

Chapter 1 — Univariate data

Exercise 1A — Categorical data

Categorical Data = observations/records (qualitative).

Numerical Data = quantitative amounts.

1 a Number of matches $\Rightarrow$ numerical

b Sex type qualitative $\Rightarrow$ categorical

c Length of cod $\Rightarrow$ numerical

d Record of weather conditions/observations $\Rightarrow$ categorical

e Rating of 1 to 5 $\Rightarrow$ numerical

2 a Strongly agree $\rightarrow$ strongly disagree $\Rightarrow$ ordinal

b Favourite rock bands $\Rightarrow$ nominal

c Mode of transport $\Rightarrow$ nominal

d Rate a book on a scale $\Rightarrow$ ordinal

e Achievement on a scale $\Rightarrow$ ordinal

3 a Heights $\Rightarrow$ continuous

b Number of Jelly beans $\Rightarrow$ discrete

c Time taken to... $\Rightarrow$ continuous

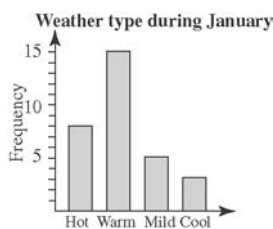
d Petrol consumption $\Rightarrow$ continuous

e IQ $\Rightarrow$ discrete

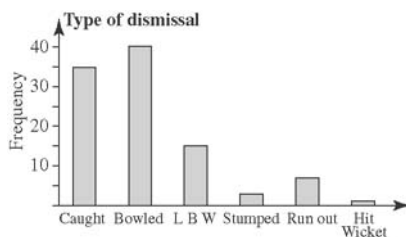
4 Ordinal data are categorical data in which there is a definite order associated with the categories used. The answer is **B**.

5 Number of people $\Rightarrow$ Numerical discrete data. The answer is **D**.

6



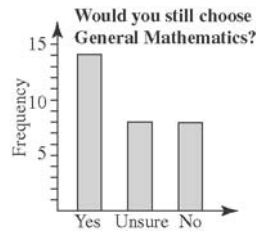
7



8 a

Category	Frequency
Yes	14
No	8
Unsure	8

b



c 14 students thought they would still choose to study General Mathematics.

$$\begin{aligned} \text{d Relative frequency} &= \frac{14}{30} \\ &= \frac{7}{15} \end{aligned}$$

e Percentage frequency

$$\begin{aligned} &= \frac{7}{15} \times 100\% \\ &= 46.67\% \end{aligned}$$

f Categorical because of qualitative answers, and then ordinal because it can be ordered.

9 Total frequency = 36

% students prefer a type of music other than rock

$$= \frac{6 + 2 + 5 + 8}{36} \times 100\%$$

$$= \frac{21}{36} \times 100\%$$

$$= 58.33\%$$

$\therefore$ answer is **C**.

10 Reading values off graph.

a Married (15 out of 40)

b 15 is the frequency of responses marked married

$$\text{c Relative frequency is } \frac{15}{40} = \frac{3}{8}$$

d Percentage frequency

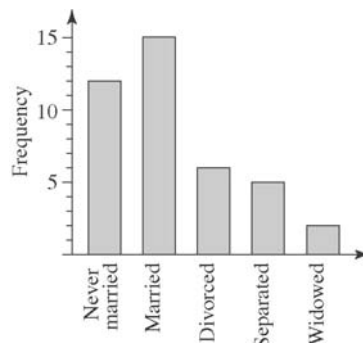
$$\begin{aligned} &= \frac{3}{8} \times 100\% \\ &= 37.5\% \end{aligned}$$

e Respondents unmarried

$$\begin{aligned} &= 12 + 6 + 5 + 2 \\ &= 25 \end{aligned}$$

f Percentage frequency

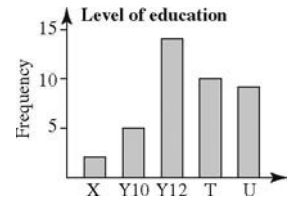
$$\begin{aligned} &= \frac{25}{40} \times 100\% \\ &= 62.5\% \end{aligned}$$



11 a

Level of Education	Frequency
X	2
Y10	5
Y12	14
T	10
U	9

b



c Education to Yr 12 = 14

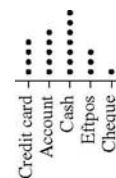
d Education continued after Yr 12 = 10 + 9 = 19.

e Percentage frequency after Yr 12

$$\begin{aligned} &= \frac{19}{40} \times 100\% \\ &= 47.5\% \end{aligned}$$

f Categorical because has qualitative answers and ordinal because categories can be placed in a definite order.

12 a



b Transactions by cash = 7

c % frequency cash transactions

$$\begin{aligned} &= \frac{7}{20} \times 100\% \\ &= 35\% \end{aligned}$$

d Categorical because has qualitative answers and nominal because no order is attached to the categories.

13

Category	Frequency
Nasty	2
Drinkable	4
Nice	7
Delicious	6
Sensational	1
	$\Sigma f = 20$

$\therefore$ answer is **C**.

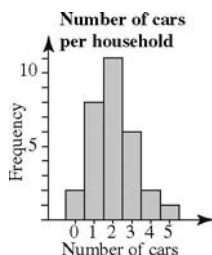
14 a



- b Highest frequency = 16
 $\therefore$ Majority opinion was good
 c Score of performance
 $= 1 \times 2 + 2 \times 4 + 3 \times 16 + 4 \times 8$
 $= 2 + 8 + 48 + 32$
 $= 90$

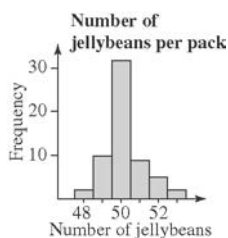
Exercise 1B — Numerical data

1 a



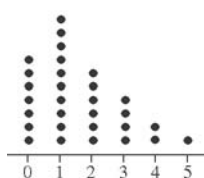
- b Frequency of 3 cars = 6
 c Relative frequency = $\frac{6}{30}$
 $= \frac{1}{5}$
 d Percentage frequency
 $= \frac{1}{5} \times 100\%$
 $= 20\%$

2 a



- b Frequency of a packet with 50 beans = 32
 c Relative frequency = $\frac{32}{60}$
 $= \frac{8}{15}$
 d Percentage frequency
 $= \frac{8}{15} \times 100\%$
 $\approx 53.33\%$

3 a

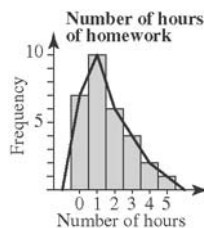


b

Number of hours of study	Frequency
0	7
1	10
2	6
3	4
4	2
5	1
	$\Sigma f = 30$

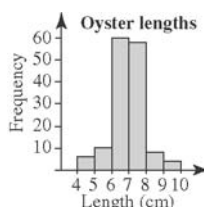
Highest frequency = 10

c and d



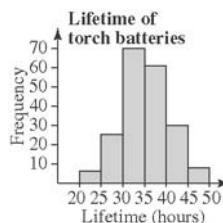
- e $\therefore$ Most common number of hours of study = 1 hour
 f Frequency of 2 hours of study = 6
 g Relative frequency (2 hours)
 $= \frac{6}{30}$
 $= \frac{1}{5}$
 h % frequency (2 hours)
 $= \frac{1}{5} \times 100\%$
 $= 20\%$

4 a



- b Oysters: 5 – 6 cm
 Frequency = 10.
 Relative frequency = $\frac{10}{146}$
 $= \frac{5}{73}$
 % frequency = $\frac{5}{73} \times 100\%$
 $\approx 6.85\%$

5 a



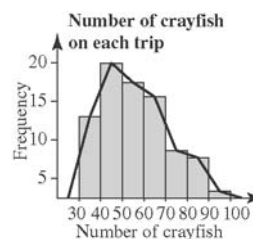
- b Batteries that lasted between 30 – 35 hours

Frequency = 70

$$\text{Relative frequency} = \frac{70}{200} = \frac{7}{20}$$

$$\begin{aligned} \text{\% frequency} &= \frac{7}{20} \times 100\% \\ &= 35\% \end{aligned}$$

6 a and b

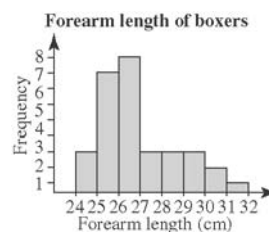


- c As the higher scores are spread over a greater range the data is 'positively skewed'.

7 a

Forearm length (cm)	Frequency
24.0 –	3
25.0 –	7
26.0 –	8
27.0 –	3
28.0 –	3
29.0 –	3
30.0 –	2
31.0 –	1

b

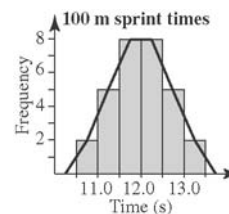


- c As the higher scores are spread over a greater range the data is 'positively skewed'.

8 a

Time in seconds	Frequency
10.5 –	2
11.0 –	5
11.5 –	8
12.0 –	8
12.5 –	5
13.0 –	2

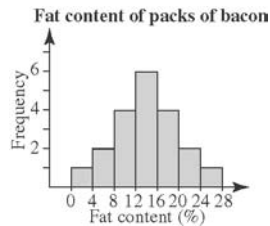
b and c



- d Distribution is symmetric around the central score, therefore the data is normally distributed.

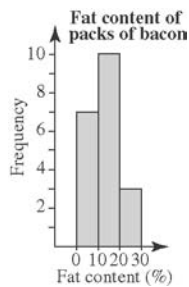
9 a

Fat content (%)	Frequency
0 – 3	1
4 – 7	2
8 – 11	4
12 – 15	6
16 – 19	4
20 – 23	2
24 – 27	1

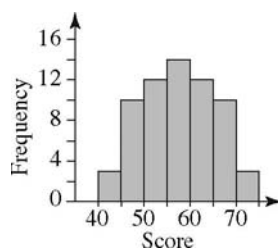


b

Fat content (%)	Frequency
0 – 9	7
10 – 19	10
20 – 29	3



- c The data is more clustered if a large group (class) size is used.
- 10 There is more than one score that is most frequent, therefore this is Bimodal Data
∴ answer is E.
- 11 Most of the data is confined to a small range 80 – 90, therefore the data is clustered.
∴ answer is C.
- 12 The most common scores occur towards the end of the distribution, with fewer scores at the start of the distribution.
∴ answer is C.
- 13



- A. True — most common score is in the centre and distribution is symmetrical around the central score
- B. True — there is an interval of 5 for the class sizes.
- C. True — the highest frequency is 14, which belongs to the class 55 – 60
- D. False. % frequency of 40 – 45 class

$$= \frac{3}{64} \times 100\%$$

$$= 4.69\% \text{ (not 3\%)}$$
- E. True:
 Sum of frequencies

$$= 3 + 10 + 12 + 14 + 12 + 10 + 3$$

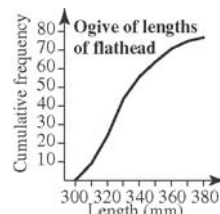
$$= 64$$
 ∴ answer is D.

Exercise 1C — Cumulative data

1 a

Length of fish (mm)	Frequency	Cumulative frequency
300 – 310	9	9
310 – 320	15	24
320 – 330	20	44
330 – 340	12	56
340 – 350	8	64
350 – 360	7	71
360 – 370	4	75
370 – 380	2	77

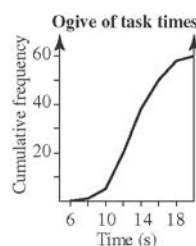
b



2 a

Time taken (sec)	Frequency	Cumulative frequency
6 – 8	1	1
8 – 10	4	5
10 – 12	15	20
12 – 14	18	38
14 – 16	12	50
16 – 18	8	58
18 – 20	2	60

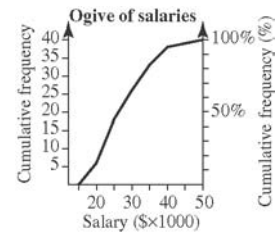
b



3 a

Salary (\$1000)	Frequency	Cumulative frequency
15 –	6	6
20 –	12	18
25 –	8	26
30 –	7	33
35 –	5	38
40 –	1	39
45 –	1	40

b

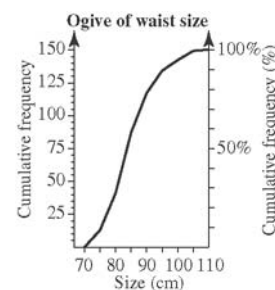


- c 11 employees are earning less than \$22 000.
- d 27 employees are earning less than \$31 000.
- e 75% of the employees earn \$33 000 or less.
- f 50% of the employees earn \$26 500 or less.
- g 25% of the employees earn \$21 500 or less.
- h An employee would have to earn \$38 000 or more to receive a pay rise.

4 a

Waist size (cm)	Frequency	Cumulative frequency
70 –	13	13
75 –	28	41
80 –	46	87
85 –	30	117
90 –	17	134
95 –	8	142
100 –	7	149
105 –	1	150

b



- c 60 young people have a waist size of less than 82 cm.
- d $150 - 130 = 20$ ∴ 20 young people have a waist size greater than 94 cm.
- e 90% of the people have a waist size less than 96 cm.
- f 50% of the people have a waist size less 84 cm.

- g Number of people with waist size between 78 cm and 100 cm
 $= 142 - 27$
 $= 115$
 $\therefore$ % of people not catered for
 $= \frac{(150 - 115)}{150} \times 100\% \approx 23\%$

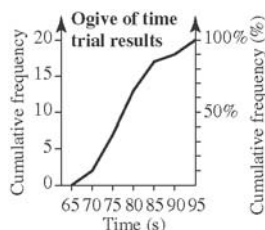
5 Reading off graph:

- a 30 or fewer seeds = 12 pumpkins.
b More than 50 seeds
 $= 11$ pumpkins.
c Fewer than 45 seeds $\approx 68\%$ of the pumpkins.
d Fewer than 20 seeds $\approx 8\%$ of the pumpkins.
e 90% the data lies below 55 seeds.
f 75% the data lies below 46 seeds.
g Half or 50% of the data lies below 37 seeds.
h 20% of pumpkins kept aside have 30 or less seeds.

6 a

Time	Tally	Frequency	Cumulative frequency
65 –		2	2
70 –		5	7
75 –		6	13
80 –		4	17
85 –		1	18
90 –		2	20

b

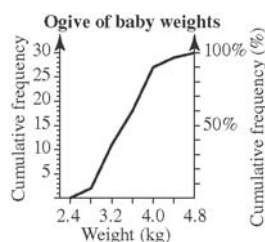


- c 14 riders finished with a time 82 seconds or better.
d 90% of the riders finished with a time 90 seconds or less.
e To qualify you need a time under 72 seconds.

7 a

Weight	Tally	Frequency	Cumulative frequency
2.4 –		2	2
2.8 –		9	11
3.2 –		7	18
3.6 –		9	27
4.0 –		2	29
4.4 –		1	30

b

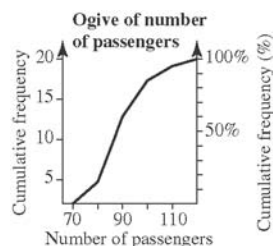


- c 75% of babies had a birth weight of 3.8 kg or less.
d 50% of the babies had a birth weight of 3.45 kg or less.
e 3% of the babies are given special attention because their weight is 2.6 kg or less.

8 a

Number of passengers	Frequency	Cumulative frequency
70 –	3	3
80 –	9	12
90 –	5	17
100 –	2	19
110 –	1	20

b



- c 13 flights had less than 95 people aboard.
d 75% of the flights had 97 passengers or less
e 50% of the flights had 89 passengers or less
f 25% of the flights had 81 passengers or less.
g 3 out of 20 flights had less than 80 passengers:

$$\frac{3}{20} \times 100\% = 15\%.$$

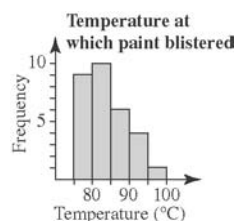
15%, is less than 20%

$\therefore$ A smaller aircraft should not be used.

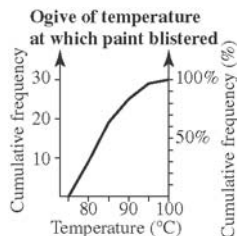
9 a

Temperature	Frequency	Cumulative frequency
75 –	9	9
80 –	10	19
85 –	6	25
90 –	4	29
95 –	1	30

b



c



- d 19 paint samples blistered at 85° or less.
e 63% of the samples blistered at 85° or less.
f 74% of the sample cannot withstand a temperature of 88°
 $(100\% - 74\% = 26\%)$
 $\therefore$ 26% of the sample could withstand a temperature of 88°.
g 90% of the samples blistered by the time the temperature reached 93°.

h The highest temperature samples can take before 20% of them blister is 78° .

10 D is untrue because the graph can be interpreted without the original data 90th percentile $\Leftrightarrow$ 90% of the data.

11 $\therefore$ 90% of the data are less than or equal to 25.
 $\therefore$ answer is C.

Exercise 1D — Measures of central tendency

1 a Mean = $\frac{6+8+4+7+6+7+6}{7}$
 $= 6.286$

4 6 6 6 7 7 8
 Median = 6
 6 occurs 3 times
 $\therefore$ Mode = 6.

b Mean
 $= \frac{102+99+106+108+101+103+102+105+102+101}{10}$
 $= 102.9$
 99, 101, 101, 102, 102, 102, 103, 105, 106, 108
 Median = $\frac{102+102}{2}$
 $= 102$
 102 occurs 3 times
 $\therefore$ Mode = 102.

c Mean
 $= \frac{5.6+5.2+5.4+5.3+5.8+5.4+5.3+5.4}{8}$
 $= 5.425$
 5.2, 5.3, 5.3, 5.4, 5.4, 5.4, 5.6, 5.8
 Median = $\frac{5.4+5.4}{2}$
 $= 5.4$
 5.4 occurs 3 times
 $\therefore$ Mode = 5.4

d Mean
 $= \frac{25+27+29+27+26+28+29+28+27+28}{10}$
 $= 27.4$
 25, 26, 27, 27, 27, 28, 28, 28, 29, 29
 Median = $\frac{27+28}{2}$
 $= 27.5$
 28 and 27 both occurs 3 times
 $\therefore$ Mode = 27 and 28

2

Number of errors (x)	f	$x \times f$	Cumulative frequency
0	9	0	9
1	18	18	27
2	13	26	40
3	6	18	46
4	3	12	49
5	1	5	50
Total	50	79	

a Mean = $\frac{79}{50}$
 $= 1.58$

b Median = $\frac{n+1}{2}$
 $= \frac{51}{2}$

$= 25.5$

Median is halfway between 25th and 26th score
 $\therefore$ Median = 1

c Highest frequency = 18
 $\therefore$ Mode = 1

3

Days of Sickness (x)	f	Midpoint $x \times f$	Cumulative frequency
0 – 4	10	$(2.5 \times 10) = 25$	10
5 – 9	10	$(7.5 \times 10) = 75$	20
10 – 14	7	$(12.5 \times 7) = 87.5$	27
15 – 19	6	$(17.5 \times 6) = 105$	33
20 – 24	5	$(22.5 \times 5) = 112.5$	38
25 – 29	2	$(27.5 \times 2) = 55$	40
30 – 34	1	$(32.5 \times 1) = 32.5$	41
Total	41	492.5	

a Mean = $\frac{492.5}{41}$
 $= 12.01$ days

b Median = $\frac{41+1}{2}$
 $= 21^{\text{st}}$ score
 Median lies in 10 – 14 class interval
 $\therefore$ Median class = 10 – 14.

c Highest frequency = 10
 $\therefore$ Modal classes are 0 – 4 and 5 – 9.

4

Depth (cm)	Frequency	Midpoint (x)	$x \times f$	cf
0 – <50	8	25	200	8
50 – <100	9	75	675	17
100 – <150	12	125	1500	29
150 – <200	15	175	2625	44
200 – <250	6	225	1350	50
250 – <300	4	275	1100	54
300 – <350	2	325	650	56
350 – <400	2	375	750	58
	58		8850	

a Mean = $\frac{8850}{58}$
 ≈ 152.59 cm

b Median = $\frac{58+1}{2}$
 $= 29.5$
 $\therefore$ Median lies between 29th and 30th scores
 $\therefore$ Median class = 150 – <200 cm

c Highest frequency = 15
 $\therefore$ Modal class = 150 – <200 cm

5 To find the mean: Total the scores, then divide by the number of scores.

$\therefore$ answer is B.

6 Seven scores:

$\therefore$ Median = $\frac{7+1}{2}$
 $= 4^{\text{th}}$ score

$\therefore$ Position of median is the 4th score
 $\therefore$ answer is B.

7 Highest frequency of scores = 3

1 and 5 have a frequency of 3

$\therefore$ Modes = 1 and 5

$\therefore$ answer is E.

8 a

Number of hours	Frequency	Midpoint (x)	$x \times f$	cf
0 –	16	2.5	40	16
5 –	6	7.5	45	22
10 –	4	12.5	50	26
15 –	2	17.5	35	28
20 –	1	22.5	22.5	29
25 –	1	27.5	27.5	30
	30		220	

$$\text{b Mean} = \frac{220}{30}$$

$$\approx 7.33$$

$$\text{c Median} = \frac{30+1}{2}$$

$$= 15.5$$

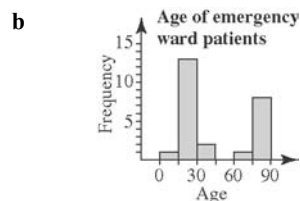
Median lies halfway between the 15th and 16th score

$\therefore$ Median class = 0–5 hours.

d Highest frequency = 16
 $\therefore$ Modal class = 0–5 hours

9

Age	Frequency	Midpoint (x)	$x \times f$	cf
0 –	1	7.5	7.5	1
15 –	13	22.5	292.5	14
30 –	2	37.5	75	16
45 –	0	52.5	0	16
60 –	1	67.5	67.5	17
75 –	8	82.5	660	25
	25		1102.5	



c Two frequent scores $\Rightarrow$ bimodal data.

$$\text{d Mean} = \frac{1102.5}{25}$$

$$= 44.1$$

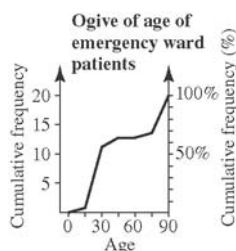
$$\text{e Median} = \frac{25+1}{2}$$

$$= 13^{\text{th}} \text{ score}$$

$\therefore$ Median class = 16 – 30 years old.

f Highest frequency = 13
 $\therefore$ Modal class is 16 – 30 years old.

g



h Median age 30 years from graph.

i No, the mean has one age group in the 31 – 45 class while the median/mode have one age group of 16 – 30

j Answers will vary. Check with teacher (e.g. Tendency for certain age groups to need emergency treatment i.e.

16 – 30 yr. old could be involved in car accidents more than other age groups.

10 a Player A mean:

$$= \frac{30 + 34 + 42 + 28 + 30 + 41}{6}$$

$$= \frac{206}{6}$$

$$\approx 34.33$$

Player B mean:

$$= \frac{0 + 0 + 1 + 0 + 250 + 0}{6}$$

$$= \frac{251}{6}$$

$$\approx 41.83$$

b Player B seems better on mean scores.

c Player A median:

28 30 31 34 41 42

$$\therefore \text{Median} = \frac{31 + 34}{2} = 32.5.$$

Player B median:

0 0 0 0 1 250

 $\therefore$ Median = 0

d Player A is better.

e Mean result is affected by extreme values so can be misleading, hence player A would be better to have in a cricket team.

11 a

Resting pulse rate	Frequency	Midpoint (x)	$x \times f$	cf
30 –	3	35	105	3
40 –	8	45	360	11
50 –	5	55	275	16
60 –	3	65	195	19
70 –	1	75	75	20
	20		1010	

$$\text{b Mean} = \frac{1010}{20}$$

$$= 50.5 \text{ seconds}$$

$$\text{c Median} = \frac{20+1}{2}$$

$$= 11 \frac{1}{2} \text{ score}$$

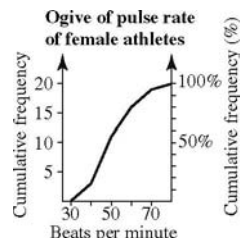
Median lies between the 11th and 12th score

$\therefore$ Median class = 40 – 50 sec

d Highest frequency = 8

Modal class = 40 – 50 sec

e



f From graph, median is 49 sec.

12 The mode is the score with the highest frequency, therefore the answer is A.

13 The median of 21 scores is 11th score, which is 2. Therefore the answer is B.

14 Extreme values affect the mean.

$\therefore$ answer is A.

15 a Mean salary is \$28 875, while the median is only \$24 000. The median is a better indication of central value

because the mean was affected by the large salaries paid to the chief executive officer and managers. So, no they weren't really being honest.

- 16 Check answer with teacher. Use knowledge of measures of central tendency.

a Mean = Median = Mode

{4, 4, 4, 4, 4}.

$$\text{Mean} = \frac{4 + 4 + 4 + 4 + 4}{5}$$

$$= \frac{20}{5}$$

$$= 4$$

$$\text{Median} = 4$$

$$\text{Mode} = 4$$

b Mean > Median > Mode

{20, 10, 8, 2, 2}

$$\text{Mean} = (20 + 10 + 8 + 2 + 2) \div 5$$

$$= \frac{42}{5}$$

$$= 8.4$$

$$\text{Median} = 8$$

$$\text{Mode} = 2$$

$$\therefore 8.4 > 8 > 2$$

c Mean > Median = Mode

{40, 10, 8, 8, 2}.

$$\text{Mean} = \frac{40 + 10 + 8 + 8 + 2}{5}$$

$$= \frac{68}{5}$$

$$= 13.6$$

$$\text{Median} = 8$$

$$\text{Mode} = 8$$

$$\therefore 13.6 > 8 = 8$$

d Median < Mean = Mode

Not possible. Let $a b c d e$ represent 5 numbers in numerical order.

c = median

$d = e$ = mean and mode

Let d and $e = x$

$$\text{Mean} = \frac{a + b + c + x + x}{5}$$

$$x = \frac{a + b + c + 2x}{5}$$

$$5x = a + b + c + 2x$$

$$3x = a + b + c$$

So, the first 3 numbers must add to a number equal to three times the value of the mean (and mode). For example let $x = 10$

$a, b, c, 10, 10$ is the data set $a + b + c$ must add up to 30, to enable the mean to be 10.

The value of c must be less than 10.

So it is not possible to keep c less than 10 and have a, b , and c sum to 30 without displacing the value of c from the median position.

Exercise 1E — Measures of variability

- 1 Range = highest — lowest score

a Range = $7 - 2$

$$= 5$$

b Range = $111 - 102$

$$= 9$$

c Range = $3.4 - 2.4$

$$= 1.0$$

d Range = $8.39 - 1.45$

$$= 6.94$$

- 2 Maximum profit = 3.6%, minimum profit = -5.4% ;
range = $3.6 - (-5.4) = 9\%$. Therefore the answer is E.

3 Range = $95 - 60$

$$= 35$$

4 a 3 4 4 (5) 7 9 10

$$n = 7$$

$$\therefore \text{Median} = \frac{7 + 1}{2}$$

$$= 4^{\text{th}} \text{ position.}$$

3, 4, 4

$$Q_1 = 4$$

7, 9, 10

$$Q_3 = 9$$

$$\therefore \text{IQR} = 9 - 4$$

$$= 5$$

b 17 19 20 25 (25) 27 28 29 29

$$n = 9$$

$$\therefore \text{Median} = \frac{9 + 1}{2}$$

$$= 5^{\text{th}} \text{ position}$$

17 19 20 25

↓

$$Q_1 = 19.5$$

27 28 29 29

↓

$$Q_3 = 28.5$$

$$\therefore \text{IQR} = 28.5 - 19.5$$

$$= 9$$

c 57 52 52 53 53 53 54 54 55 55

↓

$$n = 10$$

$$\therefore \text{Median} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = \frac{53 + 53}{2}$$

$$= 53$$

52 52 (53) 53 53

↓

$$\therefore Q_1 = 52$$

53 54 (54) 55 55

↓

$$\therefore Q_3 = 54$$

$$\therefore \text{IQR} = 54 - 52$$

$$= 2$$

d 12, 12, 14, 14, 15, ↓ 15, 17, 18, 19, 20

$$n = 10$$

$$\therefore \text{Median} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

12, 12, (14) 14, 15

↓

$$Q_1 = 14$$

15 17 (18) 19, 20

↓

$$Q_3 = 18$$

$$\text{IQR} = 18 - 14$$

$$= 4$$

- 5 From graph:

$$Q_3 = 138$$

$$Q_1 = 116$$

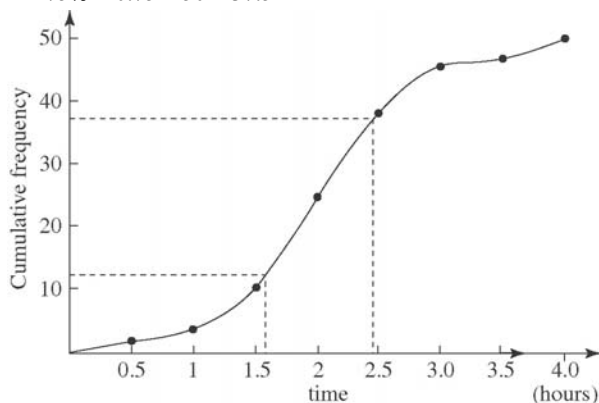
$$\begin{aligned} \text{IQR} &= 138 - 116 \\ &= 22 \end{aligned}$$

6

Time	Frequency	cf
0 –	1	1
0.5 –	2	3
1 –	7	10
1.5 –	15	25
2 –	13	38
2.5 –	8	46
3 –	2	48
3.5 –	2	50

$$25\% = 0.25 \times 50 = 12.5$$

$$75\% = 0.75 \times 50 = 37.5$$



∴ From graph,

$$Q_3 = 2.4$$

$$Q_1 = 1.6$$

$$\begin{aligned} \therefore \text{IQR} &= 2.4 - 1.6 \\ &= 0.8 \end{aligned}$$

7 17 18 18 19 (20) 21 21 23 25

$$n = 9$$

$$\text{Median} = \frac{9+1}{2}$$

$$= 5^{\text{th}} \text{ position}$$

17, 18, 18, 19

↓

$$Q_1 = 18$$

21, 21, 23, 25

↓

$$Q_3 = \frac{21+23}{2} = 22$$

$$\therefore \text{IQR} = 22 - 18 = 4$$

∴ answer is C.

8 a

Scores	Scores – Mean	Square difference
3	–1.625	2.640625
5	0.375	0.140625
8	3.375	11.390625
2	–2.625	6.890625
7	2.375	5.640625
1	–3.625	13.140625
6	1.375	1.890625
5	0.375	0.140625
TOTAL	37	41.875

$$\text{Mean} = 37 \div 8 = 4.625$$

$$\text{Variance} = 41.875 \div 7$$

$$= 5.982143$$

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\text{variance}} \\ &\approx 2.446 \end{aligned}$$

b

Scores	Scores – Mean	Square difference
11	0.571428571	0.326530612
8	–2.428571429	5.897959184
7	–3.428571429	11.75510204
12	1.571428571	2.469387755
10	–0.428571429	0.183673469
11	0.571428571	0.326530612
14	3.571428571	12.75510204
TOTAL	73	33.71428571

$$\text{Mean} = 73 \div 7 = 10.42857$$

$$\text{Variance} = 33.71428571 \div 6$$

$$= 5.6190476$$

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\text{variance}} \\ &\approx 2.370 \end{aligned}$$

c

Scores	Scores – Mean	Square difference
25	–11.375	129.390625
15	–21.375	456.890625
78	41.625	1732.640625
35	–1.375	1.890625
56	19.625	385.140625
41	4.625	21.390625
17	–19.375	375.390625
24	–12.375	153.140625
TOTAL	291	3255.875

$$\text{Mean} = 291 \div 8 = 36.375$$

$$\text{Variance} = 3255.875 \div 7$$

$$= 465.125$$

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\text{variance}} \\ &\approx 21.567 \end{aligned}$$

d

Scores	Scores – Mean	Square difference
5.2	–1.28	1.6384
4.7	–1.78	3.1684
5.1	–1.38	1.9044
12.6	6.12	37.4544
4.8	–1.68	2.8224
TOTAL	32.4	46.988

$$\text{Mean} = 32.4 \div 5 = 6.48$$

$$\text{Variance} = 46.988 \div 4$$

$$= 11.747$$

$$\begin{aligned} \text{Standard deviation} &= \sqrt{\text{variance}} \\ &\approx 3.427 \end{aligned}$$

9 As for Question 8, just completed via inbuilt calculator functions

10 The combined weight of the team = $69 \times 8 = 552$ kg. Therefore the answer is D.

11 Since the weight of each person increases by 1.8 kg, the mean weight will increase by 1.8 kg (since each weight increases by same amount). Therefore the answer is C.

12 a Group A

$$\begin{aligned}\text{Mean} &= \frac{160 + 170 \times 5 + 180}{7} \\ &= \frac{1190}{7} \\ &= 170 \text{ cm}\end{aligned}$$

160, 170, 170, (170), 170, 170, 180

∴ Median = 170 cm

Highest frequency is 5

∴ Mode = 170 cm

Group B

$$\begin{aligned}\text{Mean} &= \frac{110 + 160 + 170 \times 3 + 180 + 230}{7} \\ &= \frac{1190}{7} \\ &= 170 \text{ cm}\end{aligned}$$

110, 160, 170, (170), 170, 180, 230

∴ Median = 170 cm

Highest frequency is 3

∴ Mode = 170 cm

∴ All measures of central tendency (mean, median, mode) are the same for both groups.

b Mean, median and mode are misleading on similarity between groups.

c Group B

d Group B

e Group B

f Group A

$$\begin{aligned}\text{Range} &= 180 - 160 \\ &= 20 \text{ cm}\end{aligned}$$

160, 170, 170, (170), 170, 170, 180
 $\downarrow \quad \quad \downarrow \quad \quad \downarrow$
 $Q_1 \quad \quad \text{Median} \quad \quad Q_3$

$$\begin{aligned}\therefore \text{IQR} &= 170 - 170 \\ &= 0\end{aligned}$$

	Scores	Scores – Mean	Square difference
	160	–10	100
	170	0	0
	170	0	0
	170	0	0
	170	0	0
	170	0	0
	180	10	100
TOTAL	1190		200

$$\text{Mean} = 1190 \div 7 = 170$$

$$\text{Variance} = 200 \div 6 = 33.33$$

$$\begin{aligned}\text{Standard deviation} &= \sqrt{\text{variance}} \\ &\approx 5.774\end{aligned}$$

Group B

$$\begin{aligned}\text{Range} &= 230 - 110 \\ &= 120 \text{ cm}\end{aligned}$$

110, (160), 170, (170), 170, (180), 230
 $\downarrow \quad \quad \downarrow \quad \quad \downarrow$
 $Q_1 \quad \quad \text{Median} \quad \quad Q_3$

$$\begin{aligned}\therefore \text{IQR} &= 180 - 160 \\ &= 20 \text{ cm.}\end{aligned}$$

	Scores	Scores – Mean	Square difference
	160	–10	100
	170	0	0
	170	0	0
	110	–60	3600
	230	60	3600
	170	0	0
	180	10	100
TOTAL	1190		7400

$$\text{Mean} = 1190 \div 7 = 170$$

$$\text{Variance} = 7400 \div 6$$

$$= 1233.33$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

$$\approx 35.119$$

13 The interquartile range is not affected by extreme values so is considered to be a better measure of the variability of a set.

$\therefore$ answer is **E**.

14 a Range = 140 – 80

$$= 60$$

b and **c**

Scores	Midpoint	Frequency	Frequency $\times$ Midpoint	Scores – Mean	Square difference	Square difference $\times$ Frequency
80 –	85	1	85	– 25.48387097	649.4276795	649.4277
90 –	95	4	380	– 15.48387097	239.7502601	959.001
100 –	105	11	1155	– 5.483870968	30.07284079	330.8012
110 –	115	9	1035	4.516129032	20.39542144	183.5588
120 –	125	4	500	14.51612903	210.7180021	842.872
130 –	135	2	270	24.51612903	601.0405827	1202.081
TOTAL		31	3425		1751.404787	4167.742

$$\text{Mean} = 3425 \div 31$$

$$= 110.4839 \approx 110.48$$

$$\text{Variance} = 4167.742 \div 30$$

$$= 138.92473$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

$$\approx 11.79$$

15

Scores	Midpoint	Frequency	Frequency $\times$ Midpoint	Scores – Mean	Square difference	Square difference $\times$ Frequency
10 –	15	1	15	– 19.04761905	362.8117914	362.8118
20 –	25	6	150	– 9.047619048	81.85941043	491.1565
30 –	35	9	315	0.952380952	0.907029478	8.163265
40 –	45	4	180	10.95238095	119.9546485	479.8186
50 –	55	1	55	20.95238095	439.0022676	439.0023
TOTAL		21	715		1004.535147	1780.952

$$\text{Mean} = 715 \div 21$$

$$= 34.04762$$

$$\text{Variance} = 1780.952 \div 20$$

$$= 89.0476$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

$$= 9.437$$

answer is **C**.

- 16 a and b Range = 700 – 200
= 500 hours

Scores	Midpoint	Frequency	Frequency × Midpoint	Scores – Mean	Square difference	Square difference × Frequency
200 –	225	2	450	– 230.2857143	53 031.5102	106 063
250 –	275	5	1375	– 180.2857143	32 502.93878	162 514.7
300 –	325	12	3900	– 130.2857143	16 974.36735	203 692.4
350 –	375	25	9375	– 80.28571429	6445.795918	161 144.9
400 –	425	42	17 850	– 30.28571429	917.2244898	38 253.43
450 –	475	38	18 050	19.71428571	388.6530612	14 768.82
500 –	525	26	13 650	69.71428571	4860.081633	126 362.1
550 –	575	15	8625	119.7142857	14 331.5102	214 972.7
600 –	625	7	4375	169.7142857	28 802.93878	201 620.6
650 –	675	3	2025	219.7142857	48 274.36735	144 823.1
TOTAL		175	79 675		206 529.3878	1 374 486

$$\text{Mean} = 79\,675 \div 175$$

$$= 455.28571$$

$$= 455.29 \text{ hours}$$

$$\text{Variance} = 1\,374\,486 \div 174$$

$$= 7899.344828$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

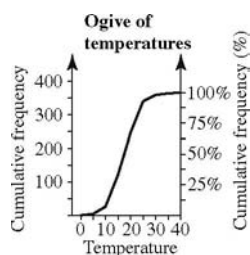
$$= 88.88$$

- 17 If you have the standard deviation you can calculate the variance by squaring the standard deviation
∴ answer is **B**.

- 18 a

Maximum temp	No. of days	cf
0 –	4	4
5 –	22	26
10 –	95	121
15 –	124	245
20 –	94	339
25 –	19	358
30 –	5	363
35 –	2	365

- b



- c Reading off graph:

Q_3 is the 75th percentile

$$Q_3 = 21$$

Q_1 is the 25th percentile

$$Q_1 = 13.5$$

$$\text{Interquartile range} = Q_3 - Q_1$$

$$= 21 - 13.5$$

$$= 7.5$$

- d Reading off graph:

Median is the 50th percentile

∴ Median ≈ 17

Scores	Midpoint	Frequency	Frequency × Midpoint	Scores – Mean	Square difference	Square difference × Frequency
0 –	2.5	4	10	– 15.05479452	226.6468381	906.5874
5 –	7.5	22	165	– 10.05479452	101.0988929	2224.176
10 –	12.5	95	1187.5	– 5.054794521	25.55094764	2427.34
15 –	17.5	124	2170	– 0.054794521	0.003002439	0.372302
20 –	22.5	94	2115	4.945205479	24.45505723	2298.775
25 –	27.5	19	522.5	9.945205479	98.90711203	1879.235
30 –	32.5	5	162.5	14.94520548	223.3591668	1116.796
35 –	37.5	2	75	19.94520548	397.8112216	795.6224
TOTAL		365	6407.5		1097.832239	11 648.9

e Mean = $\frac{6407.5}{365}$

$\approx 17.55^\circ\text{C}$

f $s = \sqrt{\frac{11648.9}{364}}$

≈ 5.657

g $s^2 = \frac{11648.9}{364}$

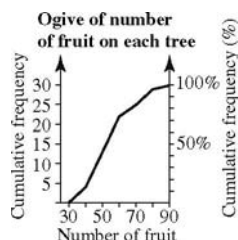
≈ 32.002

h Range = $40 - 0$
 $= 40^\circ\text{C}$

19 a

No. of fruit	Frequency	cf
30 –	4	4
40 –	9	13
50 –	9	22
60 –	3	25
70 –	4	29
80 –	1	30

b



c Reading off graph:

Median is 50th percentile

$\therefore$ Median = 55

Q_1 is 25th percentile

$\therefore Q_1 = 45$

Q_3 is 75th percentile

$\therefore Q_3 = 63$

d IQR = $63 - 45$
 $= 18$

Scores	Midpoint	Frequency	Frequency × Midpoint	Scores – Mean	Square difference	Square difference × Frequency
30 –	35	4	140	– 19	361	1444
40 –	45	9	405	– 9	81	729
50 –	55	9	495	1	1	9
60 –	65	3	195	11	121	363
70 –	75	4	300	21	441	1764
80 –	85	1	85	31	961	961
TOTAL		30	1620		1966	5270

$$\text{e Mean} = \frac{1620}{30} = 54$$

$$\text{f } s = \sqrt{\frac{5270}{30-1}} \approx 13.48$$

$$\text{g Range} = 81 - 36 = 45$$

- 20 a** Electric Mate has longest life as it has a greater value for the mean.
b Electric Mate Mean = 197 hours. Hot Wire Mean = 185 hours.
c Greatest standard deviation $\Rightarrow$ greatest spread would be Electric Mate.
d The standard deviation shows that you could be lucky to get a toaster with an element that will last, but there is also a chance that you will get an element that will not last $\Rightarrow$ overall inconsistent performance by Electric Mate.
e Check with teacher. See example of answer below:
 Hot Wire had less spread of results and a mean of 185 hours. So you can be more sure of your purchase.
 There is an element of risk due to the inconsistency of elements with Electric Mate—higher mean but greater spread of results.
21 Crinkle would be the best buy because the spread of packet weight is closer to the mean than crunch.

Exercise 1F — Stem-and-leaf plots

1 Key 0|6 = 6

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

In order of size: Key 0|6 = 6

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

2 Key 3|6 = 36 min

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

In order of size:
Key 3|6 = 36 min

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

3 Key 8*|6 = 86 w.p.m.

Stem	Leaf
8*	8 6
9	2
9*	6 6 5 5 6 5 8 9
10	2 2 3 2
10*	8 7 7 8 7 8
11	0 1 2
11*	5
12	1 4 0 1
12*	
13	0

In order: Key 8*|6 = 86 w.p.m

Stem	Leaf
8*	6 8
9	2
9*	5 5 5 6 6 6 8 9
10	2 2 2 3
10*	7 7 7 8 8 8
11	0 1 2
11*	5
12	0 1 1 4
12*	
13	0

4 Key 13*|8 = 13.8

Stem	Leaf
13*	8 9
14	0 3 3 2
14*	8 8 7 6 6 7 5
15	2 2 1
15*	7 5 6 9

In order: Key 13*|8 = 13.8

Stem	Leaf
13*	8 9
14	0 2 3 3
14*	5 6 6 7 7 8 8
15	1 2 2
15*	5 6 7 9

5 Median = $\frac{24+1}{2}$
 = 12.5th score
 Halfway between the 12th and 13th score.

$$Q_1 = \frac{12+1}{2} = 6.5^{\text{th}} \text{ score}$$

halfway between 6th and 7th score

$$Q_1 = \frac{249.6 + 249.6}{2} = 249.6$$

Q_3 — halfway between 6th and 7th score in upper section.

$$Q_3 = \frac{251.5 + 251.6}{2}$$

$$= 251.55$$

$$\therefore \text{IQR} = 251.55 - 249.6 = 1.95 \text{ grams}$$

$$\text{6 a Median} = \frac{24+1}{2}$$

$$= 12.5^{\text{th}} \text{ score}$$

or halfway between 12th and 13th score.

$$\therefore \text{Median} = \frac{8.4 + 8.4}{2} = 8.4 \text{ sec.}$$

$$\text{b } Q_1 = \frac{12+1}{2}$$

$$= 6.5^{\text{th}} \text{ score}$$

or halfway between 6th and 7th score in lower section.

$$\therefore Q_1 = \frac{7.7 + 7.9}{2}$$

$$= 7.8 \text{ sec}$$

Q_3 is between 6th and 7th score in upper half.

$$\therefore Q_3 = \frac{8.8 + 8.9}{2}$$

$$= 8.85 \text{ sec}$$

$$\text{c IQR} = 8.85 - 7.8 = 1.05 \text{ sec.}$$

$$\text{7 Class size} = 1250 - 1200 = 50$$

$\therefore$ answer is A.

$$\text{8 Number of scores} = 27$$

$\therefore$ answer is A.

$$\text{9 Median} = \frac{27+1}{2}$$

$$= 14^{\text{th}} \text{ score.}$$

$$\therefore \text{Median} = 1340$$

$\therefore$ answer is B.

$$\text{10 } Q_1 \text{ is } \frac{13+1}{2}$$

$$= 7^{\text{th}} \text{ score}$$

$$\therefore Q_1 = 1290$$

Q_3 is the 7th score in the upper half

$$\therefore Q_3 = 1390$$

$$\therefore \text{IQR} = 1390 - 1290 = 100$$

$\therefore$ answer is D.

$$\text{11 a Key: } 20|2 = 20.2$$

Stem	Leaf
20	2 7
21	4 6 7 8
22	2 8 8 8
23	0 1 2 6 6
24	6 6 7 8
25	1

$$\text{b Median} = \frac{20+1}{2}$$

$$= 10.5^{\text{th}} \text{ score}$$

or halfway between 10th and 11th score.

$$\therefore \text{Median} = \frac{22.8 + 23.0}{2}$$

$$= 22.9 \text{ cm}$$

$$\text{c } Q_1 = \frac{10 + 1}{2}$$

$$= 5 \frac{1}{2}^{\text{th}} \text{ score}$$

or halfway between lower 5th and 6th score.

$$\therefore Q_1 = \frac{21.7 + 21.8}{2}$$

$$= 21.75$$

$$Q_3 = \frac{23.6 + 24.6}{2}$$

$$= 24.1$$

$$\text{d } \text{IQR} = 24.1 - 21.75$$

$$= 2.35 \text{ cm}$$

12 a Key: 8|2 = 82 cm

Stem	Leaf
8	2 4 5 5 8
9	0 1 2 3 4 6 6 7 8 9
10	1 1 3 3 7 8
11	0 0 1 4 5 7
12	6 6
13	2

$$\text{b } \text{Median} = \frac{30 + 1}{2}$$

$$= 15 \frac{1}{2}^{\text{th}} \text{ score}$$

or halfway between 15th and 16th score.

$$\therefore \text{Median} = \frac{99 + 101}{2}$$

$$= 100 \text{ cm}$$

$$\text{c } Q_1 = \frac{15 + 1}{2}$$

$$= 8^{\text{th}} \text{ score}$$

$$\therefore Q_1 = 92 \text{ cm}$$

$$Q_3 = 110 \text{ cm}$$

$$\text{d } \text{IQR} = 110 - 92$$

$$= 18 \text{ cm}$$

Exercise 1G — Boxplots

1 6 11 (13) 16 32

$$\text{a } \text{Median} = \frac{5 + 1}{2}$$

$$= 3^{\text{rd}} \text{ score}$$

$$\therefore \text{Median} = 13$$

$$\text{b } Q_1 = 11$$

$$Q_3 = 16$$

$$\therefore \text{IQR} = 16 - 11 = 5$$

$$\text{c } \text{Range} = 32 - 6$$

$$= 26$$

2 101, 119, (122), 125, 128

$$\text{a } \text{Median} = \frac{5 + 1}{2}$$

$$= 3^{\text{rd}} \text{ score}$$

$$\therefore \text{Median} = 122$$

$$\text{b } Q_1 = 119$$

$$Q_3 = 125$$

$$\therefore \text{IQR} = 125 - 119$$

$$= 6$$

$$\text{c } \text{Range} = 128 - 101$$

$$= 27$$

3 39.2, 46.5, (49.0), 52.3, 57.8

$$\text{a } \text{Median} = \frac{5 + 1}{2}$$

$$= 3^{\text{rd}} \text{ score}$$

$$\therefore \text{Median} = 49.0$$

$$\text{b } Q_1 = 46.5$$

$$Q_3 = 52.3$$

$$\text{IQR} = 52.3 - 46.5$$

$$= 5.8$$

$$\text{c } \text{Range} = 57.8 - 39.2$$

$$= 18.6$$

4 Reading off the boxplot:

a Greatest score = 140 points

b Least points score = 56 points

c Median points score = 90 points

$$\text{d } \text{Range} = 140 - 56$$

$$= 84 \text{ points}$$

$$\text{e } Q_1 = 76$$

$$Q_3 = 102$$

$$\therefore \text{IQR} = 102 - 76$$

$$= 26 \text{ points}$$

5 Reading off the boxplot:

a Largest number of honey bears = 58;

b Smallest number of honey bears = 31

c Median number of honey bears = 43

$$\text{d } \text{Range} = 58 - 31$$

$$= 27$$

$$\text{e } Q_1 = 39$$

$$Q_3 = 46$$

$$\therefore \text{IQR} = 46 - 39$$

$$= 7$$

Reading off the boxplot 6 to 8.

6 Median = 23

$\therefore$ answer is C.

7 $Q_1 = 20$, $Q_3 = 25$

$$\therefore \text{IQR} = 25 - 20$$

$$= 5$$

$\therefore$ answer is C.

8 Most of the data are contained between the scores of 5 and 20
No – only lower 25% of the scores are contained between 5 and 20.
 $\therefore$ answer is D.

9 a 22, 25, (28), 31, 33, 37, 43, (43), 45, 48

$$\text{Median} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

$$Q_1 = \frac{5 + 1}{2}$$

$$= 3^{\text{rd}} \text{ score}$$

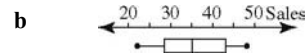
$$\therefore Q_1 = 28$$

$$Q_3 = 43$$

$$\text{min} = 32$$

$$\text{max} = 48$$

$\therefore$ five-number summary
22, 28, 35, 43, 48



10 a 10 11 12 15 21 22 22 23 30 37 39 45

$$\text{Median} = \frac{12 + 1}{2}$$

$$= 6.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 22$$

$$Q_1 = \frac{6 + 1}{2}$$

$$= 3 \frac{1}{2}^{\text{th}} \text{ score}$$

$$\therefore Q_1 = \frac{12 + 15}{2}$$

$$= 13.5$$

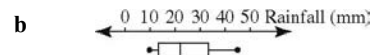
$$Q_3 = \frac{30 + 37}{2}$$

$$= 33.5$$

$$\text{max} = 45$$

$$\text{min} = 10$$

$\therefore$ five-number summary
10, 13.5, 22, 33.5, 45.



11 a From stemplot:

$$\text{Median} = \frac{25 + 1}{2}$$

$$= 13^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 26 \text{ yrs}$$

$$Q_1 = \frac{12 + 1}{2}$$

$$= 6.5^{\text{th}} \text{ score}$$

$$\therefore Q_1 = 20$$

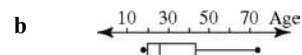
$$Q_3 = \frac{42 + 45}{2}$$

$$= 43.5$$

$$\text{min} = 18$$

$$\text{max} = 74$$

$\therefore$ five-number summary
18, 20, 26, 43.5, 74



c The data is positively skewed, and includes one extremely high value.

12 a From stemplot

$$\text{Median} = \frac{30 + 1}{2}$$

$$= 15.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = \frac{147\,000 + 149\,000}{2}$$

$$= 148\,000$$

$$Q_1 = \frac{15 + 1}{2}$$

$$= 8^{\text{th}} \text{ score}$$

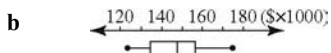
$$\therefore Q_1 = \$135\,000$$

$$Q_3 = \$157\,000$$

$$\text{min} = 124\,000$$

$$\text{max} = 175\,000$$

$\therefore$ five-number summary
124 000 135 000 148 000
157 000 175 000



13 a Key: 12|1 = 121

Stem	Leaf
12	1 5 6 9
13	1 2 4
14	3 4 8 8
15	0 2 2 2 5 7
16	3 5
17	2 9
18	1 1 1 2 3 7 8

$$\text{Median} = \frac{28 + 1}{2}$$

$$= 14.5^{\text{th}} \text{ score}$$

$\therefore$ Median = 152

$$Q_1 = \frac{14 + 1}{2}$$

$$= 7.5^{\text{th}} \text{ score}$$

$$\therefore Q_1 = \frac{134 + 143}{2}$$

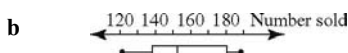
$$= 138.5$$

$$Q_3 = \frac{179 + 181}{2}$$

$$= 180$$

min = 121

max = 188



14 a Key: 1*|7 = 17 years

Stem	Leaf
1*	7 7 8 8 8 9 9
2	0 0 0 1 2 2 2 2 3 3 3 3 4 4 4
2*	5 5 8 9
3	1 2 3
3*	
4	
4*	8

b

$$\text{Median} = \frac{30 + 1}{2}$$

$$= 15.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = \frac{22 + 23}{2}$$

$$= 22.5$$

$$Q_1 = \frac{15 + 1}{2}$$

$$Q_1 = 8^{\text{th}} \text{ score}$$

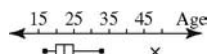
$$\therefore Q_1 = 20$$

$$Q_3 = 25$$

$$\text{Min} = 17$$

$$\text{Max}(\text{without extreme value}) = 33$$

$$(\text{Extreme value} = 48.)$$



c Data is positively skewed with one high extreme value. Most mothers have their first baby between the ages of 20 and 25.

15 Small interquartile range, long whiskers $\Rightarrow$ clustered data
 $\therefore$ answer is C.

Exercise 1H — Comparing sets of data

1 Reading off boxplot:

a Brand A has the larger maximum value.

$\therefore$ Brand A could stand highest voltage.

b Brand A has the smallest minimum value.

$\therefore$ Brand A could also withstand the least voltage.

c Brand B has a greater median value.

$\therefore$ Brand B had the better median performance.

d Brand A had the greatest range.

e Brand A also had the greatest IQR.

f Brand B would be recommended because overall it was more consistent, smaller range and IQR.

g If they wanted higher voltage they would have to take their chance with Brand A, as it can withstand greater voltage than Brand B.

2 Reading off boxplot:

a Text has the greatest maximum value

$\therefore$ Using the Text book group had the highest (best) result.

b Again the Text book group had a higher median.

c Range of Text book group was larger than no-text book groups. 20 was the IQR, suggesting that the Text book group had more variable results.

d Dependent also upon the individual as the scores in the Text book group varied a lot. The top 25% results in the Text book group were better than the new text book group.

e Individual difference — have the students completed the work or are they trying to learn on the spot.

3 Group X

Min. = 14

Max. = 33

14, (18.5), (21.5), (27.5), 33

Q_1 Median Q_3

Median = 21.5

$Q_1 = 18.5$, $Q_3 = 27.5$

Group Y

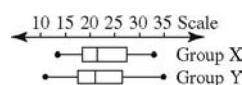
Min. = 11

Max. = 35

11, (17.5), (21), (26.5), 35

Q_1 Median Q_3

$Q_1 = 17.5$, $Q_3 = 26.5$



4 a Men

Min. = 17

Max. = 40

$$\text{Median} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = \frac{24 + 28}{2}$$

$$= \frac{55}{2}$$

$$= 27.5$$

$$\text{Quartile} = \frac{5 + 1}{2}$$

$$= 3^{\text{rd}} \text{ score}$$

$$\therefore Q_1 = 21$$

$$Q_3 = 33$$

Women

Min. = 18

Max. = 31

Median = 5.5th score

$$\therefore \text{Median} = \frac{23 + 24}{2}$$

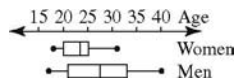
$$= \frac{47}{2}$$

$$= 23.5$$

Quartile = 3rd score

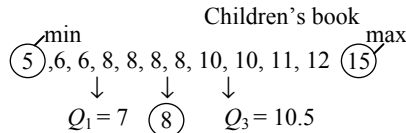
$$\therefore Q_1 = 20$$

$$Q_3 = 25$$



- b Women had a smaller range of 13, compared to men with a range of 23. IQR for women was 5, while men's was 12. The median age was smaller for women (23.5) compared to men (27.5). $\therefore$ less variability in age of marriage for women.

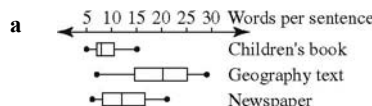
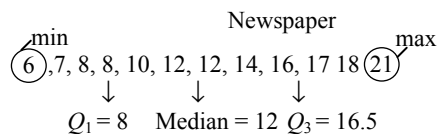
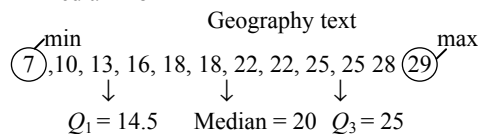
5



$$\text{Median} = \frac{12 + 1}{2}$$

$$= 6.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 8$$



- b The largest sentence was in the Geography text book of 29 words, then the newspaper 21 words, finally the children's book (15 words). The shortest sentence went children (5), Newspaper (6) and the geography text (7). The variability of results was greatest and the geography text, (range = 22, IQR = $10\frac{1}{2}$). Overall the geography text also had a large central value (median of 20) than the others, children (8) and newspaper (12).

6 a **Jones**

Max. = 61

Min. = 3

$$\text{Median} = \frac{20 + 1}{2}$$

$$= 10.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 39$$

$$\text{Quartiles} = \frac{10 + 1}{2}$$

$$= \frac{11}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

$$\therefore Q_1 = \frac{28 + 35}{2}$$

$$= 31.5$$

$$Q_3 = \frac{44 + 48}{2}$$

$$= \frac{92}{2}$$

$$= 46$$

 $\therefore$ five-number summary is:

{3, 31.5, 39, 46, 61}.

Smith

Max. = 85

Min. = 0

$$\text{Median} = \frac{57 + 58}{2}$$

$$= \frac{115}{2}$$

$$= 57.5$$

$$Q_1 = \frac{16 + 19}{2}$$

$$= \frac{35}{2}$$

$$= 17.5$$

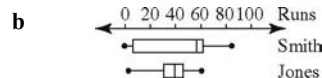
$$Q_3 = \frac{60 + 64}{2}$$

$$= \frac{124}{2}$$

$$= 62$$

 $\therefore$ five-number summary is:

{0, 17.5, 57.5, 62, 85}.



c and d Smith's range = 85

$$\text{IQR} = 44\frac{1}{2}$$

$$\text{Jones' range} = 58$$

$$\text{IQR} = 14\frac{1}{2}$$

Jones' results are more clustered than Smith's results indicating less variability in scores obtained. Both players have obtained large ranges overall. Jones' consistency may indicate that he is better as an opening batsman, while a big hitter Smith might be better at the end.

7 19 scores

$$\text{Median} = \frac{19 + 1}{2}$$

$$= 10^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 165$$

$$\therefore Q_1 = \frac{9 + 1}{2}$$

$$= 5^{\text{th}} \text{ score}$$

$$\therefore Q_1 = 155$$

$$\therefore \text{answer is C.}$$

8 A. Group A

looks normally distributed.

 $\therefore$ True

B. Group B

is skewed.

 $\therefore$ True.

C. Group A median = 155.

 $\therefore$ Group B bigger $\therefore$ True.

D. Group A more variability.

$$\begin{aligned}\text{Range} &= 181 - 122 \\ &= 59.\end{aligned}$$

$$\text{IQR} = 164 - 144 = 20$$

Group B

$$\text{Range} = 181 - 146 = 35$$

$$\text{IQR} = 174 - 155 = 19.$$

∴ True

∴ All statements true.

∴ answer is E.

9 Group X has a lower median – '25' compared to Group Y (≈ 26/27) and more variability because greater range and IQR.

∴ answer is B.

10 Interquartile range of Group X is $30 - 22 = 8$ not 25.

∴ answer is C.

11 a $n = 25$

$$\text{Median} = \frac{25 + 1}{2}$$

$$= 13^{\text{th}} \text{ score}$$

$$\text{Quartiles} = \frac{12 + 1}{2}$$

$$= 6.5^{\text{th}} \text{ score.}$$

Before adjustment:

$$\text{Median} = 21.1$$

$$Q_1 = \frac{20.7 + 20.8}{2}$$

$$= 20.75$$

$$Q_3 = \frac{21.4 + 21.5}{2}$$

$$= 21.45$$

$$\text{Min.} = 20.3, \text{Max.} = 21.8$$

∴ five-number summary is

{20.3, 20.75, 21.1, 21.45, 21.8}

After adjustment:

$$\text{Median} = 20.1$$

$$Q_1 = 19.7$$

$$Q_3 = \frac{20.9 + 21.0}{2}$$

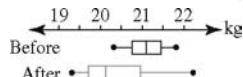
$$= 20.95$$

$$\text{Min.} = 19.3, \text{Max.} = 22.2$$

∴ five-number summary is

{19.3, 19.7, 20.1, 20.95, 22.2}.

b



c If 20.0 kg is the correct amount then after adjusting the median value is closer to 20.1 kg., however the consistency decreased because the results became more variable. Higher IQR 1.25 kg versus 0.7 kg and overall range 2.9 versus 1.5.

d The machine should be adjusted. Before the adjustment the machine was always above the expected weight. After the adjustment, the majority of weights were closer to 20 kg.

12 Key: 2|1 = 21 fruit

a

Unsprayed Leaf	Stem	Sprayed Leaf
	2	1 7 8 9 9
9 9 8 7 5	3	2 3 4 4 5 8
9 7 5 4 3 2 1 0	4	2 2 5 8
8 6 2 1 0	5	4 9
8 8 5 3 2 2	6	1 2 2 6
1	7	1 3 6
	8	5

b Sprayed data:

$$\text{Min} = 21$$

$$Q_1 = 32.5$$

$$\text{Median} = 42$$

$$Q_3 = 62$$

$$\text{Max} = 85$$

Unsprayed data:

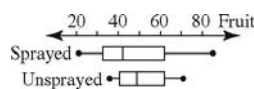
$$\text{Min} = 35$$

$$Q_1 = 40.5$$

$$\text{Median} = 49$$

$$Q_3 = 62$$

$$\text{Max} = 71$$



c General comparison.

Sprayed fruit has a greater range; $85 - 21 = 64$ and $\text{IQR} = 29.5$, showing variability in the results. The median is lower. The unsprayed fruit has a more consistent result, with a smaller range; $71 - 35 = 36$ and $\text{IQR} = 21.5$. The median is higher.

d Although the sprayed trees have some higher yields, the spraying has also resulted in some lower yields and a greater variability in results. It would probably be better for the orchardist not to spray as the smaller spread provides better 'control' over the size of the harvest.

13 a Key: 2|13 = 23 hours

Starlet Leaf	Stem	Glow-worm Leaf
8 6	1*	
4	2	3 3
9 8	2*	7 8
2 1	3	1
6 6 5	3*	5 5 8 9
3	4	1 1 2
7 6	4*	5 7 8 8 9
5 4	5	0 2 4
9	5*	
4	6	
8	6*	
2 2	7	

b $n = 20$

$$\text{Median} = \frac{20 + 1}{2}$$

$$= 10.5^{\text{th}} \text{ score}$$

$$\text{Quartiles} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

Starlet

$$\text{Min.} = 16, \text{Max.} = 72$$

$$\text{Median} = \frac{36 + 43}{2}$$

$$= 39.5$$

$$Q_1 = \frac{29 + 31}{2}, Q_3 = \frac{55 + 59}{2}$$

$$= 30 \quad = 57$$

∴ five-number summary is:

{16, 30, 39.5, 57, 72}.

Glow-worm

$$\text{Min} = 23$$

$$\text{Max} = 54$$

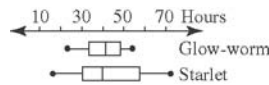
$$\text{Median} = 41$$

$$Q_1 = \frac{31 + 35}{2}$$

$$= 33$$

$$Q_3 = 48$$

∴ five-number summary is
{23, 33, 41, 48, 54}



- c The big question is “do we want consistency with our purchase?” If yes, then Glow-worm is the choice because the results do not vary as much as Starlet. They cluster more. Glow-worm-IQR = 15, overall range = 31 while Starlet IQR = 27, overall range = 56. Glow-worm

median is higher 41 versus $39\frac{1}{2}$. However, over 25% of

the Starlet bulbs have out flashed Glow-worms.

- d Recommend Glow-worm. Perform better more consistently.

14 a Key: 1|3 = 13 seconds

Placebo Leaf	Stem	Drug Leaf
8 7	1	3 7 8
8 8 5 5 4 2 1 0	2	1 5 9
8 6 5 5 4 2 2 2	3	2 2 4 5 7 8
2	4	2 2 2 4 5 7
	5	
	6	2 2

b $n = 20$

$$\text{Median} = \frac{20 + 1}{2}$$

$$= 10.5^{\text{th}} \text{ score}$$

$$\text{Quartiles} = \frac{10 + 1}{2}$$

$$= 5.5^{\text{th}} \text{ score}$$

Drug

Min = 13, Max. = 62

$$\text{Median} = \frac{35 + 37}{2}$$

$$= 36$$

$$Q_1 = \frac{25 + 29}{2}$$

$$= 27$$

$$Q_3 = \frac{42 + 44}{2}$$

$$= 43$$

∴ five number summary is:
{13, 27, 36, 43, 62}

Placebo

Min. = 17, Max. = 42

$$\text{Median} = \frac{28 + 32}{2}$$

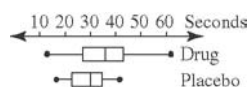
$$= 30$$

$$Q_1 = \frac{22 + 24}{2}$$

$$= 23$$

$$Q_3 = 35$$

∴ five number summary is:
{17, 23, 30, 35, 42}



c to e Variability versus Consistency.

The placebo group showed a more consistent performance with a better median time – 30 seconds versus 36 seconds, a smaller range – 25 versus 49, and a smaller IQR = 12 versus 16. Even though some people felt relief quicker in the drug group – 4 seconds faster on minimum, the drug group results varied enormously. Therefore the drug did not seem to enhance the groups performance on a whole.

Chapter review

Multiple choice

- Time = numbers/continuous
∴ answer is **D**.
- Country – categorical and can't be ordered ⇒ nominal data
∴ answer is **A**.
- Data on table responses is categorical and can be ordered
⇒ ordinal data
∴ answer is **B**.
- Relative frequency
$$= \frac{4}{2 + 6 + 7 + 4 + 1}$$
$$= \frac{4}{20}$$

∴ answer is **B**.
- Positively skewed because higher scores spread over a greater range
∴ answer is **C**.
- 20 kg or less ⇒ 15 articles
∴ answer is **C**.
- 90% mark ≈ 36 kg
∴ answer is **D**.
- Median
0 0 1 2 3 4 5 5 6 8 8 9
$$\text{Median} = \frac{12 + 1}{2}$$
$$= 6.5^{\text{th}} \text{ score}$$

∴ Median = $\frac{4 + 5}{2}$
$$= 4.5$$

answer is **D**.

9

Score	Frequency	cf
15	12	12
16	15	27
17	10	37
18	6	43
19	3	46
20	2	48

$$\text{Median is } \frac{48 + 1}{2}$$

$$= 24.5^{\text{th}} \text{ score}$$

$$\therefore \text{Median} = 16$$

answer is **C**.

10

Score	(x)	(f)	xf
20 –	25	3	75
30 –	35	6	210
40 –	45	12	540
50 –	55	13	715
60 –	65	7	455
70 –	75	2	150
		43	2145

$$\text{Mean} = \frac{2145}{43} \\ \approx 49.88.$$

$\therefore$ answer is **D**.

- 11** The percentage of students who obtained 30 marks or more = $7 + 11 + 14 + 16 + 18 + 12 = 78$.

Therefore, the answer is **D**.

- 12** $2 + 3 + 2 + 3 + 6 + 6 + 7 + 11 + 14 = 54$. Therefore, the median mark lies between 40 and 45. So the answer is **B**.
- 13** The mean would exceed mode because graph is positively skewed.
 $\therefore$ answer is **B**.
- 14** The interquartile range is not generally affected by extreme values.
 $\therefore$ answer is **C**.
- 15** From graph,
 $Q_1 = 12$
 $Q_3 = 24$
 $\therefore \text{IQR} = 24 - 12$
 $= 12$
 $\therefore$ answer is **E**.
- 16** Variability looks at range/spread $\Rightarrow$ the mean is not a measure of variability.
 $\therefore$ answer is **B**.
- 17** If Median = 13th score, this leaves 12 below so
 $Q_1 = \frac{12+1}{2}$
 $= 6.5^{\text{th}}$ score
 $\therefore$ answer is **C**.
- 18** From boxplot, $Q_3 = 36$.
 $\therefore$ answer is **D**.
- 19** From boxplot, $Q_3 = 36$ and $Q_1 = 24$
 $\therefore \text{IQR} = 36 - 24$
 $= 12$
 $\therefore$ answer is **B**.
- 20** $n = 20$
Median = $\frac{20+1}{2}$
 $= 10.5^{\text{th}}$ score
 $\therefore \text{Median} = \frac{125+129}{2}$
 $= 127$
 $Q_1 = \frac{10+1}{2}$
 $= 5.5^{\text{th}}$ score
 $\therefore Q_1 = \frac{116+118}{2}$
 $= 117$
 $Q_3 = \frac{137+138}{2}$
 $= 137.5$
 $\therefore \text{IQR} = 137.5 - 117$
 $= 20.5$
 $\therefore$ answer is **A**.

- 21** $n = 23$
Median = $\frac{23+1}{2}$
 $= 12^{\text{th}}$ score
 $= 7.5$
 $\therefore$ answer is **D**.

- 22** The standard deviation cannot be easily determined from a boxplot.

$\therefore$ answer is **D**.

- 23** Positively skewed shown by the histogram.
 $\therefore$ answer is **A**.

- 24** 36 in 28 36 50 62 64 is Q_1 and
Range = $64 - 28$
 $= 36$

$\therefore$ answer is **C**.

25 Group A

$n = 20$

Min	Q_1	Median	Q_3	Max
230	253.5	266	277.5	291

$$\text{IQR} = 277.5 - 253.5 \\ = 24$$

$$\text{Range} = 291 - 230 \\ = 61$$

Group B

Min	Q_1	Median	Q_3	Max
231	244.5	249.5	261	292

$$\text{IQR} = 261 - 244.5 \\ = 16.5$$

$$\text{Range} = 292 - 231 \\ = 61$$

Both groups have the same range and Group A has the large median.

$\therefore$ Group A has larger IQR

$\therefore$ answer is **C**.

- 26** Data from Group B are positively skewed – higher values are spread over greater range.

$\therefore$ answer is **D**.

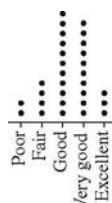
- 27** $Q_3 = 14$, so 25% of jellyfish taken from location A had a diameter greater than 14 mm. Therefore, the answer is **C**.

- 28** The median for location A is smaller, while both range and IQR are greater than for the location B. Therefore, the diameters of the jellyfish taken from location A are generally less than the diameters of the jellyfish taken from location B and more variable. So the answer is **C**.

Short answer

- 1 a** Time – continuous
b No. of goals – discrete
c Type of play – nominal
d Income – discrete
e Pages in novel – discrete
f Level of contentment as ordered response – ordinal
- 2** Boxplot (a) $\rightarrow$ **G**
Stem-and-leaf plot (b) $\rightarrow$ **E**
Dot plot (c) $\rightarrow$ **C**
Bar chart (d) $\rightarrow$ **A**
Histogram (e) $\rightarrow$ **F**
Polygon (f) $\rightarrow$ **B**
Ogive (g) $\rightarrow$ **D**
- 3 a** The histogram is positively skewed.
b i From the graph, the number of years in which the mean yearly rainfall was 500 mm or more = $13 + 7 + 3 + 1 + 1 + 1 = 26$.
ii From the graph, the number of years in which the mean yearly rainfall was between 500 mm and 600 mm is $13 + 7 = 20$. Therefore, the percentage of years in which the mean yearly rainfall was between 500 mm and 600 mm is $\frac{20}{103} \times 100\% = 19.4\%$.

4 a



b $11 + 10 + 3 = 24$

c Relative frequency good or better

$$= \frac{24}{30}$$

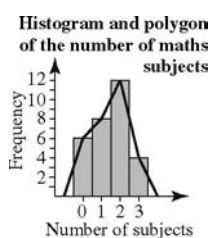
$$= \frac{4}{5}$$

d $\frac{24}{30} \times 100 = 80\%$

5 a

Number of subjects	Frequency
0	6
1	8
2	12
3	4

b and c



d Relative frequency for 2 maths subjects

$$= \frac{12}{30}$$

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

e % frequency for 3 maths subjects

$$= \frac{4}{30} \times 100\%$$

$$= 13\frac{1}{3}\%$$

6 Reading off ogive:

a 17 people spent \$50 or less

b 5 people spent more than \$100

c % people spent less than \$80

$$= \frac{36}{50} \times 100\%$$

$$= 72\%$$

d 80th percentile is 40 people and it occurs at \$88 indicating that 80% of the people spent \$88 dollars or less on shopping.

e Lowest 10% of 50 is 5. Five people spent a maximum of \$20.

7

Scores	Scores — mean	Square difference
1	-3.9	15.21
2	-2.9	8.41
2	-2.9	8.41
3	-1.9	3.61
5	0.1	0.01
6	1.1	1.21
6	1.1	1.21
7	2.1	4.41
8	3.1	9.61
9	4.1	16.81
49		68.9

a Mean = $\frac{49}{10} = 4.9$

$$\begin{aligned}\text{b The median} &= \frac{10+1}{2} \\ &= 5.5^{\text{th}} \text{ score} \\ \therefore \text{Median} &= \frac{5+6}{2} \\ &= 5.5\end{aligned}$$

c The scores 2 and 6 both occur twice
Mode = 2 and 6

$$\text{d Range} = 9 - 1 = 8$$

$$\begin{aligned}\text{e } Q_1 &= 2 \text{ and } Q_3 = 7 \\ \text{IQR} &= Q_3 - Q_1 \\ &= 7 - 2 \\ &= 5\end{aligned}$$

$$\begin{aligned}\text{f Standard deviation} &= \sqrt{\frac{68.9}{9}} \\ &\approx 2.767\end{aligned}$$

$$\begin{aligned}\text{g Variance} &= \text{standard deviation squared} \\ &= \frac{68.9}{9} \\ &= 7.656\end{aligned}$$

8

Scores	Mid-point	Frequency	Frequency $\times$ Midpoint	Scores – mean	Square difference	Square difference $\times$ Frequency
30 –	35	3	105	–30.78125	947.4853516	2842.456
40 –	45	6	270	–20.78125	431.8603516	2591.162
50 –	55	12	660	–10.78125	116.2353516	1394.824
60 –	65	15	975	–0.78125	0.610351563	9.155273
70 –	75	18	1350	9.21875	84.98535156	1529.736
80 –	85	10	850	19.21875	369.3603516	3693.604
TOTAL		64	4210		1950.537109	12 060.94

$$\begin{aligned}\text{a Mean} &= \frac{4210}{64} \\ &= 65.781\end{aligned}$$

$$\begin{aligned}\text{b Median} &= \frac{64+1}{2} \\ &= 32.5^{\text{th}} \text{ score} \\ \therefore \text{Median group} &= 60 -\end{aligned}$$

c Highest frequency = 18
 $\therefore$ Modal group 70 –

$$\begin{aligned}\text{d Range} &= 90 - 30 \\ &= 60\end{aligned}$$

$$\begin{aligned}\text{e Standard deviation} &= \sqrt{\frac{12\,060.94}{63}} \\ &= 13.836\end{aligned}$$

$$\begin{aligned}\text{f Variance} &= \text{standard deviation squared} \\ &= \frac{12\,060.94}{63} \\ &\approx 191.443\end{aligned}$$

9 a

Non dreaming		Dreaming
	0	7 7
	1	0 4 5 8
1	2	0 1 4 7
2	3	4 6 9
	4	1
2	5	
6 3 3	6	
6	7	
3 2	8	
1	9	
9	10	
0 0	11	
	12	
	13	
	14	
8	15	

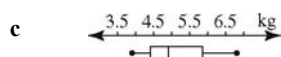
$$2 \mid 1 \quad \approx 2.1$$

- b** Using calculator, for the data for dreaming sleep mean = 2.2 hours and standard deviation = 1.2 hours.
- c** Both, the center (mean) and spread (standard deviation) of hours of non-dreaming sleep are greater than the center and the spread for hours of dreaming sleep. So overall, the hours for non-dreaming sleep are larger and more spread-out.

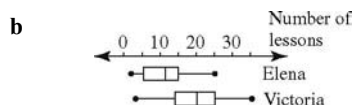
10 a Key: 3*9 = 3.9

Stem	Leaf
3*	9
4	0 0 2 3
4*	5 6 7 8 8
5	0 3
5*	5 8 8 9
6	1 2 2
6*	8

- b** $n = 20$
Median = $\frac{21}{2}$
= 10.5th score
 $\therefore$ Median = $\frac{4.8 + 5.0}{2}$
= 4.9
Min = 3.9, Max = 6.8
Quartiles = $\frac{10 + 1}{2}$
= 5.5th score
 $\therefore Q_1 = \frac{4.3 + 4.5}{2}$
= 4.4
 $Q_3 = \frac{5.8 + 5.9}{2}$
= 5.85
 $\therefore$ five-number summary
= {3.9, 4.4, 4.9, 5.85, 6.8}



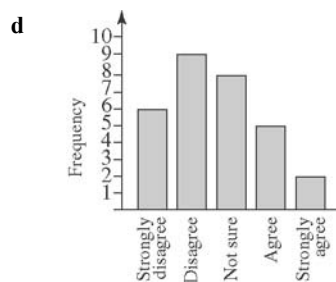
- 11 a** Elena: Median = $\frac{11 + 12}{2}$,
= 11.5
 $Q_1 = 6, Q_3 = 15$
Victoria: Median = 20
 $Q_1 = 14$
 $Q_3 = 25$



- c** The number of lessons given by Victoria has varied compared to Elena. Elena's number has been more consistent.
Victoria's range was 32 while Elena's was 9
Victoria's IQR was 11, while Elena's was 9
Victoria's median value was 20 lessons while Elena's was $11\frac{1}{2}$ lessons

Extended response

- 1 a** Counting the dots in the dot plot, 30 people participated in the survey.
b The data is categorical.
c Since the data is ordered, the data is ordinal.



e

Opinion	Frequency
Strongly disagree	6
Disagree	9
Not sure	8
Agree	5
Strongly agree	2

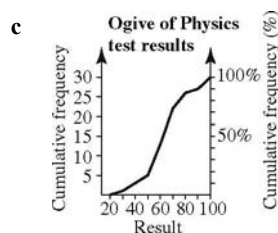
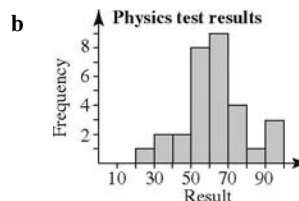
f Relative frequency = $\frac{5}{30}$
= $\frac{1}{6}$

g % frequency = $\frac{8}{30} \times 100$
= $26\frac{2}{3}\%$

- h** 20% of 30
= 0.20×30
= 6
Strongly disagree

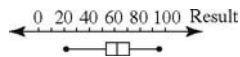
2 a

Physics Test Mark	Frequency	cf
20 –	1	1
30 –	2	3
40 –	2	5
50 –	8	13
60 –	9	22
70 –	4	26
80 –	1	27
90 –	3	30



- d** From graph
 $Q_1 = 25\% = 53$
Median = $50\% = 62$
 $Q_3 = 75\% = 72$
- e** Top 10% $\Rightarrow$ 90% up off graph
90% = 90
 $\therefore$ A score of 90 or more.

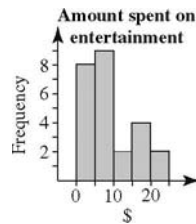
f



3 a

Spending \$	Frequency
0 –	8
5 –	9
10 –	2
15 –	4
20 –	2

b



Scores	Mid-point	Frequency	Frequency $\times$ Midpoint	Scores – mean	Square difference	Square difference $\times$ Frequency
0 –	2.5	8	20	–6.6	43.56	348.48
5 –	7.5	9	67.5	–1.6	2.56	23.04
10 –	12.5	2	25	3.4	11.56	23.12
15 –	17.5	4	70	8.4	70.56	282.24
20 –	22.5	2	45	13.4	179.56	359.12
TOTAL		25	227.5		307.8	1036

$$\text{c Mean} = \frac{227.5}{25} = 9.1$$

d Median out of 25

$$= \frac{26}{2}$$

= 13th score $\therefore$ Median class = \$5 – <\$10

e Highest frequency = 9

 $\therefore$ Modal class \$5 – <\$10

$$\text{f } s = \sqrt{\frac{1036}{25-1}}$$

$$s = \sqrt{\frac{1036}{24}} = 6.57$$

$$\text{g } s^2 = \frac{1036}{24} = 43.17$$

4 a Designer tiles appears to be negatively skewed.
Nathan's tiles appears to be positively skewed.

b Designer tiles

Price per m ² (\$)	Frequency
20–	2
30–	2
40–	5
50–	8
60–	10
70–	10
80–	12
90–	9

Nathan's tiles

Price per m ² (\$)	Frequency
20–	6
30–	11
40–	13
50–	12
60–	8
70–	6
80–	1
90–	1

c

Statistics	Designer tiles	Nathan's tiles
Mean	$\frac{4060}{58} = 70$	$\frac{2930}{58} = 50.52$
Median class	70 – <80	40 – <50
Modal class	80 – <90	40 – <50
Range	100 – 20 = 80	100 – 20 = 80
Standard deviation	18.85	16.56

d The measures of central tendency are all higher for Designer tiles. The spread is almost the same with the standard deviation for Designer tiles slightly higher. For Designer Tiles, people tend to buy the more expensive tiles.

e The more expensive tiles from Nathan's tiles would be better placed at Designer tiles. Increasing the range of cheaper tiles at Nathan's tiles would be beneficial, as the statistics suggest that people tend to buy the less expensive tiles at Nathan's tiles.

5 a Key: 1|3 = 13

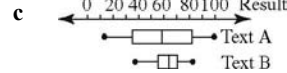
Text B	Text A
Leaf	Leaf
	1
	2
9 7	3
8 2	4
9 9 8 8 5 5 3 2	5
8 8 8 6 5 4 4 3 2	6
9 9 5 2 2	7
2 1	8
	9

b Text A $n = 25$

Min	Q_1	Median	Q_3	Max
13	35	59	82	98

Text B $n = 28$

Min	Q_1	Median	Q_3	Max
37	55	63.5	70	82



d Performance on test. Students who used Text A had varied results while the students who used Text B were more consistent and they had a higher median score.

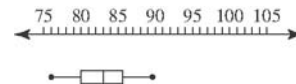
Text A: range 85, IQR = 47, Median = 59

Text B: range 44, IQR = 15, Median = 63.5

e and f Students individual differences — that is, studying, doing the work, ability can affect the results. However, you would probably go with the text that produced more consistent results $\Rightarrow$ Text B

6 a Using calculator, the standard deviation of the heights of the 18-month-old boys is 3.8

b Using calculator, five number summary is: $X_{\min} = 76$; $Q_1 = 80$; Med = 83; $Q_3 = 85.8$; $X_{\max} = 89.8$



c From the box-plot, median height of the 27-month-old boys is 89.5 cm.

d It is apparent from the three parallel boxplots that *height* and *age* (18 months, 27 months, 36 months) are **positively** related, as the median height of boys increases with age.