrac{3+9+5+5+6+4+a+6+b+8}{10} = 6$

$$\begin{aligned}\therefore\ \frac{46+a+b}{10} &= 6\\ \therefore\ 46+a+b &= 60\\ \therefore\ a+b &= 14\\ \therefore\ b &= 14-a\end{aligned}$$

and $\sigma = \sqrt{\dfrac{(-3)^2+3^2+(-1)^2+(-1)^2+(-2)^2+(a-6)^2+(b-6)^2+2^2}{10}} = \sqrt{3.2}$

$$\begin{aligned}\therefore\ 9+9+1+1+4+4+(a-6)^2+(14-a-6)^2 &= 32\\ \therefore\ 28+a^2-12a+36+64-16a+a^2 &= 32\\ \therefore\ 2a^2-28a+96 &= 0\\ \therefore\ a^2-14a+48 &= 0\\ \therefore\ (a-6)(a-8) &= 0\end{aligned}$$

$\therefore\ a = 6$ or 8 and $b = 8$ or 6

But $a > b$ $\quad\therefore\ a = 8,\ b = 6$

10 **a** $$\sum_{i=1}^{n}(x_i-\mu)^2 = \sum_{i=1}^{n}(x_i{}^2) - 2\mu\sum_{i=1}^{n}x_i + \sum_{i=1}^{n}\mu^2$$
$$= \sum_{i=1}^{n}(x_i{}^2) - 2\mu(x_1+x_2+....+x_n) + n\mu^2$$
$$= \sum_{i=1}^{n}(x_i{}^2) - 2\mu(n\mu) + n\mu^2 \qquad \{\mu = \frac{x_1+x_2+....+x_n}{n}\}$$
$$= \sum_{i=1}^{n}(x_i{}^2) - n\mu^2$$

b $$\sigma = \sqrt{\frac{\sum_{i=1}^{25}(x_i-\mu)^2}{25}} = 5.2$$
$$\therefore \frac{\sum_{i=1}^{25}(x_i-\mu)^2}{25} = 27.04$$
$$\therefore \sum_{i=1}^{25}(x_i-\mu)^2 = 676$$

$$\therefore \sum_{i=1}^{25}(x_i{}^2) - 25\mu^2 = 676 \qquad \{\text{using } \mathbf{a}\}$$
$$\therefore 2568.25 - 25\mu^2 = 676$$
$$\therefore 1892.25 = 25\mu^2$$
$$\therefore \mu^2 = 75.69$$
$$\therefore \mu = 8.7 \qquad \{\text{assuming positive data}\}$$

EXERCISE 23C.2

1

x	f	fx	$x-\mu$	$f(x-\mu)^2$
0	14	0	-1.7241	41.62
1	18	18	-0.7241	9.44
2	13	26	0.2759	0.99
3	5	15	1.2759	8.14
4	3	12	2.2759	15.54
5	2	10	3.2759	21.46
6	2	12	4.2759	36.57
7	1	7	5.2759	27.83
$\sum$	58	100		161.59

$$\mu = \frac{\sum fx}{\sum f} = \frac{100}{58} \approx 1.72 \text{ children}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} \approx \sqrt{\frac{161.59}{58}} \approx 1.67 \text{ children}$$

2

x	f	fx	$x-\mu$	$f(x-\mu)^2$
11	2	22	-3.48	24.22
12	1	12	-2.48	6.150
13	4	52	-1.48	8.762
14	5	70	-0.48	1.152
15	6	90	0.52	1.622
16	4	64	1.52	9.242
17	2	34	2.52	12.70
18	1	18	3.52	12.39
$\sum$	25	362		76.24

$$\mu = \frac{\sum fx}{\sum f} = \frac{362}{25} \approx 14.5 \text{ years}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} = \sqrt{\frac{76.24}{25}} \approx 1.75 \text{ years}$$

3

x	f	fx	$x-\mu$	$f(x-\mu)^2$
33	1	33	-4.2708	18.24
35	5	175	-2.2708	25.78
36	7	252	-1.2708	11.30
37	13	481	-0.2708	0.95
38	12	456	0.7292	6.38
39	8	312	1.7292	23.92
40	2	80	2.7292	14.90
$\sum$	48	1789		101.47

$$\mu = \frac{\sum fx}{\sum f} = \frac{1789}{48} \approx 37.3 \text{ toothpicks}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} = \sqrt{\frac{101.47}{48}} \approx 1.45 \text{ toothpicks}$$

4

Midpoint (x)	f	fx	$f(x-\mu)^2$
41	1	41	52.80
43	1	43	27.74
45	3	135	32.01
47	7	329	11.23
49	11	539	5.91
51	5	255	37.35
53	2	106	44.80
$\sum$	30	1448	211.87

$$\mu = \frac{\sum fx}{\sum f} = \frac{1448}{30} \approx 48.3 \text{ cm}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} \approx \sqrt{\frac{211.87}{30}} \approx 2.66 \text{ cm}$$

5

Midpoint (x)	f	fx	$f(x-\mu)^2$
364.995	17	6204.9	10 881.53
374.995	38	14 249.8	8895.42
384.995	47	18 094.8	1320.23
394.995	57	22 514.7	1259.13
404.995	18	7289.9	3889.62
414.995	10	4150.0	6100.9
424.995	10	4250.0	12 040.9
434.995	3	1305.0	5994.27
$\sum$	200	78 059	50 382

$$\mu = \frac{\sum fx}{\sum f} = \frac{78\,059}{200} \approx \$390.30$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} = \sqrt{\frac{50\,382}{200}} \approx \$15.87$$

REVIEW SET 23A

1 **a** **i** There are 30 colonies.

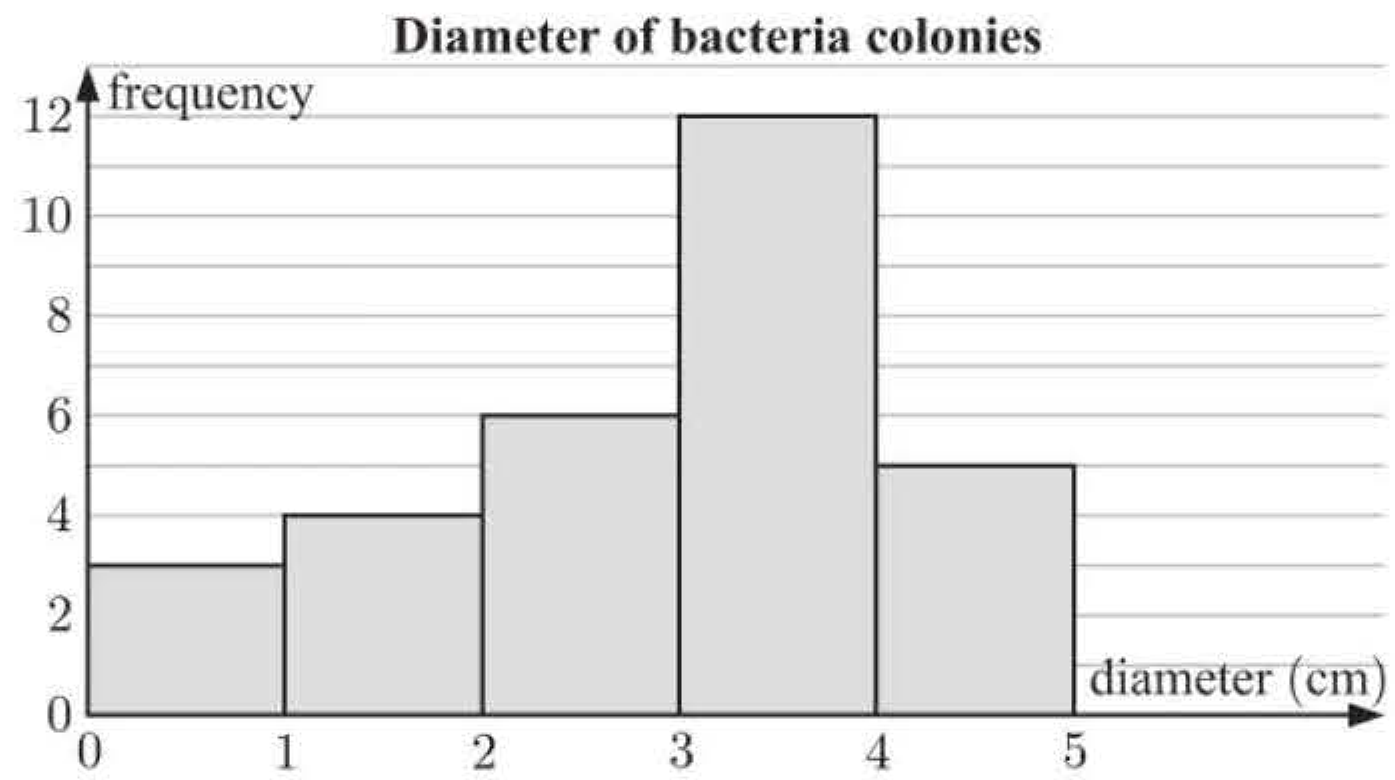
$\therefore$ median = average of 15th and 16th colonies

$$= \frac{3.1+3.2}{2} = 3.15 \text{ cm}$$

ii range $= 4.9 - 0.4 = 4.5$ cm

b

Diameter	*Tally*	*Frequency*
0.0 - 0.9	|||	3
1.0 - 1.9	||||	4
2.0 - 2.9	~~||||~~ |	6
3.0 - 3.9	~~||||~~ ~~||||~~ ||	12
4.0 - 4.9	~~||||~~	5



c The distribution is slightly negatively skewed.

2 If the mode is 6 then one of the unknown numbers must be 6.
Suppose the other unknown number is x.

$$\therefore \frac{4+6+9+6+3+x}{6} = 6$$

$$\therefore 28 + x = 36$$

$$\therefore x = 8$$

Since $a > b$, $a = 8$ and $b = 6$.

3 **a** We first organise the data into tables:

Girls:

Time x (s)	f	fx
33	1	33
34	3	102
35	5	175
36	4	144
37	4	148
38	1	38
39	1	39
40	0	0
41	1	41
Total	20	720

Boys:

Time x (s)	f	fx
32	1	32
33	4	132
34	5	170
35	6	210
36	3	108
37	1	37
Total	20	689

Both boys and girls have 20 member squads, so the median is the average of the 10th and 11th swimmer.

Girls:

$$\text{median} = \frac{36+36}{2} = 36 \text{ s}$$

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{720}{20} = 36 \text{ s}$$

The tallest column on the *Girls* histogram is the '35' column.

$\therefore$ the modal class is 34.5 - 35.5 s.

Boys:

$$\text{median} = \frac{34+35}{2} = 34.5 \text{ s}$$

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{689}{20} = 34.45 \text{ s}$$

The tallest column on the *Boys* histogram is the '35' column.

$\therefore$ the modal class is 34.5 - 35.5 s.

So,

Distribution	*Girls*	*Boys*
shape	positively skewed	approximately symmetrical
median	36 s	34.5 s
mean	36 s	34.45 s
modal class	34.5 - 35.5 s	34.5 - 35.5 s

b The girls' distribution is positively skewed and the boys' distribution is approximately symmetrical. The median and mean swim times for boys are both about 1.5 seconds lower than for girls. Despite this, the distributions have the same modal class because of the skewness in the girls' distribution. The analysis supports the conjecture that boys generally swim faster than girls with less spread of times.

4 $\text{mean} = \dfrac{2+5+k+k+3+k+7+4}{8} = \dfrac{21+3k}{8}$

$$\therefore \frac{21+3k}{8} = 6$$
$$\therefore 21+3k = 48$$
$$\therefore 3k = 27$$
$$\therefore k = 9$$

5 **a** **i** mode $= 4$ {occurs most often}

ii total $= 2+6+12+8+6+5+3+2 = 44$

median = average of 22nd and 23rd data values

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

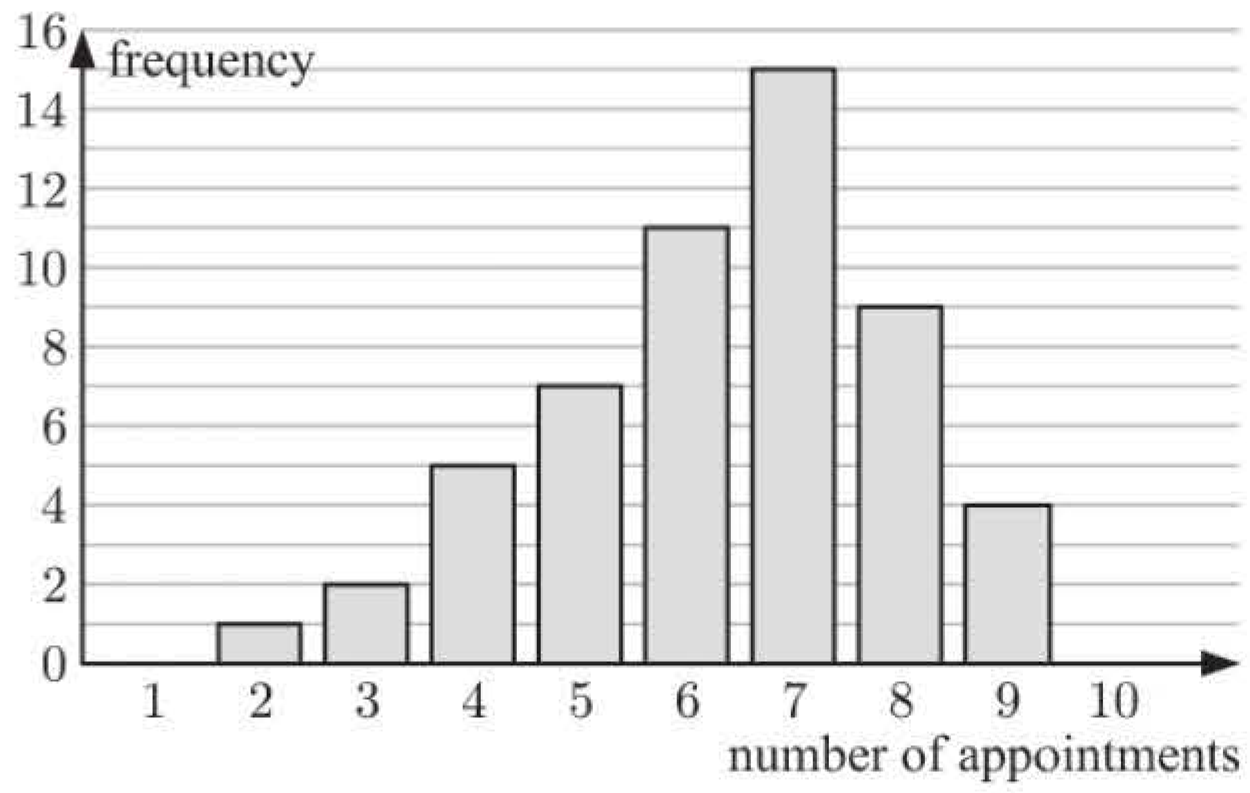
iii

x	f	fx
2	2	4
3	6	18
4	12	48
5	8	40
6	6	36
7	5	35
8	3	24
9	2	18
$\sum$	44	223

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{223}{44} \approx 5.07$$

b The data is positively skewed.

6 **a**



b The data is negatively skewed.

c **i** mode = 7 appointments {occurs most often}

ii

Number of appointments x	*Frequency* f	fx
2	1	2
3	2	6
4	5	20
5	7	35
6	11	66
7	15	105
8	9	72
9	4	36
$\sum$	54	342

$$\text{mean} = \frac{\sum fx}{\sum f} = \frac{342}{54} \approx 6.33 \text{ appointments}$$

7 $$\frac{a+b+c+d+e}{5} = 8$$

$$\therefore\ a+b+c+d+e = 40 \quad .... (1)$$

$$\mu = \frac{(10-a)+(10-b)+(20-c)+(20-d)+(50-e)}{5} = \frac{110-(a+b+c+d+e)}{5}$$
$$= \frac{110-40}{5} \quad \{\text{using (1)}\}$$
$$= \frac{70}{5}$$
$$= 14$$

8 $$\mu = \frac{12+13+8+10+14+7+a+b}{8} = 10$$
$$\therefore\ 64+a+b = 80$$
$$\therefore\ a+b = 16$$
$$\therefore\ b = 16-a$$

and $\sigma = \sqrt{\dfrac{2^2+3^2+(-2)^2+0^2+4^2+(-3)^2+(a-10)^2+(b-10)^2}{8}} = \sqrt{8.5}$

$$\therefore \frac{4+9+4+16+9+(a-10)^2+(16-a-10)^2}{8} = 8.5$$

$$\therefore 42 + a^2 - 20a + 100 + 36 - 12a + a^2 = 68$$

$$\therefore 2a^2 - 32a + 110 = 0$$

$$\therefore a^2 - 16a + 55 = 0$$

$$\therefore (a-5)(a-11) = 0$$

$$\therefore a = 5 \text{ or } 11 \quad \text{and} \quad b = 11 \text{ or } 5$$

But $a < b \quad \therefore a = 5,\ b = 11$

9 **a** $\mu = \dfrac{8+11+12+9+a}{5}$

$= \dfrac{40+a}{5}$

$= 8 + \frac{1}{5}a$

b

$$\sigma^2 = \frac{\sum {x_i}^2}{n} - \mu^2 = 6$$

$$\therefore \frac{8^2+11^2+12^2+9^2+a^2}{5} - \mu^2 = 6$$

$$\therefore \frac{410+a^2}{5} - \left(\frac{40+a}{5}\right)^2 = 6$$

$$\therefore \frac{5(410+a^2) - (1600+80a+a^2)}{25} = 6$$

$$\therefore 2050 + 5a^2 - 1600 - 80a - a^2 = 150$$

$$\therefore 4a^2 - 80a + 300 = 0$$

$$\therefore a^2 - 20a + 75 = 0$$

$$\therefore (a-5)(a-15) = 0$$

$$\therefore a = 5 \text{ or } 15$$

10 Let the five consecutive integers be $(x-2)$, $(x-1)$, x, $(x+1)$, and $(x+2)$.

$\mu = \dfrac{(x-2)+(x-1)+x+(x+1)+(x+2)}{5}$

$= \dfrac{5x}{5}$

$= x$

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{n} = \frac{(x-2-x)^2+(x-1-x)^2+(x-x)^2+(x+1-x)^2+(x+2-x)^2}{5}$$

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

$$= \frac{4+1+1+4}{5}$$

$$= \frac{10}{5}$$

$$= 2$$

REVIEW SET 23B

1 **a** highest $= 97.5$ m, lowest $= 64.6$ m

b The range $= 97.5 - 64.6 = 32.9$

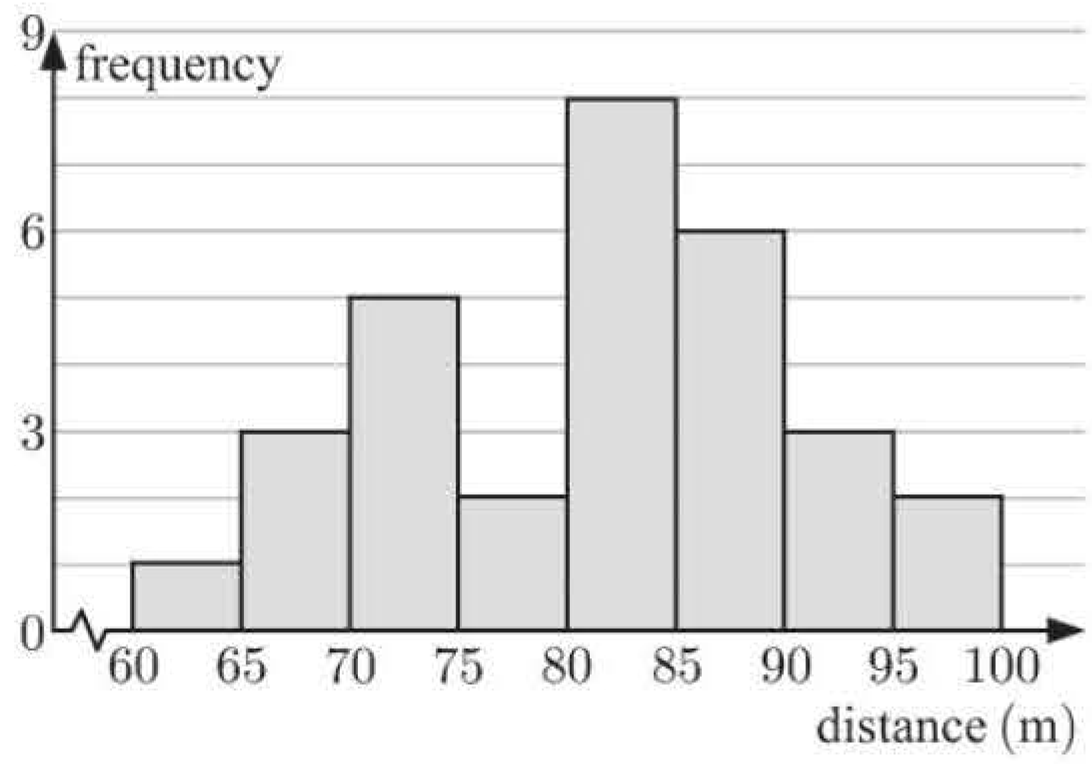
So, if intervals of length 5 are used we need about 7 of them.

We choose $60 \leqslant d < 65$, $65 \leqslant d < 70$, $70 \leqslant d < 75$, and so on.

c **Distances thrown by Thabiso**

Distance (m)	*Tally*	*Frequency* (f)
$60 \leqslant d < 65$	$\vert$	1
$65 \leqslant d < 70$	$\vert\vert\vert$	3
$70 \leqslant d < 75$	$\cancel{\vert\vert\vert\vert}$	5
$75 \leqslant d < 80$	$\vert\vert$	2
$80 \leqslant d < 85$	$\cancel{\vert\vert\vert\vert}\ \vert\vert\vert$	8
$85 \leqslant d < 90$	$\cancel{\vert\vert\vert\vert}\ \vert$	6
$90 \leqslant d < 95$	$\vert\vert\vert$	3
$95 \leqslant d < 100$	$\vert\vert$	2
	Total	30

d **Frequency histogram displaying the distance Thabiso throws a baseball**



e Using technology:

i $\mu \approx 81.1$ m **ii** median ≈ 83.1 m

2 **a** $$\text{mean} = \frac{(k-2)+k+(k+3)+(k+3)}{4} = \frac{4k+4}{4} = \frac{\cancel{4}(k+1)}{\cancel{4}} = k+1 \quad \text{as required}$$

b The members are now
$k,\ k+2,\ k+5,$ and $k+5$.

$\therefore$ new mean
$$= \frac{k+(k+2)+(k+5)+(k+5)}{4} = \frac{4k+12}{4} = \frac{\cancel{4}(k+3)}{\cancel{4}} = k+3$$

3

Scores	f	midpt x	fx	$x-\mu$	$f(x-\mu)^2$
$0 \leqslant x < 10$	1	5	5	-21	441
$10 \leqslant x < 20$	13	15	195	-11	1573
$20 \leqslant x < 30$	27	25	675	-1	27
$30 \leqslant x < 40$	17	35	595	9	1377
$40 \leqslant x < 50$	2	45	90	19	722
$\sum$	60		1560		4140

$$\mu = \frac{\sum fx}{\sum f} = \frac{1560}{60} = 26$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} = \sqrt{\frac{4140}{60}} \approx 8.31$$

4

Litres (x)	f	fx	$f(x-\mu)^2$
17.5	5	87.5	1299.38
22.5	13	292.5	1607.71
27.5	17	467.5	636.87
32.5	29	942.5	36.42
37.5	27	1012.5	406.32
42.5	18	765	1419.16
47.5	7	332.5	1348.45
$\sum$	116	3900	6754.31

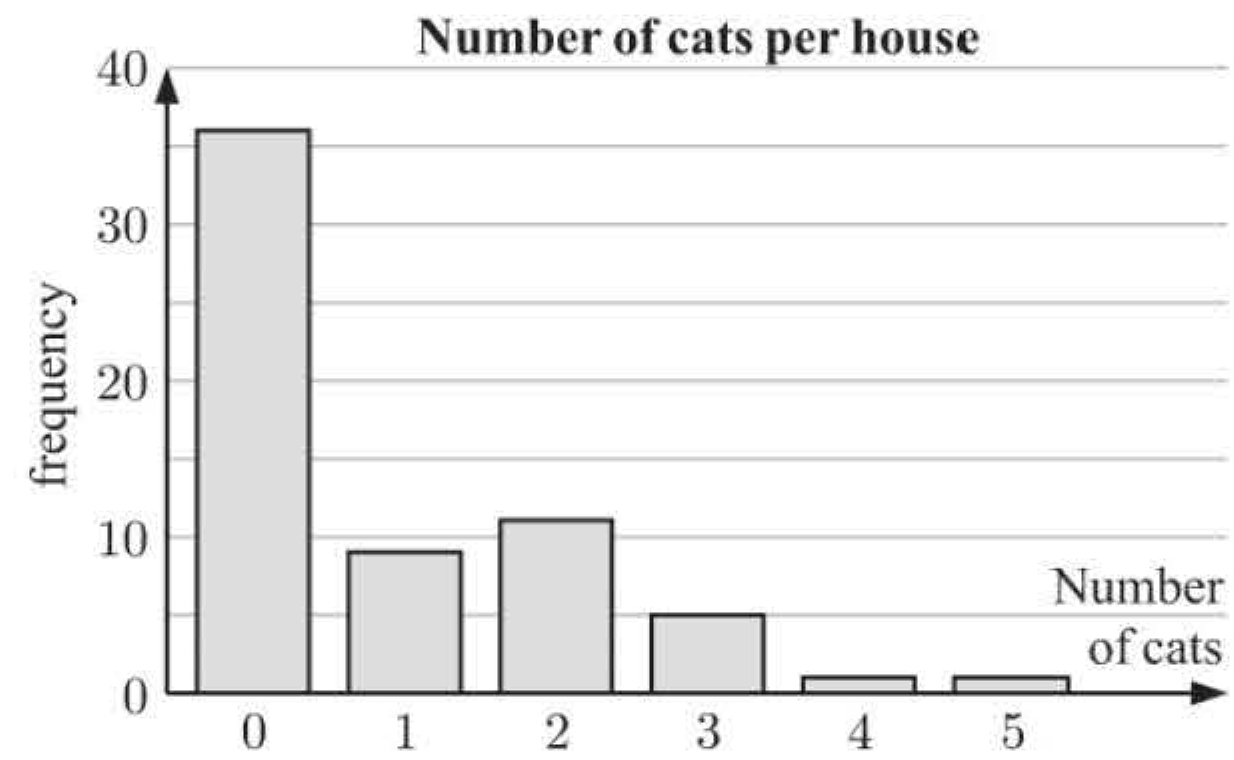
$$\mu = \frac{\sum fx}{\sum f} = \frac{3900}{116} \approx 33.6 \text{ litres}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} \approx \sqrt{\frac{6754.31}{116}} \approx 7.63 \text{ litres}$$

5 **a** Using technology, $\mu \approx 49.6$, $\sigma \approx 1.60$.

b This does not justify the claim. A larger sample is needed.

6 **a**



b The distribution is positively skewed.

c **i** The mode is 0 cats.

ii

Number of cats	*Frequency*	fx
0	36	0
1	9	9
2	11	22
3	5	15
4	1	4
5	1	5
Total	63	55

$$\mu = \frac{\sum fx}{\sum f} = \frac{55}{63} \approx 0.873$$

iii There are 63 values, so the median is the $\dfrac{63+1}{2} = 32$nd value.

$\therefore$ median $= 0$ cats.

d The mean, as it suggests that some people have cats. (The mode and median are both 0.)

7 **a** total cars $= 32 + 85 + 123 + 97 + 62 + 27 = 426$ cars

b

Midpoint (x)	f	fx	$f(x-\mu)^2$
0.5	32	16	178.100
1.5	85	127.5	157.021
2.5	123	307.5	15.866
3.5	97	339.5	39.836
4.5	62	279	166.927
5.5	27	148.5	188.300
$\sum$	426	1218	746.05

$$\mu = \frac{\sum fx}{\sum f} = \frac{1218}{426} \approx 2.86 \text{ cars}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} = \sqrt{\frac{746.05}{426}} \approx 1.32 \text{ cars}$$

8 **a** $\mu = \dfrac{42 + 58 + 74 + 62 + 51 + 45 + 73 + 54 + 66 + 84}{10}$

$= 60.9$

b median = average of 5th and 6th values (when in order)

$= \dfrac{58+62}{2} = 60$

c

x	$(x-\mu)^2$
42	357.21
58	8.41
74	171.61
62	1.21
51	98.01
45	252.81
73	146.41
54	47.61
66	26.01
84	533.61
$\sum$	1642.9

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{1642.9}{10}} \approx 12.8$$

9 **a** Kevin: $\mu = 41.2$, Felicity: $\mu = 39.5$ {using technology}

b Kevin: $\sigma \approx 7.61$, Felicity: $\sigma \approx 9.22$ {using technology}

c Felicity has a lower mean, so she generally solved the puzzles faster.

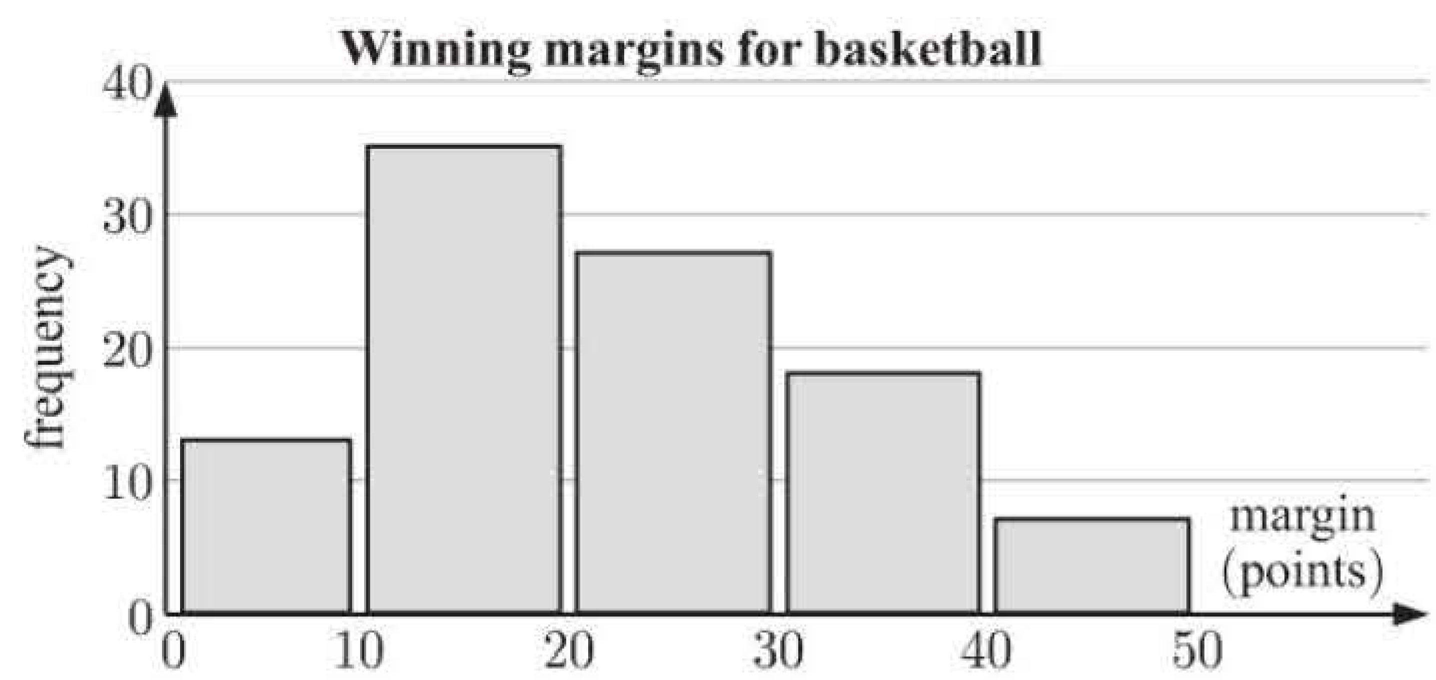
d Kevin has a lower standard deviation, so he was more consistent.

10 $$\begin{aligned}\sigma &= \frac{\sum {x_i}^2}{n} - \mu^2\\ &= \frac{\sum\limits_{i=1}^{20} {x_i}^2}{20} - \mu^2\\ &= \frac{2872}{20} - 11^2\\ &= 22.6\end{aligned}$$

REVIEW SET 23C

1 **a** The data is discrete.

b



c No, as we do not know each individual data value, only the intervals they fall in.

2 **a**

Score	f	*Product*
2	3	6
5	2	10
x	4	$4x$
$x+6$	1	$x+6$
Total	10	$5x+22$

$$\begin{aligned}\text{mean} &= 5.7\\ \therefore \frac{5x+22}{10} &= 5.7\\ \therefore 5x+22 &= 57\\ \therefore 5x &= 35\\ \therefore x &= 7\end{aligned}$$

b The data set is:

Score x	f	$x-\mu$	$(x-\mu)^2$	$f(x-\mu)^2$
2	3	-3.7	13.69	41.07
5	2	-0.7	0.49	0.98
7	4	1.3	1.69	6.76
13	1	7.3	53.29	53.29
$\sum f$	10		$\sum f(x-\mu)^2$	102.1

$$\begin{aligned}\text{Variance} &= \frac{\sum f(x-\mu)^2}{\sum f}\\ &= \frac{102.1}{10}\\ &= 10.21\\ \therefore \sigma^2 &\approx 10.2\end{aligned}$$

3 Use technology or

Midpoint (x)	f	fx
274.5	14	3843
324.5	34	11 033
374.5	68	25 466
424.5	72	30 564
474.5	54	25 623
524.5	23	12 063.5
574.5	7	4021.5
$\sum$	272	112 614

$$\begin{aligned}\mu &= \frac{\sum fx}{\sum f}\\ &= \frac{112\,614}{272}\\ &\approx 414 \text{ customers}\end{aligned}$$

4 The mode is 36, so at least one of m and n must be 36.

$$\mu = \frac{33+18+25+40+36+41+36+x}{8} = 32$$

$$\therefore\ 229 + x = 256$$

$$\therefore\ x = 27 \text{ is the other number.}$$

Now $m < n$, $\quad\therefore\ m = 27,\ n = 36$

5 **a** In order, the data set is $\{m-3,\ m-2,\ m,\ m+1,\ m+4,\ m+6\}$

median = average of 3rd and 4th values

$$= \frac{m+(m+1)}{2}$$
$$= m + \tfrac{1}{2}$$

$$\text{mean} = \frac{(m-3)+(m-2)+m+(m+1)+(m+4)+(m+6)}{6}$$
$$= \frac{6m+6}{6}$$
$$= m+1$$

b $$\sigma^2 = \frac{\sum(x_i-\mu)^2}{n} = \frac{(m-3-(m+1))^2+(m-2-(m+1))^2+\dots.+(m+6-(m+1))^2}{6}$$
$$= \frac{(-4)^2+(-3)^2+(-1)^2+0^2+3^2+5^2}{6}$$
$$= \frac{60}{6}$$
$$= 10$$

6 Using technology with x values 74.995, 84.995, 94.995, and so on, $\mu \approx$ €103.51 and $\sigma \approx$ €19.40

7 **a** No, it will not be the same. Extreme values have less effect on the standard deviation of a larger population.

b **i** The mean would be used. **ii** The standard deviation would be used.

c A low standard deviation means that the weight of biscuits in each packet is, on average, close to 250 g.

8

Midpoint (x)	f	fx	$f(x-\mu)^2$
6.5	5	32.5	21.126
7.5	19	142.5	21.170
8.5	38	323	0.117
9.5	22	209	19.623
10.5	6	63	22.685
$\sum$	90	770	84.72

$$\mu = \frac{\sum fx}{\sum f} = \frac{770}{90} \approx 8.56 \text{ hours}$$

$$\sigma = \sqrt{\frac{\sum f(x-\mu)^2}{\sum f}} \approx \sqrt{\frac{84.72}{90}} \approx 0.970 \text{ hours}$$

9 **a** Roger: $\mu \approx 84.1,\ \sigma \approx 6.60$

Clinton: $\mu \approx 76.8,\ \sigma \approx 3.83$ {using technology}

b Clinton has a lower mean, so he generally has the lower score.

c Roger has a higher standard variation, so he has greater variation in his scores.