NEW CENTURY MATHS 11
MATHEMATICS STANDARD (PATHWAY 2)

FULLY WORKED SOLUTIONS

Chapter 10

SkillCheck

Question 1

a Add up all the leaves. There are 29 visitors.

b The highest number in the stem is 5, and its highest leaf is 8, so the highest number is 58.

c Look through the numbers and 2 appears 3 times in the '1' stem, so 12 occurs most frequently.

d All numbers left of the 6 in the stem '1', and previous stem(s), i.e. the three numbers opposite '0'.
Do not include the 16, so there are 3 + 5 = 8

e Not counting the middle age, there are 28 other numbers.
14 will be before the middle age and 14 after it, so the middle age will be the 15th age, which is 24.

Question 2

A column graph.

Question 3

a Count the dots. There are 10.

b The shoe size with the most dots is size 7.

c The outlier must be clearly separate from the rest of the data.
In this case, it is size 12. The person has large feet.

d There are no dots above size 10, so 0.

e 3 of the 10 (30%) students had a size greater than 8.

Question 4

a Opposite 0 cars is a frequency of 4, meaning there were 4 families who did not own a car.

b Find the highest frequency, which is 16, which corresponds to families with 1 car, so the most common number of cars owned is 1.

c The highest number of cars with a frequency of 1 or more is 4.

d Add all the frequencies.
$4 + 16 + 11 + 0 + 1 = 32$ students.

e Work out for all families with cars.
16 with 1 car (16), 11 with 2 cars (22), 1 with 4 cars (4).
Total cars = $16 + 22 + 4 = 42$ cars.

Question 5

Since the data is continuous, the first class centre will be half way between 50 and 60,

i.e. $\frac{50+60}{2} = 55$.

The other class centres will be 65, 75, 85, 95 and 105.
Find the frequencies by counting, or using tally marks. The frequencies are 4, 12, 7, 8, 3, and 6.

50 – < 60	55	4
60 – < 70	65	12
70 – < 80	75	7
80 – < 90	85	8
90 – < 100	95	3
100 – < 110	105	6

Exercise 10.01 The mean, median and mode

Question 1

a **i** Add up all the numbers and divide by how many numbers there are.

$\text{mean} = (1 + 1 + 2 + 5 + 5 + 7 + 9 + 10) \div 8$

$= 40 \div 8$

$= 5$

ii The median is the middle number. As both middle numbers are 5, the median is 5.

iii The mode is the most common score. In this case there are two modes, 1 and 5.
The data is bimodal.

b **i** Add up all the numbers and divide by how many numbers there are.

$\text{mean} = (37 + 31 + ... + 35 + 38) \div 10$

$= 344 \div 10$

$= 34.4$

ii First put the numbers in order.
31 31 32 32 34 35 35 37 38 39
The median is the average of the two middle numbers.
$\therefore$ median = (34 + 35) ÷ 2 = 34.5

iii The mode is the most common score. In this case, there are three modes, 31, 32 and 35.
This data is multimodal.

c **i** Add up all the numbers and divide by how many numbers there are.

$\text{mean} = (28 + 40 + ... + 41 + 30) \div 9$

$= 324 \div 9$

$= 36$

ii First put the numbers in order.
28 29 30 31 38 40 41 42 45
The median is the middle number, 38.

iii There is no mode, as each number only occurs once

d **i** Add up all the numbers and divide by how many numbers there are.

$$\text{mean} = (5+8+...+7+5) \div 11$$
$$= 99 \div 11$$
$$= 9$$

ii First put the numbers in order.
5 5 7 7 8 8 9 10 11 14 15
The median is the middle number, 8.

iii The mode is the most common score. In this case, there are three modes, 5, 7 and 8.
This data is multimodal.

Question 2

a Count all the leaves. There were 24 rounds.

b Add the actual scores and divide by number of rounds.

$$\text{mean} = (6+6+12+13+....+37+40+62) \div 24$$
$$= 580 \div 24$$
$$= 24.1666...$$
$$\approx 24$$

c There are 24 scores. The middle scores are the 12th and 13th scores.
Both the 12th and 13th scores are 20. So, the median is the average of 20 and 20, which is 20.

d In this case there are three of both 14 and 20. So there are two modes, 14 and 20.
The data is bimodal.

Question 3

a Add the actual scores and divide by number of rounds.

$$\text{mean} = (28+34+....+24+27) \div 10$$
$$= 260 \div 10$$
$$= 26$$
$\therefore$ D

b First put the numbers in order.
22 24 24 24 25 26 26 27 28 34
The median is the average of the two middle numbers.
The median = (25 + 26) ÷ 2 = 25.5, $\therefore$ **C**.

c The mode is the most common score, 24, $\therefore$ **A**.

Question 4

a No need to add up the frequencies as we're told there were 160 matchboxes.
Calculate fx by multiplying the number of matches (x) by the frequencies (f). They are as follows:

$48 \times 10 = 480$

$49 \times 45 = 2205$

$50 \times 52 = 2600$

$51 \times 39 = 1989$

$52 \times 9 = 468$

$53 \times 5 = 265$

Add fx's:

$480 + 2205 + 2600 + 1989 + 468 + 265 = 8007$

$\therefore$ mean $= 8007 \div 160$

$= 50.04...$

≈ 50.0

b Yes as the mean is 50.0.

c The highest frequency is 52. This is for 50 matches. So, the mode is 50.

d There are 160 matchboxes The median (or middle value) is the average of the 80th and 81st value. They are both 50, so the median is 50.

Question 5

a Count the dots. There are 20 families.

b There is 1 zero, 3 ones, etc. Include an fx column.

No. of children in family	f	fx
0	1	0
1	3	3
2	8	16
3	4	12
4	2	8
5	1	5
6	0	0
7	0	0
8	1	8
Total:	20	52

$\therefore$ Mean = 52 ÷ 20 = 2.6

c Count the dots from left to right. The 10th and the 11th scores are both 2.
$\therefore$ the median is 2.

d There are eight 2's. This is the highest frequency
$\therefore$ the mode is 2.

e The outlier must be clearly separate from the rest of the data.
In this case, it is the family with 8 children (8).

f **i** The outlier, 8, is above the mean (average), so if it was removed, the mean would decrease.

ii If the outlier, 8, is removed, there will be 19 families.
There will be the median in the middle with 9 families on each side.
The median will be the 10th score, which is still 2, so in this case the median is not affected.

iii If the outlier, 8, is removed, the highest column, the mode, will still be 2,
so the mode is not affected.

Question 6

a

No. of calls, *x*	***f***	***fx***
2	1	2
3	2	6
4	3	12
5	5	25
6	2	12
7	2	14
Total:	15	71

b This is the total of the frequencies = 15

c There are five 5's. This is the highest frequency.
$\therefore$ the mode is 5.

d There are 15 calls. The median would be the 8th value.
$\therefore$ the median is 5.

e mean = total no. of calls $\div$ no. of days
$= 71 \div 15$
$= 4.733...$
≈ 4.7 phone calls per day

Question 7

a There will also need to be a column for *fx*.

Speed (km/h)	Class centre (x)	f	fx
36–40	38	64	2432
41–45	43	36	1548
46–50	48	18	864
51–55	53	15	795
56–60	58	11	638
61–65	63	5	315
	Total:	149	6592

Mean $= 6592 \div 149$

$= 44.2416...$

≈ 44.24 phone calls per day

b This is the total of the frequencies = 149

Question 8

a Add the frequencies to find there are 234 scores.
This is two lots of 117 scores, so the median will be the average of the 117th and the 118th scores.
Find these scores by adding the frequencies. The first to 28th scores are all 20–29, the 29th to 73rd scores are all 30–39, and the 147th to 21st scores are all 40–49.
The 117th and the 118th scores are both in the class 40-49, so the median class is 40–49.

b The modal is the class with the highest frequency, 40–49.

Question 9

Count the dots to check that there are 21 (3 weeks) temperatures.

a The mode is the most common score, i.e. the highest frequency, or highest column of dots, –2 and 2.

b There are 10 temperatures on each side of the median, so the median will be the 11th score.
Count the dots from left to right. The median is 2.

c Multiply each temperature by its frequency, to find the *fx*. Then find the total of these values.
$3 \times (-2) + 2 \times (-1) + 1 \times 0 + 2 \times 1 + 3 \times 2 + 1 \times 4 + 2 \times 5 + 2 \times 6 + 2 \times 7 + 1 \times 8 = 54$

Mean $= 54 \div 21$

$= 2.571...$

≈ 2.6 phone calls per day

Question 10

a The modal is the class with the highest frequency, 300–< 400.

b Add the frequencies to find there are 44 scores.
This is two lots of 22 scores, so the median will be the average of the 22nd and the 23rd scores.
Find these scores by adding the frequencies. The first five scores are all 100–<200,
the 6th to 16th scores are all 200–<300, and the 17th to 36th scores are all 300–<400.
The 22nd and the 23rd scores are both in the class 300–<400, so the median class is 300–<400.

Question 11

a The mean takes in all scores. The median only uses the middle score or scores and the mode only uses one score, the most common score. The mean.

b The mean, or average, is often not one of the scores. The mode is always one of the scores.
The median will always be a score (the middle score) if there is an odd number of scores,
but may not be one of the scores if there is an even number of scores. The median.

c Half the scores may not be above/below the mean or mode, but (roughly speaking) will be for the median. The median.

d There is only ever one mean and one mode for any set of data, but if more than one score ‘dead heats’ for most common score, there can be more than one mode. The mode.

e The mode is the most frequent score, and so doesn’t need rounding. The median is one score, or the average of two scores, and so doesn’t usually need rounding. The mean is the total divided by the number of scores, and will often need rounding, particularly when a recurring decimal is involved. The mean.

f The mean, or average, needs numerical data. The median, which involves order, can be used for ordinal (qualitative) data, but not categorical as there can be no order. You can have the category with the highest frequency, so the mode can be used for categorical data. The mode.

g The mode is usually unaffected by outliers, as an outlier is very unlikely to occur most frequently. The median may be affected, but only slightly, as removing outliers may slightly change the middle value. As the mean uses all the data, extreme values will have the greatest effect on the mean.
The mean.

h The mean often works out not to be one of the scores. If there is an even number of scores and the two middle scores are different, the median will not be one of the scores. The mode is the score with the highest frequency, and so will always be one of the scores. The mode.

Question 12

a The mean as it takes in all the marks. You could argue for the median if you thought there may be one or two students who made little effort, or had illness or other problems and would unfairly bring the class average down.

b For clothing sizes the mode is used.

c The median should be used as there may be outliers, e.g. wealthy people renting luxury accommodation, or low-rent public housing.

d The mean as it includes all the data and there are unlikely to be outliers. The mode could be used if interest is more in the most popular sizes.

e The mean, as it includes all the data and there are unlikely to be outliers.

f This data is categorical, and so only the mode can be used.

Question 13

a $1 + 3 + 1 + 2 = 7$

b Calculate the total of all the wages and divide by the total number of wages.

$$\begin{aligned}\text{Mean} &= (\$158\ 300 + ... + 2 \times \$68\ 500) \div 7 \\ &= \$573\ 800 \div 7 \\ &= \$81\ 971.428... \\ &\approx \$82\ 000\end{aligned}$$

c The median will be the fourth score. This will be one of the two clerical officers.
$\therefore$ median = \$68 500.

d The mean is higher because the general manager's income pushes it up.

e The median, because the mean is influenced by the general manager's salary, which is an outlier.

Question 14

a **i** $\text{Mean} = (49 + 32 + ... + 50) \div 8$
$= 305 \div 8$
$= 38.125$

ii Put the ages in order. The median will be the average of the middle two scores, 37 and 39.
∴ Median = (37 + 39) ÷ 2 = 38

iii The mode is the most common score, 32.

b **i** Because the new teacher is younger than the teacher being replaced, the total will be less and so the mean will decrease.

ii The 39 will be replaced by 22 and the order will be as follows.
22 25 32 32 37 41 49 50
The median will now be the average of 32 and 37.
Median = (32 + 37) ÷ 2 = 34.5
∴ The median has decreased.

iii The most common score will still be 32. The mode is not affected.

Question 15

a Add the frequencies.
4 + 7 + 7 + 9 + 12 = 39.

b The most frequent colour is white.

c The mean and median need numerical data (or at least ordinal for the median).
The data is categorical (neither numeric, not able to be put in order).
Hence, the mode is the only valid measure of central tendency here.

Question 16

a **i** There are 11 amounts.

$$\text{Mean} = (\$370 + 628 + ... + 481) \div 11$$
$$= \$5757 \div 11$$
$$= \$523.3636...$$
$$\approx \$523$$

ii Put the repayments in order. The middle repayment is the 6th one.
$\therefore$ the median is \$481.

iii As the repayments are all different, there is no mode.

b The mean is inappropriate because it is distorted (affected) by the outlier \$1027.
The mode cannot be used as it does not exist, mainly because the repayments can vary too much.

c Without the outlier the new mean is:

$$\text{Mean} = 4730 \div 10$$
$$= \$473$$

The new median is the average of the two middle values: $(435 + 481) \div 2 = 458$.

Difference the new mean and median = 473 – 458
= 15

Difference the old mean and median = 523 – 481
= 42

$\therefore$ The new mean and median are much closer than the old mean and median were.

Question 17

a Count the dots to find there are 10 scores.

i $\text{Mean} = (2\times 6 + 3\times 7 + ... + 1\times 12) \div 10$

$= 79 \div 10$

$= 7.9$

ii Count the dots from left to right.
The median will be the average of the 5th and the 6th scores = (7 + 8) ÷ 2 = 7.5

iii The mode is the most common score, i.e. the highest frequency, or highest column of dots, 7.

b The outlier must be clearly separate from the rest of the data. In this case, it is size 12.

i Since the outlier is larger than the mean, the mean will decrease.

ii The mode will be unaffected as size 7 will still have the highest frequency (3).

c The mode, because the store wants to know which size they will sell the most of.

Question 18

a **i** Count the leaves to check that there are 14 temperatures for the two weeks.

$\text{Mean} = (22 + 24 + ... + 34) \div 14$

$= 378 \div 14$

$= 27°\text{C}$

ii There are 14 marks, 7 on each side of the median.
The median will be the average of the 7th and 8th scores, which are both 27.
∴ Median = 27°C.

iii Looking through the scores, the most common score is 27°C.

b The mean, as it includes all the data and there are no outliers.

Exercise 10.02 Quartiles, deciles and percentiles

Question 1

a Put the data in order.

8.7 8.9 9.0 9.1 9.2 9.6 9.9 10.1

10.6 10.6 10.8 11.0 12.3 13.1 13.5

The second quartile (median) is the middle value. $\therefore$ $Q_2 = 10.1$

The first quartile is the middle of the first half of the data, i.e. of 8.7 to 9.9. $\therefore$ $Q_1 = 9.1$

The third quartile is the middle of the second half of the data, i.e. of 10.6 to 13.5 $\therefore Q_3 = 11.0$

$\therefore$ Q_1, Q_2, Q_3 are: 9.1, 10.1, 11.0.

b Put the data in order.

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

The median is 50. $Q_2 = 50$

The other quartiles are the middle of the two halves of the data.

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

$Q_1 = (49 + 50) \div 2 = 49.5$, $Q_3 = 51$

$\therefore$ Q_1, Q_2, Q_3 are: 49.5, 50, 51

c Put the data in order.

The median = $(3.40 + 3.50) \div 2 = 3.45$, $Q_2 = 3.45$

The other quartiles are the middles of the two halves of the data.

2.80 3.00 3.00 3.20 3.40 | 3.50 3.50 3.50 3.90 4.10

$Q_1 = 3.00$, $Q_3 = 3.50$

$\therefore$ Q_1, Q_2, Q_3 are: 3.00, 3.45, 3.50

d Put the data in order.

The median is 18. $Q_1 = 18$

The other quartiles are the middle of the two halves of the data. Do not include the median, 18.

0 5 7 | 9 16 17 $\boxed{18}$ 21 21 22 | 24 26 27

$Q_1 = 8$, $Q_3 = 23$

$\therefore$ Q_1, Q_2, Q_3 are: 8, 18, 23

Question 2

a Count the leaves. There are 32 game scores. This splits the data into two sets of 16 scores.
The median will be the average of the 16th and the 17th scores.
$\therefore$ Median = $Q_2 = (113 + 114) \div 2 = 113.5$

b The first 16 scores split into two sets of 8 scores. The lower quartile will be the average of the 8th and the 9th scores.
$\therefore\ Q_1 = (99 + 104) \div 2 = 101.5$

c The last 16 scores also split into two sets of 8 scores. If you count backwards, the upper quartile will be the average of the 8th and the 9th scores **from the end**.
$\therefore\ Q_3 = (126 + 125) \div 2 = 125.5$
(Alternatively, there will be 16 scores, then the 8th and 9th scores of the second half of the data. This means you will average the 24th and 25th scores.)

Question 3

Count the dots to check that there are 20 scores. This splits the data into two sets of 10 scores. The upper quartile will be the middle of the upper (highest) 10 scores. This will be the average of the 15th and the 16th scores.
These are both 8. So, $Q_3 = 8$, $\therefore$ **D**.

Question 4

Count the scores. There are 30. This means we must sort the data into ten groups (as deciles), each of three scores.

Put the scores in order. You may find it easier not to split any groups of three.

45 46 50 | 57 60 61 | 61 63 64 | 64 66 67 | 69 71 74|

75 75 78 | 80 81 83 | 84 85 87 | 89 92 95 | 95 97 100

a $D_8 = (87 + 89) \div 2 = 88$

b $D_3 = 64$

c $P_{40} = D_4 = (67 + 69) \div 2 = 68$

d $20\% = \frac{2}{10}$
From the data, this is 61.

e Looking at the data, this is the last four tenths, which is 40%.

Question 5

There are 20 scores. This means we must sort the data into ten groups, each of two scores. The scores are already in order, from lowest to highest.

a $D_1 = (4 + 5) \div 2 = 4.5$

b $D_5 = 7$

c This will be D_7. The 70% mark will cut off after the 14th score.
The 14th and 15th scores are both 8, $\therefore 8$.

d This will be D_6. The 60% mark will cut off after the 12th score.
The 12th score is 7, and this is the end of the top 60% of scores, so 7.

e $P_{90} = D_9 = 9$. (The 18th and 19th scores are both 9.)

Question 6

a Median = $P_{50} = 20.6$

b This is the first or lower quartile, so it is low for his age.
As it is the lower quartile, 25% (one quarter) of boys will have a BMI lower than him.

c He is at the 9th decile, which is 25.4.

d 29.4 is the 97th percentile, which is very high for his age.
Only 3% of boys his age will have a higher BMI.

e 17.6 is the first decile, i.e. the 10th percentile.

Question 7

a Range = Highest temp – lowest temp
$= 41.0 - 15.6$
$= 25.4$

b $28.6°C$ is the 9th decile, i.e. only one tenth, or 10% of temperatures would be higher.

c Assuming temperatures were symmetrically distributed, you would expect the median to be about the same as the mean, $23.5°C$.

d The first decile, $18.9°C$.

e $41.0 - 28.6 = 12.4°C$.

Question 8

a False. P_{75} is the third quartile, Q_3, not Q_1.

b True because $\frac{6}{10} = 60\%$.

c True because 50% is the middle (i.e. median)

d True. P_{75} is the third quartile, Q_3.

e False. D_8 is the 80th percentile P_{80}, not P_{20}.

f True. Q_2 is the median, and D_5 is the 5th decile, the middle, or the median, too.

Question 9

a **i** ATAR of 75.25 is at the 60th percentile.
This means that 60% of students scored an ATAR 'below 75.25'.

ii ATAR of 61.65 is at the 40th percentile.
This means that 60% of students scored an ATAR 'above 61.65'.

iii ATAR of 93.95 is at the 90th percentile. ATAR of 99.40 is at the 99th percentile.
This means that $99 - 90 = 9\%$ of students scored an ATAR 'between 93.95 and 99.40'.

b From the table, 93.95% corresponds to an ATAR of 90, so $100\% - 93.95\% = 6.05\%$ scored an ATAR above the 90th percentile.

c If 9.1% scored above this ATAR, then 100% – 9.1% or 90.9% scored below it.
From the table, 90.9% corresponds to an ATAR of 85, so only 9.1% scored an ATAR above 85.

d The median is the 50th percentile. So an ATAR of 68.65.

e An ATAR of 81.6 matches up to the 70th percentile.

Question 10

a Median = P_{50} = 102.7 cm (from table).

b For 2.5 years, 88.3 cm is P_{25}. This is below average (median) so no, she is not tall for her age.
Only 25% of girls are shorter than her.

c If she is still at the 25th percentile, look up P_{25} for 5-year-olds, which is 106.2 cm.

d If 15% are taller than her, then 85% are shorter than her. This is P_{85}, or 103.3 cm for a 3.5-year-old.

e For 2 years old, 93.2 cm is P_{99}, $\therefore$ no, she is not short for her age.
99% of girls her age are shorter than her, so only 1% are taller.

f The third quartile is P_{75}, which for age 3 is 97.6 cm. In 18 months, aged 4.5 years, assuming she is still at the third quartile (P_{75}), she will be 109.2 cm tall.

Question 11

a Find 9 on the horizontal (Age) axis and 129 on the vertical (Height) axis.
Go up from 9 and across right from 129, to find the point (9, 129).
This point lies on the line for the 25th percentile, so 25% of boys are shorter than him.

b Find 11 on the horizontal (Age) axis and 155 on the vertical (Height) axis.
Go up from 11 and across right from 155, to find the point (11, 155).
This point lies on the line for the 95th percentile, so 95% of boys are shorter than him.

c Find 18 on the horizontal axis and move up to the 95th percentile.
Moving left, this shows 188 cm on the vertical axis.

d The first decile is the tenth percentile.
Find 103 cm on the vertical axis and move across to the 10th percentile.
Moving down, this shows age 5 years on the horizontal axis.

e **i** The third quartile is the 75th percentile.
Find 16 on the horizontal axis and move up to the 75th percentile.
Moving left, it shows a value between 178 cm and 179 cm on the vertical axis.
The answer could be 178 cm or 179 cm or even 178.5 cm.

ii The third quartile is P_{75}, which for age 3 is 97.6, assuming he is still at the 75th percentile.
Find 20 on the horizontal axis and move up to the 75th percentile.
Move left to find just under 182 cm on the vertical axis.
The answer is just under 182 cm or 182.5 cm.

Exercise 10.03 The range and interquartile range

Question 1

a Range = highest score – lowest score
= 6 – 0
= 6

b Range = highest score – lowest score
= 8 – 3
= 5

c Range = 1027 – 299 = 728

d Range = 34 – 22 = 12

Question 2

a Place the 12 data in order. The median will split the data into two equal parts, and each part will be split into two equal parts. The middle of these parts are the upper and lower quartiles.
0 0 0 | 0 0 1 | 1 1 2 | 2 3 6
Clearly, $Q_1 = 0$ and $Q_3 = 2$
∴ Interquartile range = 2 – 0 = 2

b Place the nine scores in order. The median is the middle score (5). The upper and lower quartiles are the middle of each group of four scores on each side of the median, NOT including 5.
3 3 | 4 4 5 6 6 | 6 8
$Q_1 = (3 + 4) \div 2 = 3.5$ and $Q_3 = 6$
∴ Interquartile range = 6 – 3.5 = 2.5

c Place the 11 repayments in order.
The median is the middle score (481).
The upper and lower quartiles are the middle of each group of four scores on each side of the median, NOT including 481.
299 354 370 417 435 481 509 585 628 652 1027
$Q_1 = 370$, $Q_3 = 628$
∴ Interquartile range = 628 – 370 = 258

d Place the 10 times in order. The median will split the data into two equal parts, and each part will be split into two equal parts. The middle of these parts are the upper and lower quartiles.
22 24 24 24 25 | 26 26 27 28 34
$Q_1 = 24$, $Q_3 = 27$
∴ Interquartile range = 27 – 24 = 3

Question 3

Count the dots to check there are 20 scores. The median will divide the data into two parts of 10 scores. Each of these will be divided into 5 scores.
Q_1 will be the average of the 5th and 6th scores. Counting the dots from the left,
$Q_1 = (5 + 6) \div 2 = 5.5$
Q_3 will be the average of the 15th and 16th scores. These are both 8, so $Q_3 = 8$.
Interquartile range = 8 – 5.5 = 2.5, so **A**.

Question 4

a Range = highest score – lowest score
$= 93 - 30$
$= 63$

b Count the leaves. There are 27 marks. There will be 13 marks on either side of the median (the 14th mark).
Q_1 will be the middle of the first 13 marks, i.e. the 7th mark. $Q_1 = 48$
Q_3 will be the middle of the last 13 marks, 21st mark.
(It is easier to count the 7th back from the end.) $Q_3 = 75$.
Interquartile range = 75 – 48 = 27.

Question 5

a Range = highest score – lowest score
$= 61 - 15$
$= 46$

b Put the 15 times in order. In this case, two rows work well. The median is clearly 38, and the quartiles are the middle times in the top and bottom rows.
15 26 30 34 35 37 37 38
43 44 45 46 48 52 61
$Q_1 = 34$ and $Q_3 = 46$
Interquartile range = 46 – 34 = 12.

c The person who took 15 seconds to complete the test may know all the work very well or may have guessed the answers.

Question 6

Count the leaves. There are 22 scores. The median will split the data into two equal parts of 11.
Q_1 will be the middle of the first 11 marks, i.e. the 6th mark. $Q_1 = 89$.
Q_3 will be the middle of the last 11 marks, 17th mark. (It is easier to count the 6th back from the end.)
$Q_3 = 127$.
Interquartile range = 127 – 89 = 38
$\therefore$ B

Question 7

$$\begin{aligned}\text{Range} &= \text{highest score} - \text{lowest score} \\ &= 158 - 78 \\ &= 80\end{aligned}$$

Exercise 10.04 The effect of outliers

Question 1

a 0 0 1 1 2 | 2 2 2 2 2 | 3 3 3 3 3 | 4 4 4 5 8

$Q_1 = 2$

$Q_3 = \dfrac{3+4}{2} = 3.5$

$\text{Interquartile range} = 3.5 - 2$

$= 1.5$

b **i** $Q_1 - 1.5 \times \text{IQR} = 2 - 1.5 \times 1.5$

$= -0.25$

ii $Q_3 + 1.5 \times \text{IQR} = 3.5 + 1.5 \times 1.5$

$= 5.75$

c 8 goals is an outlier, because it is greater than 5.75 (which is $Q_3 + 1.5 \times \text{IQR}$).

Question 2

a 2 5 | 6 6 $\boxed{7}$ 8 10 | 10 15

$Q_1 = \dfrac{5+6}{2}, \quad Q_3 = 10$

$= 5.5$

$\text{Interquartile range} = 10 - 5.5$

$= 4.5$

$Q_1 - 1.5 \times \text{IQR} = 5.5 - 1.5 \times 4.5$

$= -1.25$

$Q_3 + 1.5 \times \text{IQR} = 10 + 1.5 \times 4.5$

$= 16.75$

$\therefore$ There are no outliers because all data values lie between –1.25 and 16.75.

b 9 13 13 14 | 14 15 15 15 $\boxed{15}$ 16 16 16 16 | 16 17 17 18

$Q_1 = 14, \qquad Q_3 = 16$

$\text{Interquartile range} = 16 - 14$

$= 2$

$Q_1 - 1.5 \times \text{IQR} = 14 - 1.5 \times 2$

$= 11$

$Q_3 + 1.5 \times \text{IQR} = 16 + 1.5 \times 2$

$= 19$

$\therefore$ 9 is an outlier because this value is less than 11 (which is $Q_1 - 1.5 \times \text{IQR}$).

c

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

$$Q_1 = \frac{23+24}{2} \qquad Q_3 = \frac{44+49}{2}$$
$$= 23.5 \qquad = 46.5$$

$$\text{Interquartile range} = 46.5 - 23.5$$
$$= 23$$
$$Q_1 - 1.5 \times \text{IQR} = 23.5 - 1.5 \times 23$$
$$= -11$$
$$Q_3 + 1.5 \times \text{IQR} = 46.5 + 1.5 \times 23$$
$$= 81$$

$\therefore$ There are no outliers because all data values lie between –11 and 81.

d

4 4 4 5 5 **5** 5 5 5 5 5 **5**
5 5 5 6 6 **6** 6 7 7 7 9

$Q_1 = 5, \quad Q_3 = 6$

$$\text{Interquartile range} = 6 - 5$$
$$= 1$$
$$Q_1 - 1.5 \times \text{IQR} = 5 - 1.5 \times 1$$
$$= 3.5$$
$$Q_3 + 1.5 \times \text{IQR} = 6 + 1.5 \times 1$$
$$= 7.5$$

$\therefore$ 9 is an outlier because this value is greater than 7.5 (which is $Q_3 + 1.5 \times \text{IQR}$).

Question 3

a $\bar{x} = (450 + 520 + ... + 590) \div 7$

$= 3720 \div 7$

$= 531.428...$

≈ 531.43

$\therefore$ The mean wage is \$531.43.

b 230 $\boxed{420}$ 450 $\boxed{520}$ 590 $\boxed{610}$ 900

$\therefore$ Median wage is \$520.

c $Q_1 = \$420 \qquad Q_3 = \$610$

$\therefore \text{Interquartile range} = 610 - 420 = \$190$

d The manager's wage must be \$900 because this is significantly greater than any other wage

$Q_3 + 1.5 \times \text{IQR} = 610 + 1.5 \times 190$

$= \$895$

$\therefore$ The \$900 wage is an outlier, because it is greater than \$895.

e $\bar{x} = \dfrac{450 + 520 + 610 + 230 + 420 + 590}{6} = 470$

The mean wage is reduced to \$470.

230 420 450 | 520 590 610

$\text{Median} = \dfrac{450 + 520}{2}$

$= 485$

$\therefore$ The median wage is reduced to \$485.

f The new wages are found by multiplying the existing wages by 1.1.

253 462 495 572 649 671 990

$\bar{x} = \dfrac{253 + 462 + 495 + 572 + 649 + 671 + 990}{7}$

≈ 584.57

The new mean wage is \$584.57.

$1.1\bar{x}_{\text{old}} = 1.1 \times 531.43$

$= 584.57$

$\therefore$ The new mean is 10% more than the old mean.

253 462 495 $\boxed{572}$ 649 671 990

The new median = 572

$1.1 \times \text{Median}_{\text{old}} = 1.1 \times 520$

$= 572$

$\therefore$ The new median is 10% more than the old median.

So, the new mean and the new median are both 10% more than the previous amounts.

Question 4

a Total number of markers $= 1 + 4 + 5 + 9 + 1$
$= 20$

b 2 3 3 3 3 | 4 4 4 4 4 | 5 5 5 5 5 | 5 5 5 5 8

$Q_1 = 3.5$, $Q_3 = 5$

Interquartile range $= 5 - 3.5$
$= 1.5$

$Q_1 - 1.5 \times \text{IQR} = 3.5 - 1.5 \times 1.5$
$= 1.25$

$Q_3 + 1.5 \times \text{IQR} = 5 + 1.5 \times 1.5$
$= 7.25$

$\therefore$ 8 is the outlier as it is greater than 7.25.

c **i** Using a scientific calculator with:

x	f
2	1
3	4
4	5
5	9
8	1

$\therefore \ \bar{x} = 4.35$

ii Using a scientific calculator with:

x	f
2	1
3	4
4	5
5	9

$\therefore \ \bar{x} = 4.16$

d If the outlier is included, the mean number of cups of coffee drank increases.

Question 5

If Sally brings her 5-year-old sister as well, the mean age decreases (see following).

Without Sally's sister:

$$\overline{x} = \frac{13+12+11+14+12+15+14+13}{8}$$
$$= 13$$

With Sally's sister:

$$\overline{x} = \frac{13+12+11+14+12+15+14+13+5}{9}$$
$$= 12.111...$$

The median age remains unchanged (see following).

Without Sally's sister:
11 12 12 13 | 13 14 14 15
Median = 13

With Sally's sister:
5 11 12 12 $\boxed{13}$ 13 14 14 15
Median = 13

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

Question 6

a *Wombats*:

6 14 $\boxed{18}$ 22 24

Median = 18

$$\bar{x} = \frac{24+18+14+6+22}{5}$$

$$= 16.8$$

Mean = 16.8 points

Possums:

15 15 $\boxed{16}$ 16 18

Median = 16

$$\bar{x} = \frac{16+16+15+18+15}{5}$$

$$= 16$$

Mean = 16 points

Koalas:

8 12 $\boxed{14}$ 16 36

Median = 14

$$\bar{x} = \frac{36+8+14+16+12}{5}$$

$$= 17.2$$

Mean = 17.2 points

b The Possums are the most consistent because the mean and median are the same.

c 14 16 $\boxed{18}$ 22 24

Median = 18

$$\bar{x} = \frac{24+18+14+16+22}{5}$$

$$= 18.8$$

$\therefore$ Mean = 18.8 points

d The Possums are still the most consistent.

Question 7

a Sam mode = 3 Terri mode = 3

b If only the mode was known for each person, we would say that they are equally good salespeople.

c *Sam*: 1 2 3 3 5 | 6 7 8 12 25
Median = 5.5 copiers

Terri: 3 3 3 14 16 | 18 18 24 32 35
Median = 17 copiers

d *Sam*:

$$\bar{x} = \frac{1+2+3+3+5+6+7+8+12+25}{10}$$
$$= 7.2 \text{ copiers}$$

Terri:

$$\bar{x} = \frac{3+3+3+14+16+18+18+24+32+35}{10}$$
$$= 16.6 \text{ copiers}$$

e The median is best for comparing the performances because there is an outlier of 25 in Sam's data.

f Terri is the better salesperson because his median number of copiers sold is much higher than Sam's.

Question 8

a $\text{Mean} = \dfrac{0\times 2+1\times 2+\ldots+9\times 1}{2+2+1+3+2+1+1}$

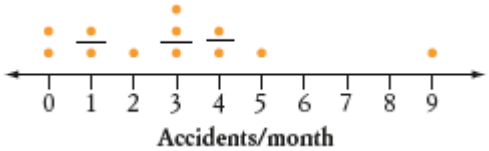
$= \dfrac{35}{12}$

$= 2.92$

Mode = 3 (the highest frequency)

Median = 3

b



$Q_1 = 1,\ Q_3 = 4$

$\text{Interquartile range} = 4-1$

$= 3$

$Q_1 - 1.5\times \text{IQR} = 1-1.5\times 3$

$= -3.5$

$Q_3 + 1.5\times \text{IQR} = 4+1.5\times 3$

$= 8.5$

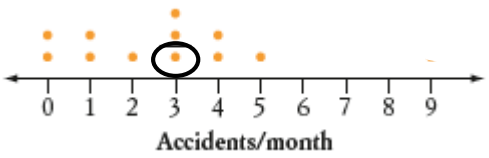
$\therefore$ 9 is the outlier number of accidents, since it is greater than 8.5.

c $\text{Mean} = \dfrac{0\times 2+1\times 2+\ldots+5\times 1}{2+2+1+3+2+1}$

$= \dfrac{26}{11}$

$= 2.36$

Mode = 3 (the highest frequency)



Median = 3

d The outlier has no effect on the mode or median but it increases the mean.

Question 9

a $$\text{Mean} = \frac{13\,500 \times 2 + 46\,500 \times 3 + \ldots + 73\,800}{10}$$

$$= \$36\ 570$$

13 500 13 500 28 500 28 500 34 200 | 34 200 46 500 46 500 46 500 73 800

Median = $34 200

Mode = $46 500 (since three employees have this salary)

b Rupert would use the mode to make the salaries seem higher because the mode is greater than both the mean and the median for this set of data.

c The median best represents the average wage for an employee at the bookstore because it is not affected by the one high salary of $73 800.

Exercise 10.05 Cumulative frequency graphs

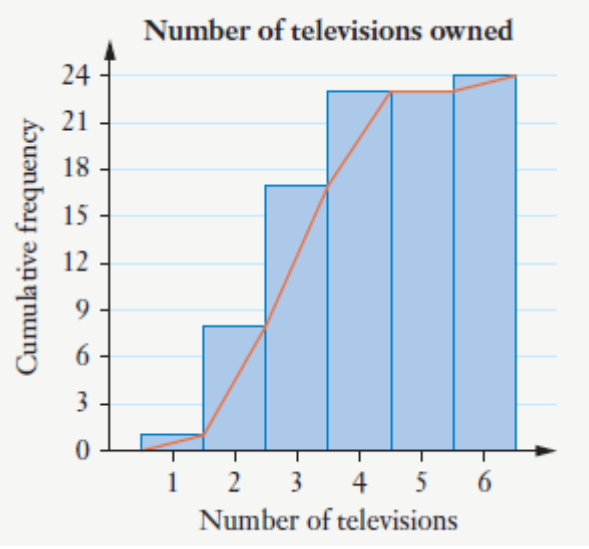
Question 1

a Add the frequencies as you go

No. of TVs	Frequency	Cumulative frequency
1	1	1
2	7	1 + 7 = 8
3	9	8 + 9 = 17
4	6	17 + 6 = 23
5	0	23 + 0 = 23
6	1	23 + 1 = 24

There are 24 scores, so the median will be the average of the 12th and 13th scores.
It can be seen from the table that the 9th to 17th scores are all 3, so both these scores will be 3.
∴ The median will be 3.

b Choose a suitable graph and plot the frequencies at the end of each interval.



c **i** On the vertical axis draw a line from the halfway mark (12) across to the ogive (line graph) to find the median on the horizontal axis.
Read down to the horizontal axis, this is clearly three TV sets.

ii On the vertical axis draw a line across from the quarter mark (6) to find Q_1 on the horizontal axis. This is 2 TV sets.
On the vertical axis draw a line across from the three quarter mark (18) to find Q_3 on the horizontal axis. This is 4 TV sets.
∴ Interquartile range = 4 – 2 = 2.

Question 2

a Look at the vertical value at the end of the graph to find there were 25 vehicles in the survey.

b On the vertical axis draw a line across from the halfway mark (12.5, midway between 10 and 15) to find the median of 45 km/h.

c On the vertical axis draw a line across from the quarter mark (6.25) to find Q_1 is about 27 km/h. Note: the unmarked line midway between 5 and 10 is 7.5, and 6.25 is exactly halfway between this line and 5.
On the vertical axis draw a line across from the three-quarter mark (18.75) to find Q_3 is about 56 km/h. Note: the unmarked line midway between 15 and 20 is 17.5, and 18.75 is exactly halfway between this line and 20.
Interquartile range $\approx 56 - 27 = 29$ km/h.
Your answers may vary from this by 1 or 2 km/h.

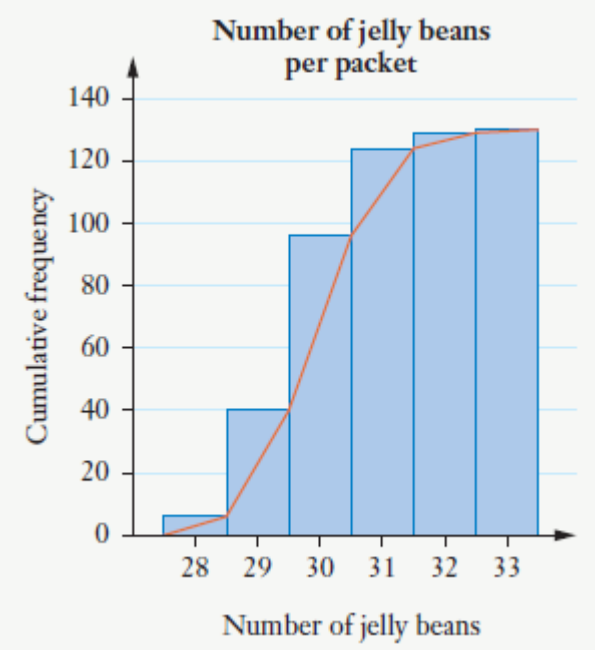
d The 9th decile is 90% of the vertical axis, so want 90% of 25 = 22.5.
Draw a line across from 22.5 and read down to the horizontal axis.
The 9th decile is about 65 km/h.

Question 3

a Add the frequencies as you go.

No. of jelly beans	Frequency	Cumulative frequency
28	6	6
29	34	6 + 34 = 40
30	56	40 + 56 = 96
31	28	96 + 28 = 124
32	5	124 + 5 = 129
33	1	129 + 1 = 130

b Choose a suitable graph and plot the frequencies at the end of each interval.



i On the vertical axis draw a line across from the halfway mark (65) to find the median of 30 jelly beans on the horizontal axis.

ii On the vertical axis draw a line across from the quarter mark (32.5) to find Q_1 is 29 jelly beans on the horizontal axis.
On the vertical axis draw a line across from the three quarter mark (97.5) to find Q_3 is 31 jelly beans on the horizontal axis.
Interquartile range = 31 – 29 = 2.

iii 4th decile is 40% so, 40% of 130 = 52
On the vertical axis draw a line across from 52 to find D_4 is 30 jelly beans on the horizontal axis.

Question 4

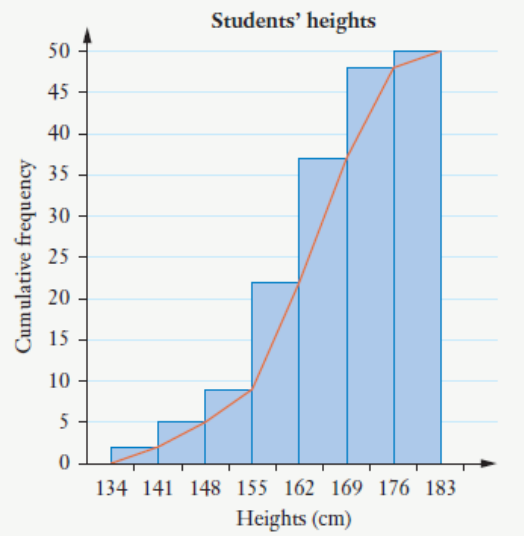
a As the data is continuous, the class centre will be the middle of the heights, so the first will be (134 + 141) ÷ 2 = 137.5 etc. Add the frequencies as you go to calculate the cumulative frequencies.

Height (cm)	Class centre	Frequency	Cumulative frequency
134–<141	137.5	2	2
141–<148	144.5	3	2 + 3 = 5
148–<155	151.5	4	5 + 4 = 9
155–<162	158.5	13	9 + 13 = 22
162–<169	165.5	15	22 + 15 = 37
169–<176	172.5	11	37 + 11 = 48
176–<183	179.5	2	48 + 2 = 50

b The modal class is the class with the highest frequency, 162–<169.

c The median class will be at the middle two scores, the 25th and 26th scores. These are both in 162–<169, which is the median class.

d Choose a suitable graph and plot the frequencies at the end of each interval.



i Draw a line across from the halfway mark (25) to find the median of about 164 cm on the horizontal axis.

ii Draw a line across from the quarter mark (12.5) to find Q_1 is about 158 on the horizontal axis.
Draw a line across from the three quarter mark (37.5) to find Q_3 is about 171 cm on the horizontal axis.
Interquartile range = 171 – 158 = 13 cm

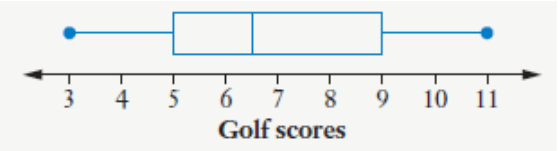
iii 70% of 50 = 35
Draw a line across from 35 to find D_7 is about 168 cm on the horizontal axis.

Exercise 10.06 Box plots

Question 1

a Place the 18 scores in order. The median splits the data into two equal parts.
The quartiles will be the middle score in each half.
3 4 4 5 5 5 5 6 6 | 7 7 8 8 9 9 9 10 11
The median = (6 + 7) ÷ 2 = 6.5
$Q_1 = 5$, $Q_3 = 9$
The smallest score is 3 and the largest is 11.
Five number summary: 3, 5, 6.5, 9, 11

b Draw a number line which at least includes the lowest and highest numbers, 3 and 11.
Label it and draw the box plot above it.



Question 2

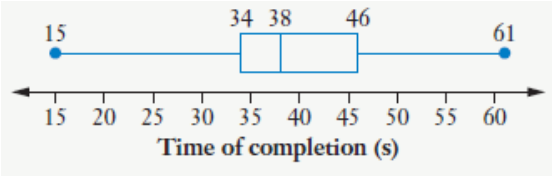
Place the 15 times in order. In this case, two rows works well. The median is the middle value and the quartiles are the middle times in each half.

15 26 30 34 35 37 37 38 43 44 45 46 48 52 61

Median = 38, $Q_1 = 34$ and $Q_3 = 46$

The smallest score is 15 and the largest score is 61.

Five number summary: 15, 34, 38, 46, 61.
Draw a number line which at least includes the lowest and highest numbers, 15 and 61. Label it and draw the box plot above it.



Question 3

Count the leaves. There are 24 ages represented.

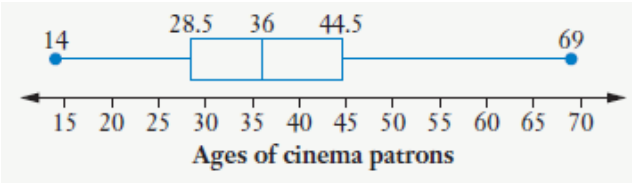
The smallest age is 14 and the largest is 69.

The median will split the data into two equal parts with 12 ages in each, so the median will be the average of the 12th and 13th ages.
Median = (35 + 37) ÷ 2 = 36

The quartiles will split each half into two equal parts with 6 ages in each. The first quartile will be the average of the 6th and 7th ages.
Q_1 = (28 + 29) ÷ 2 = 28.5

The third quartile will be the average of the 18th and 19th ages. (You could also count the 6th and 7th back from the end.) Q_3 = (44 + 45) ÷ 2 = 44.5

Five number summary: 14, 28.5, 36, 44.5, 69
Draw a number line which at least includes the lowest and highest numbers, 14 and 69.
Label it and draw the box plot above it.



Question 4

Read the numbers from the box plot.

a The median is the line in the box, 20.

b The quartiles are the ends of the box, 25 and 13.
∴ Interquartile range = 25 – 13 = 12

c The end of the whisker is 3.

d This is the bottom half of the box. This is between Q_1 and the median.
So, this is one quarter, 25%, of the graph.
∴ 25% of 60 = 15.

e This is from the middle of the box down. This is from the median down.
So, this is one half, 50%, of the graph.
∴ 50% of 60 = 30

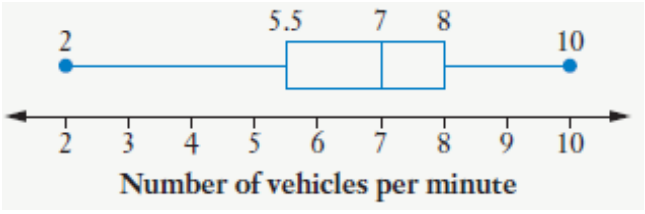
Question 5

a Count the dots. There are 20 ages represented.
The smallest age is 2 and the largest is 10.
The median will split the data into two equal parts with 10 ages in each, so the median will be the average of the 10th and 11th ages. Both are 7.
∴ Median = 7

The quartiles will split each half into two equal parts with 5 ages in each.
The first quartile will be the average of the 5th and 6th ages.
∴ $Q_1 = (5 + 6) \div 2$
$= 5.5$

The third quartile will be the average of the 15th and 16th ages. (You could also count the 5th and 6th back from the end.) Both are 8.
∴ $Q_3 = 8$

Five number summary: 2, 5.5, 7, 8, 10
Draw a number line which at least includes the lowest and highest numbers, 2 and 10.
Label it and draw the box plot above it.



b Different answers possible. The dot plot shows individual scores and any clusters and outliers, while the box-and-whisker plot identifies the quartiles, including the median.

Question 6

a The outlier must be clearly separate from the rest of the data. In this case, it is the dot on the end, 75, which represents \$75 000.

b The median is the line in the box, 41, which represents \$41 000.

c Range = highest score – lowest score
$= 58 - 17 = 41$
$\therefore$ This represents \$41 000.

d Range = highest score – lowest score
$= 75 - 17 = 58$
$\therefore$ This represents \$58 000.

e This represents the box. The lower part of the box is \$28 000 and the upper part is \$51 000, so the middle 50% of staff wages was between \$28 000 and \$51 000.

f \$28 000 is the lower quartile, or lower 25%, so 25% will earn below \$28 000.

Question 7

a Reading from the boxplot, the five number summaries are:
Geography: 45, 55, 60, 75, 90
Mod History: 40, 60, 70, 75, 85

b **i** Geography: Range $= 90 - 45 = 45$
Mod History: Range $= 85 - 40 = 45$

ii Geography: IQR $= 75 - 55 = 20$
Mod History: IQR $= 75 - 60 = 15$

c Geography: Median = 60
Mod History: Median = 70

d Modern History, as range for both subjects is the same but IQR for History is lower.

e **i** This is between the median and Q_3
$\therefore 75\% - 50\% = 25\%$ of students
$\therefore 25\%$ of $40 = 10$ students

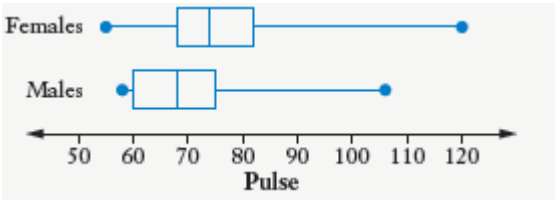
ii This is between Q_1 and Q_3
$\therefore 75\% - 25\% = 50\%$ of students
$\therefore 50\%$ of $40 = 20$ students

f Modern History, as 30 students scored above 60 compared with 20 students in Geography.

Question 8

a **Male**: Place the 15 scores in order. The median splits the data into two equal parts.
The quartiles will be the middle score in each half.
58 59 60 [60] 63 64 66 [68] 69 70 70 [75] 84 88 106
The median = 68, $Q_1 = 60$, $Q_3 = 75$
The smallest score is 58 and the largest is 106.
$\therefore$ Five number summary: 58, 60, 68, 75, 106

Female: Place the 15 scores in order. The median splits the data into two equal parts.
The quartiles will be the middle score in each half.
55 59 60 [68] 71 73 74 [74] 75 77 79 [82] 82 91 120
The median = 74, $Q_1 = 68$, $Q_3 = 82$
The smallest score is 55 and the largest is 120.
$\therefore$ Five number summary: 55, 68, 74, 82, 120.



b Male: Range $= 106 - 58 = 48$ IQR $= 75 - 60 = 15$
Female: Range $= 120 - 55 = 65$ IQR $= 82 - 68 = 14$

c Range of females is higher due to the outlier of 120.
The IQRs are similar, so no significant difference in the spread of both groups.

d Males: 25% of males had pulses above 75, while 50% of female pulses were above 74.

Question 9

Physics: median = 50
$\therefore$ 50% scored above 50.

Chemistry: $Q_1 = 50$
$\therefore$ 75% scored above 50

$\therefore$ Number of students = 75% of 48
$= 0.75 \times 48$
$= 36$
The same number of students scored a mark that was 50 or above in both. So, we know that 50% of the number of students who did Physics is 36.
$\therefore$ 72 students did the Physics exam.
$\therefore$ D

Question 10

a **Dominant**: Place the 15 scores in order. The median splits the data into two equal parts. The quartiles will be the middle score in each half.

0.29 0.31 0.31 0.32 0.33 0.36 0.37 0.38

0.38 0.41 0.43 0.44 0.46 0.50 0.52

The median = 0.38, $Q_1 = 0.32$, $Q_3 = 0.44$

The smallest score is 0.29 and the largest is 0.52.

∴ Five no. summary: 0.29, 0.32, 0.38, 0.44, 0.52

Non-dominant: Place the 15 scores in order. The median splits the data into two equal parts. The quartiles will be the middle score in each half.

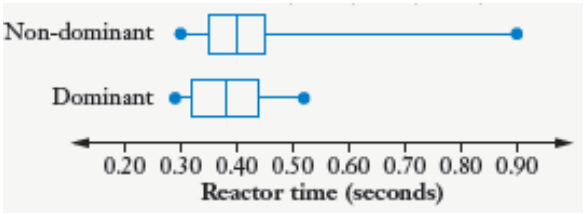
0.30 0.34 0.34 0.35 0.35 0.38 0.38

0.40 0.41 0.41 0.45 0.45 0.48 0.50 0.9

The median = 0.40, $Q_1 = 0.35$, $Q_3 = 0.45$

The smallest score is 0.30 and the largest is 0.9.

∴ Five no. summary: 0.30, 0.35, 0.40, 0.45, 0.9



b Dominant: Range $= 0.52 - 0.29 = 0.23$ IQR $= 0.44 - 0.32 = 0.12$

Non-dominant: Range $= 0.9 - 0.3 = 0.6$ IQR $= 0.45 - 0.35 = 0.10$

c If outlier of 0.9 is omitted, the spread of both groups is similar with dominant hand having slightly better (shorter) reaction times.

Exercise 10.07 Standard deviation

Question 1

a Mean = Total of scores ÷ Number of scores
$= (3+0+4+...+2+2) \div 8$
$= 16 \div 8$
$= 2$

b Use your calculator. 1.322 … ≈ 1.3

Question 2

a Mean = Total of scores ÷ No. of scores
$= (6+0+3+...+7+1) \div 12$
$= 27 \div 12$
$= 2.25$

b Use your calculator. 3.0310 … ≈ 3.03

c $\bar{x} + \sigma \approx 2.25 + 3.03 = 5.28$
$\bar{x} - \sigma \approx 2.25 - 3.03 = -0.78$

d This will be all scores from 0 up to 5. (Note that –1 will not be included).
These are 0, 3, 5, 0, 3, 1.
∴ There are six scores.

e $\frac{6}{12} \times 100\% = 50\%$

Question 3

a Mean = Total of scores ÷ No. of scores
$= (60+73+...81) \div 12$
$= 900 \div 12$
$= 75$

b s = 9.8442…
≈ 9.84

Question 4

a Mean = Total of scores ÷ No. of scores

$= (540 + 510 + ... + 1080) \div 9$

$= 7520 \div 9$

$= 835.55...$

$\approx \$836$

b \$277.41 …≈ \$277

Question 5

a Include an *fx* column.

Score (*x*)	**Frequency (*f*)**	***fx***
0	6	0
1	7	7
2	8	16
3	10	30
4	9	36
5	5	25
6	5	30
Total:	50	144

Mean = 144 ÷ 50 = 2.88

b 1.807 … ≈ 1.8

c $\bar{x} + \sigma \approx 2.88 + 1.86 = 4.74$

$\bar{x} - \sigma \approx 2.88 - 1.86 = 1.02$

So, any scores between 1.02 and 4.74, i.e. scores of 2, 3 and 4. Add the frequencies of 2, 3 and 4.

∴ 8 + 10 + 9 = 27 scores within one standard deviation.

d $\frac{27}{50} \times 100\% = 54\%$

(Note: This may be smaller than the expected two thirds because the score 1 is very narrowly excluded. If it was included there would be 34 scores, which is exactly 68% of the scores.)

Question 6

a Mean = Total of scores ÷ No. of scores

$= (1\times 2 + 1\times 4 + 3\times 5 + ... + 1\times 10) \div 20$

$= 136 \div 20$

$= 6.8$

b $1.860\ldots \approx 1.86$

c $\bar{x} + \sigma \approx 6.8 + 1.86 = 8.66$

$\bar{x} - \sigma \approx 6.8 - 1.86 = 4.94$

So, the scores between these values, i.e. 5, 6, 7, 8. Add the frequencies of 5, 6, 7 and 8.

$\therefore$ 3 + 2 + 5 + 5 = 15 scores.

d $\frac{15}{20} \times 100\% = 75\%$

Question 7

a Average the weekly wages for each wage group, i.e. $(500 + 600) \div 2 = 550$

$\therefore$ \$550, \$650, \$750, \$850, \$950, \$1050

b You can treat the data as a continuous variable. Add a column for *fx*.

Wage ($)	**Class centre (*x*)**	***f***	***fx***
\$500–<\$600	\$550	7	\$3850
\$600–<\$700	\$650	20	\$13 000
\$700–<\$800	\$750	36	\$27 000
\$800–<\$900	\$850	17	\$14 450
\$900–<\$1000	\$950	11	\$10 450
\$1000–<\$1100	\$1050	3	\$3150
	Total:	94	\$71 900

i Mean = \$71 900 ÷ 94

= \$764.8936...

≈ \$764.89

ii Use your calculator. 119.3577... ≈ \$119.36

Question 8

a Use your calculator:

Males: $\bar{x} = 175.5$

$s = 7.868... \approx 7.87$

Females: $\bar{x} = 164.58... \approx 164.6$

$s = 7.242... \approx 7.24$

b Males are significantly taller than females (difference in means is 10.9).
However, the spread of heights of females is slightly less than the heights of the males.

Question 9

a Test 1: $\bar{x} = 65.95$ $\sigma = 14.361... \approx 14.36$

Test 2: $\bar{x} = 71.55$ $\sigma = 14.451... \approx 14.45$

b The standard deviations are nearly equal, indicating the spread for both tests is the same but the mean for test 2 is 5.6 marks better than the mean for test 1.

c Because the mean for test 2 is significantly higher, the students performed better in test 2.

Question 10

a Men: $\bar{x} = 340.55$ $\sigma = 436.028... \approx 436.03$

Women: $\bar{x} = 399.65$ $\sigma = 559.580... \approx 559.58$

b Men: $\bar{x} = 224.222... \approx 224.22$

$\sigma = 269.114... \approx 269.11$

Women: $\bar{x} = 234.055... \approx 234.06$

$\sigma = 248.208... \approx 248.21$

c Women make longer calls than men (by approx. 60 s) and the spread of the time of their calls is much more than the time for calls made by men. However, if the outliers are excluded, there is no significant difference, as the mean time of calls for women is only 10 s longer than that for men. Also, the spread of times for calls is now more for men (their standard deviation over 20 more than that for women) but again it is not significant.

Questions 11 – 17

Teacher to check.

Exercise 10.08 The shape of a distribution

Question 1

The data are negatively skewed since the tail points to the left, and there is one mode, which is 8.
The correct answer is **C**.

Question 2

a **i** The tail points to the left, so the data are negatively skewed.

ii Unimodal, peak at 6.

iii Cluster from 6 to 8.

b **i** The tail points to the right so the data are positively skewed.

ii Multimodal

iii No clusters

c **i** The data are positively skewed because the tail points to the right.
(Hint: turn the stem-and-leaf plot on its side so stems are at the bottom of the page.)

ii Unimodal, peak at 16.

iii Clustering in the 10s and 30s.

d **i** The data are negatively skewed because the tail points to the left.

ii Multimodal, peaks at 12, 14, 16 and 18.

iii No clusters

e **i** The data are symmetrical.

ii Multimodal, modes at 55, 65 and 85.

iii Clustering in the 60s

f **i** The data are not symmetrical or skewed.

ii Bimodal, peaks at 15–20 and 30–35.

iii Clustering at 10 to 20 and 30 to 40.

Question 3

a Count the leaves to find out how many Oztel stores were surveyed.
Number of stores surveyed = 36

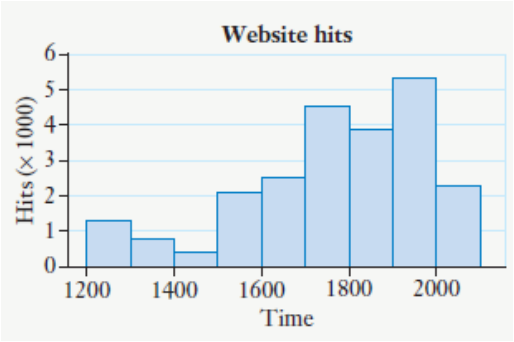
b The tail points to the left, so the data are negatively skewed.
(Hint: turn the stem-and-leaf plot on its side so stems are at the bottom of the page.)

c There is clustering in the 70s and 80s, where the scores are bunched close together.

d The mode is 88.

Question 4

a



b The data are negatively skewed with a peak at 1900–2000 and a cluster from 1700 to 2000.

c A possible reason for the skewness of the data is that more visits to Internet sites occur outside of working or school hours.

Question 5

Each score in set X occurs with the same frequency, so set X does not have a mode.
∴ The correct answer is B.

Question 6

a

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

b The tail points to the right so the data are positively skewed.
(Hint: turn the stem and leaf plot on its side so stems are at the bottom of the page.)
There is a cluster from 15 years to 19 years and the data has a mode of 15 (so a peak at 15).
With the data grouped in class intervals of 10, as displayed in the stem-and-leaf plot, the stem of 1 corresponds to a peak.

Question 7

a The tail points to the right, so the data displayed in the dot plot are positively skewed.

b The mode is 3 accidents/month. It has the highest frequency.

c $\bar{x} = \dfrac{\sum fx}{\sum f}$

$= \dfrac{0 \times 2 + 1 \times 2 + ... + 9 \times 1}{12} = \dfrac{35}{12} = 2.9$

The mean is 2.9 accidents/month, which is very close to the mode of 3, but slightly less.

Question 8

a

	11BS1	**11BS2**
Mean	$\dfrac{1177}{19} \approx 61.9$	65
Median	59	66.5
Mode	47, 56, 57	64, 71

b First find the 5 number summary for each class.

11BS1.

There are 19 results. So this breaks the data into two groups of 9 scores with one in the middle. The median is $\therefore$ the 10th score. The lower quartile is the median of the bottom 9 scores, i.e. the 5th score. The upper quartile is the median of the top 9 scores, i.e. the 15th score (may be easier to count backward to the 5th from the end).

So, five number summary is: 38, 53, 59, 72, 85.

$\therefore$ Range $= 85 - 38 = 47$, IQR $= 72 - 53 = 19$ and using a calculator $\sigma = 12.432... \approx 12.4$

11BS2

There are 20 results. So this breaks the data into two groups of 10 scores. The median is the middle, $\therefore$ the average of the 10th and 11th scores. The lower quartile is the median of the bottom 10 scores, i.e. the average of the 5th and 6th scores. The upper quartile is the median of the top 10 scores, i.e. the average of the 15th and 16th scores.

Median $= (66 + 67) \div 2 = 66.5$

$Q_1 = (57 + 59) \div 2 = 58$, $Q_3 = (73 + 74) \div 2 = 73.5$

So, 5 number summary is: 34, 58, 66.5, 73.5, 84.

$\therefore$ Range $= 84 - 34 = 50$, IQR $= 73.5 - 58 = 15.5$ and using a calculator $\sigma = 12.521... \approx 12.5$

c 11BS1: positively skewed, 11BS2: negatively skewed

d 11BS2 achieved better results since the mean and median are both greater than the mean and median for 11BS1 (by 3.1, 7.5 respectively). Although 11BS2 had the larger range, the IQR was less and the standard deviation was similar so the spreads were approximately the same.

Question 9

a 12W: Median = 50
12X: Median = 40

b 12W: Range $= 76 - 38 = 38$
IQR $= 56 - 44 = 12$
12X: Range $= 80 - 24 = 56$
IQR $= 58 - 30 = 28$

c Both sets of data are positively skewed.

d 12W had the better results as the median score was 50 compared to 40 for 12X.
More than 75% of the students from 12W scored more than 50% of the students from 12X.

Question 10

a

Estimates		Test results
3	3	2
	4	6 9
8	5	2
5 5	6	4
8 7 7 5 3 0	7	2 3 3 3 8 9 9 9
9 9 8 7 7 6 5 3 2 2 2 0	8	0 2 5 6 6 7 7
3 2	9	1 1 1 5

b The estimates are negatively skewed, has one mode (82) and are clustered in the 80s. The test results are slightly positively skewed, have 3 modes (73, 79 and 91) and are clustered in the 70s and 80s.

c Median is the average of the 12th and 13th scores.

	Estimate	**Results**
Mean	$78.16... \approx 78.2$	$75.41... \approx 75.4$
Median	$(82 + 82) \div 2 = 82$	$(79 + 79) \div 2 = 79$
Mode	82	73, 79, 91

d There are 24 results. This breaks the data into two groups of 12 scores. The lower quartile is the median of the bottom 12 scores, i.e. the average of the 6th and 7th scores. The upper quartile is the median of the top 12 scores, i.e. the average of the 18th and 19th scores.

$Q_1 = (73 + 75) \div 2 = 74$
$Q_3 = (87 + 87) \div 2 = 87$

	Estimate	**Results**
Range	$93 - 33 = 60$	$95 - 32 = 63$
IQR	$87 - 74 = 13$	$86.5 - 72.5 = 14$
σ	$12.87... \approx 12.9$	$15.76... \approx 15.8$

e The students did overestimate their results, but only by 3 when comparing means and medians. Also, the spread of the two sets of scores are similar with the ranges (90 and 93) and IQRs (13 and 14) being very close.

Sample HSC problem

a Add up all the leaves. There were 20 patients.

$$\text{Mean} = (12 + 12 + ... + 81) \div 20$$
$$= 785 \div 20$$
$$= 39.25$$

b The median will be the average of the 10th and 11th ages.

$\therefore$ Median = (30 + 33) ÷ 2
= 31.5

c The mean, as it includes every score and there are no outliers.

d The first quartile is the average of the 5th and 6th ages.

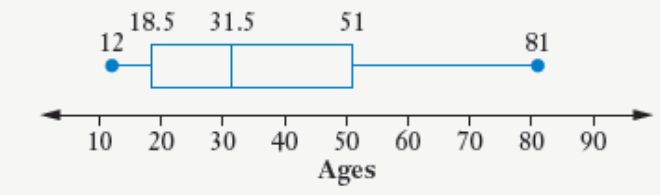
$\therefore$ Q_1 = (16 + 21) ÷ 2 = 18.5

The third quartile is the average of the 15th and 16th ages. Both are 51.

$\therefore$ Q_3 = 51

$\therefore$ Interquartile range (IQR) = 51 – 18.5
= 32.5

e Draw a number line which at least includes the lowest and highest numbers, 12 and 81. Label it and draw the box plot above it.



Test yourself 10

Question 1

a Add up all the heights.
$165 + 183 + 170 + \ldots + 175 + 179 = 2771$
Count the heights. There are 16.
$2771 \div 16 = 173.1875 \approx 173.2$ cm

b Put the heights in ascending order:
165 168 168 168 170 170 170 171 172 175 175 177 179 179 181 183
There are 16 scores. The median is the average of the 8th and 9th heights.
$\text{Median} = (171 + 172) \div 2$
$= 171.5$

c 168 and 170

Question 2

a You can treat the data as a continuous variable. For part **b**, add a column for fx.

Speed (km/h)	**Class centre (x)**	f	fx
60–<70	65	7	665
70–<80	75	13	975
80–<90	85	23	1955
90–<100	95	7	665
	Total:	50	4050

b Mean = $4050 \div 50 = 81$ km/h

c There are 50 values. The median is the average of the 25th and 26th value.
The 21st to the 43rd value were in the speed class of 80–<90.
This contains both the 25th and 26th value.
$\therefore$ The median class of speeds is 80–< 90 km/h.

d The mode is the highest frequency.
$\therefore$ The modal class is 80–< 90 km/h.

Question 3

Count the dots to check there are 20.
$\text{Mean} = (2 + 3 + 4 + 4 + \ldots + 12) \div 20 = 6.8$
The median will be the average of the 10th and 11th scores. Count from the left.
$\therefore$ Median = 7
The mode is the most common score, 7.

Question 4

a Add up all the house prices.
$642 000 + $585 000 + … + $740 000 = $5 599 000
Count the house prices. There are 8.
$5 599 000 ÷ 8 = $699 875

b Put the prices in order.
$352 000 $415 000 $585 000 $642 000
$680 000 $705 000 $740 000 $1 480 000
Median is the average of the middle two prices.
($642 000 + $680 000) ÷ 2 = $661 000

c The median, as it is not affected by the outlier of $1 480 000 as much as the mean is.

Question 5

a The mode is used for things like shoe sizes, where the most common size is required.

b The mean is used as it includes all the data and there are no outliers.

c The median would be better as there may be outliers.

Question 6

The **c**lass centre for each class is found by averaging the end values of the class intervals.
The class centres are: 13, 18, 23, 28, 33.
The total number of scores is $4+7+...+15=62$
The total of the data is $13\times4+18\times7+...+33\times15=1621$
$\therefore$ Mean $=1621\div62=26.14...$
$\therefore$ **C**

Question 7

a Over $\frac{7}{10}$, or more than 70% of students would have scored below her.

b The 78th percentile would include 78% of the scores below her. This means 100% – 78%, or 22% scored above her.

Question 8

a The interquartile range is the difference between the third (upper) quartile and the first (lower) quartile. This will be the range of the middle half (50%) of the scores.

b Put the scores in order.

48 49 49 49 50 50 50 50
50 51 51 51 51 52 53

The range = highest – lowest
= 53 – 48 = 5

The median is the middle score, 50.

The quartiles will be the middle scores of the top and bottom halves of the data, NOT including the median.

48 49 49 49 50 50 50 50
50 51 51 51 51 52 53

$Q_1 = 49$, $Q_3 = 51$

Interquartile range = 51 – 49 = 2

Question 9

a $\text{Range} = 62 - 6 = 56$

b There are 24 results. This breaks the data into two groups of 12 scores.
The lower quartile is the median of the bottom 12 scores, i.e. the average of the 6th and 7th scores.
The upper quartile is the median of the top 12 scores, i.e. the average of the 18th and 19th scores.

$Q_1 = (14 + 14) \div 2 = 14$

$Q_3 = (32 + 34) \div 2 = 33$

$\therefore IQR = 33 - 14 = 19$

Question 10

Put the values in ascending order:

\$874 \$950 \$950 \$980 \$1026 \$1140 \$1216 \$1710

There are 8 scores. This breaks the data into two groups of 4 scores. The lower quartile is the median of the bottom 4 scores. The upper quartile is the median of the top 4 scores.

$Q_1 = (950 + 950) \div 2 = 950$

$Q_3 = (1140 + 1216) \div 2 = 1178$

$\therefore IQR = 1178 - 950$

$= 228$

$Q_3 + 1.5 \times \text{IQR} = 1178 + 1.5 \times 228 = 1520$

$\therefore$ Any values above 1520 are outliers.

$\therefore$ \$1710 is an outlier for this set of data.

Question 11

a Current data set: $\text{Mean} = (4+7+...+20) \div 8 = 11.625$

$\text{Median} = (8+12) \div 2 = 10$

Data set with outlier: $\text{Mean} = (4+7+...+40) \div 9 = 14.\dot{7}$

$\text{Median} = 12$

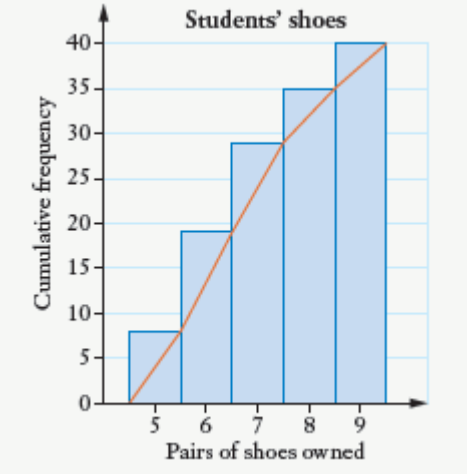
$\therefore$ Increases the mean from 11.625 to 14.778 and also increases the median from 10 to 12.

b The median, as it is less affected by the outlier.

Question 12

a Add the frequencies as you go.

Pairs of shoes	Frequency	Cumulative frequency
5	8	8
6	11	8 + 11 = 19
7	10	19 + 10 = 29
8	6	29 + 6 = 35
9	5	35 + 5 = 40



b On the graph

i Draw a line across from the halfway mark on the vertical axis (20) to find the median of 7 on the horizontal axis.

ii Draw a line across from the quarter mark on the vertical axis (10) to find Q_1 is 6 on the horizontal axis.
Draw a line across from the three quarter mark (30) to find Q_3 is 8 on the horizontal axis.
Interquartile range = 8 – 6 = 2.

iii $0.3 \times 40 = 12$
Draw a line across from 12 on the vertical scale to the line and then down to find the third decile of about 5.9.
$\therefore$ 6 pairs of shoes.

Question 13

a Looking at the last value for the cumulative frequency, 36 students completed the assignment.

b On the graph

i Draw a line across from the halfway mark on the vertical axis (18) to find the median of 4.5 marks on the horizontal axis.

ii Draw a line across from the quarter mark on the vertical axis (9) to find Q_1 is 3 marks on the horizontal axis.
Draw a line across from the three-quarter mark (27) to find Q_3 is 7 marks on the horizontal axis.
Interquartile range = 7 – 3 = 4

iii $0.6 \times 36 = 21.6$
Draw a line across from 21.6 on the vertical scale to the line and then down to find the 6th decile of about 5.4
$\therefore$ 5 marks.

iv $0.45 \times 36 = 16.2$
Draw a line across from 16.2 on the vertical scale to the line and then down to find the 45th percentile of about 4.4 marks
$\therefore$ 4 marks.

Question 14

a Read the answer from the start of the box plot: the lowest score is 0.

b The interquartile range is the difference between the two ends of the box.
Interquartile range = 10 – 5 = 5.

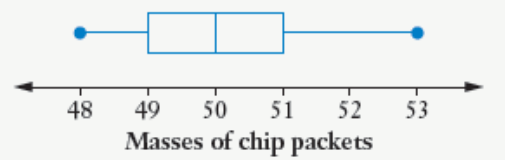
c 8 is the median, so 50% or $\frac{1}{2}$ of games.

d 5 is the first quartile, so 25% of games

Question 15

a Refer to Question **8** where we have ordered the data and found the median and the quartiles.
48 49 49 49 50 50 50 50
50 51 51 51 51 52 53
Median = 50, $Q_1 = 49$, $Q_3 = 51$
The smallest mass is clearly 48 and the largest is 53.
List these in order. 48, 49, 50, 51, 53

b Draw a number line which at least includes the lowest and highest numbers, 2 and 10.
Label it and draw the box plot above it.



Question 16

a English: Range $= 100 - 20 = 80$
IQR $= 80 - 60 = 20$
History: Range $= 100 - 10 = 90$
IQR $= 70 - 40 = 30$
$\therefore$ English has the smaller spread as it has the smaller range and IQR.

b 50% of students scored 70 or less in the English exam.
As 144 students did the English exam, this is 50% of 144 = 72 students.
$\therefore$ 72 students also scored 70 or less in the History exam.
70 or less in History was 75% of the students.
So, 75% of History students = 72.

$$75\% = 72 \text{ students}$$

$$1\% = \frac{72}{75}$$

$$100\% = \frac{72}{75} \times 100$$

$$= 96 \text{ students did the history exam}$$

Question 17

a Total = 2.00 + 1.99 + 65 + … + 2.00 = 19.98
Mean = 19.98 ÷ 10 = 1.998 cm

b Use your calculator. 0.0172 … ≈ 0.02

Question 18

a Use an *fx* column.

Pairs of shoes	Frequency	fx
5	8	40
6	10	60
7	9	63
8	5	40
9	5	45
Total:	37	248

Mean = 248 ÷ 37 = 6.7027 … ≈ 6.7

b Use your calculator. 1.312 … ≈ 1.3

Question 19

a Number of values for each can be found by adding the frequencies of each up.

Test 1: No. scores $= 1+3+3+...+2 = 30$

$\overline{x}_{\text{Test 1}} = (2\times1+3\times3+...+9\times2)\div30$

$= 169\div30$

$= 5.63...$

≈ 5.6

Test 2: No. scores $= 1+1+2+...+1 = 30$

$\overline{x}_{\text{Test 2}} = (2\times1+3\times1+...+10\times1)\div30$

$= 190\div30$

$= 6.33...$

≈ 6.3

Median: The median of each set is the average of the 15th and 16th scores.

$\text{median}_{\text{Test 1}} = (6+6)\div2$

$= 6$

$\text{median}_{\text{Test 2}} = (7+7)\div2$

$= 7$

Mode: The one with the greatest frequency.

$\text{mode}_{\text{Test 1}} = 6$

$\text{mode}_{\text{Test 2}} = 7$

b Test 1: Results are symmetrical. Test 2: The results are negatively skewed.

c **i** Test 1: Range = 9 – 2 = 7 Test 2: Range = 10 – 2 = 8

ii There are 30 results. This breaks the data into two groups of 15 scores.
The lower quartile is the median of the bottom 15 scores, i.e. the 8th score.
The upper quartile is the median of the top 15 scores, i.e. 23rd score.

Test 1: $Q_1 = 5,\ Q_3 = 7$

$\therefore IQR = 7-5 = 2$

Test 2: $Q_1 = 5,\ Q_3 = 7$

$\therefore IQR = 7-5 = 2$

iii From the calculator:

Test 1: $\sigma = 1.741... \approx 1.74$

Test 2: $\sigma = 1.795... \approx 1.80$

d Results of Test 2 are better than Test 1 as mean, median and mode of Test 2 are higher than for Test 1. The spread for both tests are similar as there is only a difference of 1 between ranges, the IQRs are equal and the standard deviations are approximately equal.

