



GuideMyGrades

— YOUR GOALS. OUR EXPERTISE. —

EDEXCEL INTERNATIONAL A LEVEL MATHS: QUESTIONS BY TOPIC

Representation & Summary of Data (S1.1)

International A Level Mathematics · S1: Statistics 1 (WST01)

22 questions · 241 marks · Mark schemes included

Past papers October 2020 - June 2025 · 2 sub-topics

 guidemygrades.com

 @Guidemygrades

 @Guidemygrades

 @Guidemygrades

 info@guidemygrades.com

 Piccadilly Business Centre, Unit C, Aldow Enterprise Park
Blackett Street, Manchester, M12 6AE

HOW TO USE THIS BOOKLET

Questions are grouped by sub-topic and run from the lowest to the highest mark tariff, then oldest paper first.

Click any row below to jump straight to that sub-topic or question.

On every question page, click ANSWER in the top-right corner to jump to its mark scheme; < QUESTION brings you back.

A gold bar in the left margin marks exactly where the question sits on the original page.

Diagrams, graphs, grids and tables are printed exactly as they appear in the original papers.

Each question is filed under its main topic; longer questions may also test skills from other topics.

QUESTION	SOURCE	MARKS	PAGE
S1.1a MEASURES OF LOCATION & SPREAD (INCL. CODING)		5 Qs · 59 marks	2
1. Q2	IAL S1 (WST01) June 2025	7	3
2. Q3	IAL S1 (WST01) October 2022	10	5
3. Q3	IAL S1 (WST01) June 2021	14	9
4. Q3	IAL S1 (WST01) June 2022	14	13
5. Q3	IAL S1 (WST01) June 2024	14	17
S1.1b DIAGRAMS: HISTOGRAMS, BOX PLOTS & STEM-AND-LEAF		17 Qs · 182 marks	21
6. Q1	IAL S1 (WST01) June 2023	8	22
7. Q1	IAL S1 (WST01) January 2024	8	26
8. Q5	IAL S1 (WST01) October 2024	8	30
9. Q7	IAL S1 (WST01) January 2025	8	34
10. Q2	IAL S1 (WST01) January 2021	9	37
11. Q3	IAL S1 (WST01) June 2023	9	39
12. Q3	IAL S1 (WST01) January 2022	10	43
13. Q1	IAL S1 (WST01) January 2023	10	47
14. Q1	IAL S1 (WST01) October 2024	10	51
15. Q1	IAL S1 (WST01) June 2022	11	55
16. Q1	IAL S1 (WST01) October 2022	11	59
17. Q2	IAL S1 (WST01) January 2025	12	63
18. Q2	IAL S1 (WST01) October 2023	13	67
19. Q1	IAL S1 (WST01) June 2024	13	71
20. Q4	IAL S1 (WST01) October 2020	14	75
21. Q3	IAL S1 (WST01) October 2021	14	79
22. Q5	IAL S1 (WST01) June 2025	14	83



QUESTIONS

22 questions · 241 marks · 2 sub-topics

SUB-TOPIC S1.1a

MEASURES OF LOCATION & SPREAD (INCL. CODING)

Sub-topic S1.1a · 5 questions · 59 marks · lowest tariff first



2. Students in class A and class B sit a statistics test.

There are 24 students in class A and 30 students in class B

The marks, x , for students in each class are summarised in the table below.

	n	$\bar{x}$	$\sum x^2$
Class A	24	47	66 876
Class B	30	b	73 826

The two classes are combined into one group of 54 students.

The mean mark for all 54 students is 45

- (a) Show that $b = 43.4$

(2)

- (b) Find the standard deviation of the marks for all 54 students.

(2)

Following moderation, each student in class B has their mark increased by 2

- (c) Without further calculations state, giving a reason in each case, the effect this will have on

- (i) the variance of the marks for class B

- (ii) the mean mark for all 54 students

- (iii) the standard deviation of the marks for all 54 students.

(3)

(Total for Question 2 is / marks)







(Total for Question 3 is 10 marks)



Leave
blank

3. A random sample of 100 carrots is taken from a farm and their lengths, L cm, recorded. The data are summarised in the following table.

Length, L cm	Frequency, f	Class mid point, x cm
$5 \leq L < 8$	5	6.5
$8 \leq L < 10$	13	9
$10 \leq L < 12$	16	11
$12 \leq L < 15$	25	13.5
$15 \leq L < 20$	30	17.5
$20 \leq L < 28$	11	24

A histogram is drawn to represent these data.

The bar representing the class $5 \leq L < 8$ is 1.5 cm wide and 1 cm high.

- (a) Find the width and height of the bar representing the class $15 \leq L < 20$ (3)

- (b) Use linear interpolation to estimate the median length of these carrots. (2)

- (c) Estimate

- (i) the mean length of these carrots, (2)

- (ii) the standard deviation of the lengths of these carrots. (3)

A supermarket will only buy carrots with length between 9 cm and 22 cm.

- (d) Estimate the proportion of carrots from the farm that the supermarket will buy. (2)

Any carrots that the supermarket does not buy are sold as animal feed.

The farm makes a profit of 2.2 pence on each carrot sold to the supermarket, a profit of 0.8 pence on each carrot longer than 22 cm and a loss of 1.2 pence on each carrot shorter than 9 cm.

- (e) Find an estimate of the mean profit per carrot made by the farm. (2)

Leave
blank

Question 3 continued

Leave
blank

Question 3 continued

Leave
blank

(Total 14 marks)

Q3



3. Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	Frequency (f)
$15 < l \leq 20$	19
$20 < l \leq 25$	35
$25 < l \leq 27$	16
$27 < l \leq 30$	15
$30 < l \leq 40$	3

A histogram is drawn to represent these data.

The bar representing logs with length $27 < l \leq 30$ has a width of 1.5 cm and a height of 4 cm.

- (a) Calculate the width and height of the bar representing log lengths of $20 < l \leq 25$ (3)

- (b) Use linear interpolation to estimate the median of l (2)

The maximum length of log Gill can use in her stove is 26 cm.

Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

- (c) Show that $x = 62$ (1)

Gill randomly selects 4 logs from the bag.

- (d) Using $x = 62$, find the probability that all 4 logs will fit into her stove. (2)

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \quad \sum y = 924 \quad \sum y^2 = 12\,862$$

- (e) Calculate

- (i) the mean of W (3)

- (ii) the variance of W (3)





3. The lengths, x mm, of 50 pebbles are summarised in the table below.

Length	Frequency
$20 \leq x < 30$	2
$30 \leq x < 32$	16
$32 \leq x < 36$	20
$36 \leq x < 40$	8
$40 \leq x < 45$	3
$45 \leq x < 50$	1

A histogram is drawn to represent these data.

The bar representing the class $32 \leq x < 36$ is 2.5 cm wide and 7.5 cm tall.

- (a) Calculate the width and the height of the bar representing the class $30 \leq x < 32$ (3)
- (b) Using linear interpolation, estimate the median of x (2)

The weight, w grams, of each of the 50 pebbles is coded using $10y = w - 20$
These coded data are summarised by

$$\sum y = 104 \qquad \sum y^2 = 233.54$$

- (c) Show that the mean of w is 40.8 (2)
- (d) Calculate the standard deviation of w (4)
- The weight of a pebble recorded as 40.8 grams is added to the sample.
- (e) Without carrying out any further calculations, state, giving a reason, what effect this would have on the value of
- (i) the mean of w
- (ii) the standard deviation of w (3)

(Total for Question 9 is 14 marks)



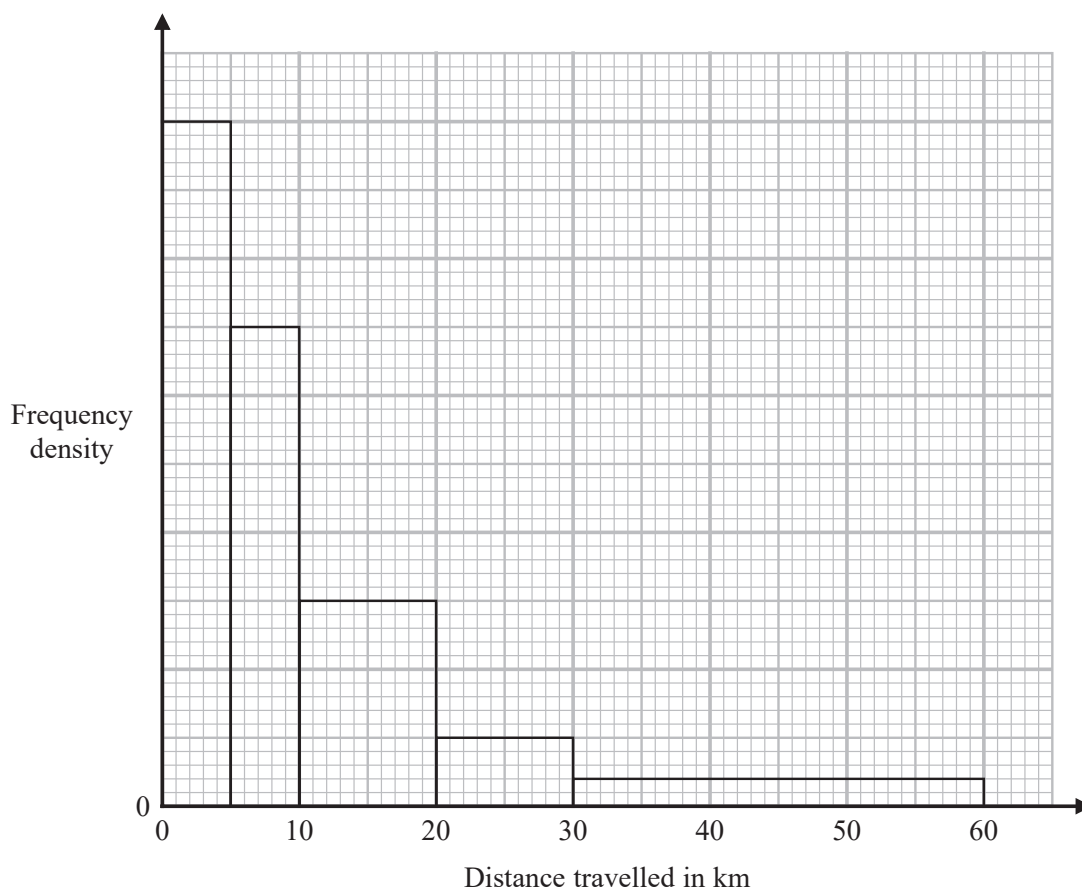
SUB-TOPIC S1.1b

DIAGRAMS: HISTOGRAMS, BOX PLOTS & STEM-AND-LEAF

Sub-topic S1.1b · 17 questions · 182 marks · lowest tariff first



1. The histogram shows the distances, in km, that 274 people travel to work.



Given that 60 of these people travel between 10 km and 20 km to work, estimate

- (a) the number of people who travel between 22 km and 45 km to work, (3)
- (b) the median distance travelled to work by these 274 people, (2)
- (c) the mean distance travelled to work by these 274 people. (3)

Question 1 continued

Lined area for writing the answer to Question 1 continued.



Question 1 continued

Lined area for writing the answer to Question 1.

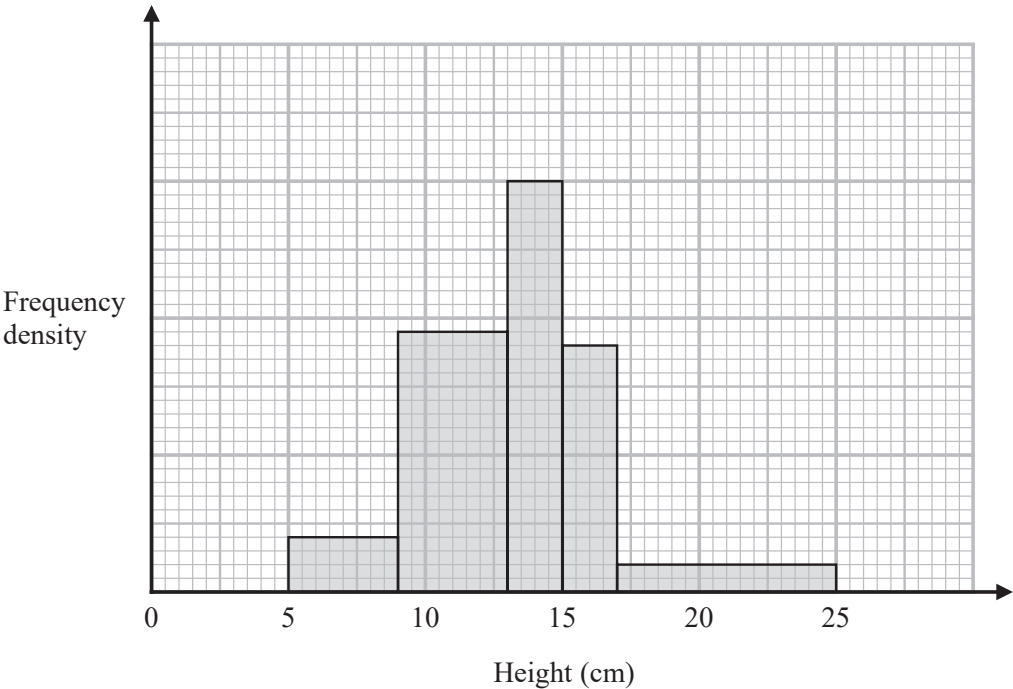


Question 1 continued

(Total for Question 1 is 8 marks)



1. The histogram below shows the distribution of the heights, to the nearest cm, of 408 plants.



(a) Use the histogram to complete the following table.

Height (<i>h</i> cm)	$5 \leq h < 9$	$9 \leq h < 13$	$13 \leq h < 15$	$15 \leq h < 17$	$17 \leq h < 25$
Frequency	32	152	120		

(2)

(b) Use interpolation to estimate the median.

(2)

The mean height of these plants is 13.2 cm correct to one decimal place.

(c) Describe the skew of these data. Give a reason for your answer.

(1)

Two of these plants are chosen at random.

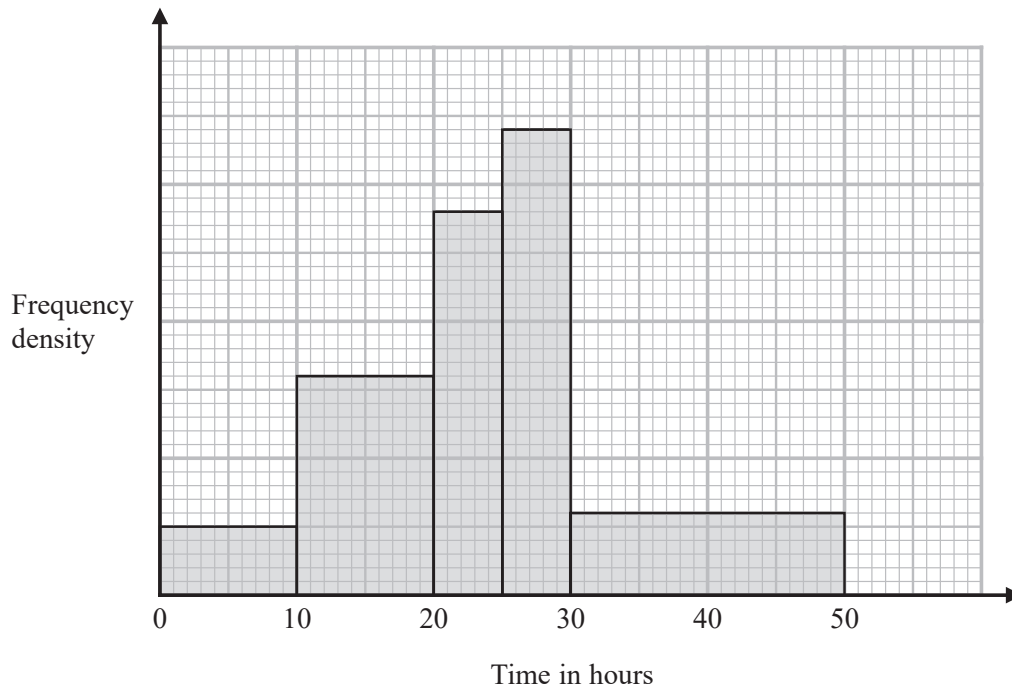
(d) Estimate the probability that both of their heights are between 8 cm and 14 cm

(3)

Question 1 continued

Question 1 continued

5.



The histogram shows the number of hours worked in a given week by a group of 64 freelance photographers.

(a) Give a reason to justify the use of a histogram to represent these data.

(1)

Given that 16 of these freelance photographers spent between 10 and 20 hours working in this week,

(b) estimate the number that spent between 12 and 24 hours working in this week.

(3)

(c) Find an estimate for the median time spent working in this week by these 64 freelance photographers.

(2)

Charlie decides to model these data using a normal distribution.

Charlie calculates an estimate of the mean to be 23.9 hours to one decimal place.

(d) Comment on Charlie's decision to use a normal distribution.
Give a justification for your answer.

(2)

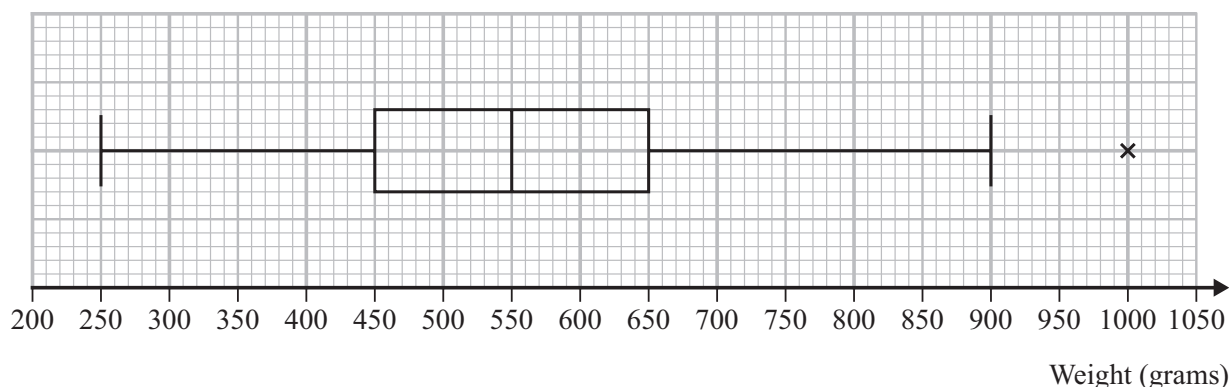
Question 5 continued



Question 5 continued

7. A farmer supplies cabbages to a local supermarket. The supermarket manager takes a random sample of 200 of these cabbages and the weight of each cabbage is recorded.

The weights of the cabbages in the sample are summarised in the following box plot.



- (a) Use linear interpolation to estimate the probability that a randomly chosen cabbage from this sample has a weight of less than 570 grams.

(2)

The supermarket manager takes the quartiles from this sample and now decides to classify any cabbage whose weight is at least $Q_3 + 1.5 \times (Q_3 - Q_1)$ as extra-large.

- (b) Find the minimum weight of a cabbage that the supermarket manager would classify as extra-large.

(1)

The supermarket manager assumes that the weights of cabbages supplied to the supermarket may be modelled by a normal distribution.

- (c) Explain whether or not the box plot supports this assumption.

(1)

The farmer records the weight of every cabbage harvested from a field and classifies any cabbage whose weight is more than 2 standard deviations above the mean as an outlier.

Assuming that the weights of the cabbages from this field may be modelled by a normal distribution,

- (d) find the probability that a randomly selected cabbage harvested from the field would be classified as an outlier using the farmer's method.

(2)

The mean weight of the cabbages harvested from the field is 560 grams.

Given that a cabbage with a weight of 1000 grams is classified as an outlier by the farmer,

- (e) find the maximum possible value of the standard deviation of the weights of cabbages harvested from the field.

(2)

Question 7 continued

Question 7 continued

Lined area for writing the answer to Question 7.

(Total for Question 7 is 8 marks)

TOTAL FOR PAPER IS 75 MARKS



Leave
blank

2. The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

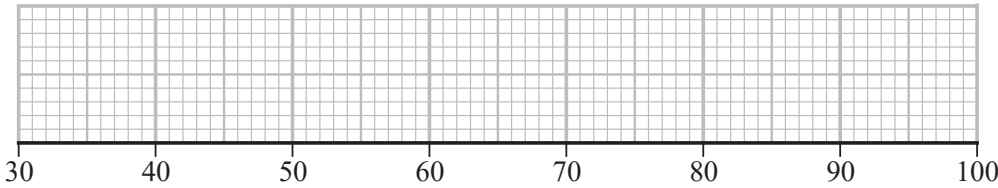
Age		Key: 4 3 is an age of 43											
4	3											(1)	
5	4											(1)	
6	2	3	5	6	8	8	8	9	9			(9)	
7	1	1	3	4	4	6	6	6	8	8	9	(11)	
8	0	0	2	7	8	8	9						(7)
9	3	7											(2)

- (a) Find the median age of the residents. (1)
- (b) Find the interquartile range (IQR) of the ages of the residents. (2)
- An outlier is defined as a value that is either
- more than $1.5 \times (\text{IQR})$ below the lower quartile or
- more than $1.5 \times (\text{IQR})$ above the upper quartile.
- (c) Determine any outliers in these data. Show clearly any calculations that you use. (3)
- (d) On the grid on page 5, draw a box plot to summarise these data. (3)

Question 2 continued

Leave
blank

Blank area for writing the answer to Question 2.



Ages

Q2

(Total 9 marks)



Question 3 continued

Lined area for writing the answer to Question 3 continued.



Question 3 continued

Lined area for writing the answer to Question 3 continued.



Leave
blank

3. The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

Key: 10 | 8 represents 108 deliveries

10	8	9									(2)	
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	a	b	c									(3)

where a, b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than $1.5 \times$ interquartile range above the upper quartile.

Given that there is only one outlier for these data,

- (a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

- (b) Find the mean number of deliveries.

(3)

- (c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

- (d) Find $P(D > 118 | X < 0)$

(2)

Question 3 continued

Leave
blank



Question 3 continued

Leave
blank



Question 3 continued

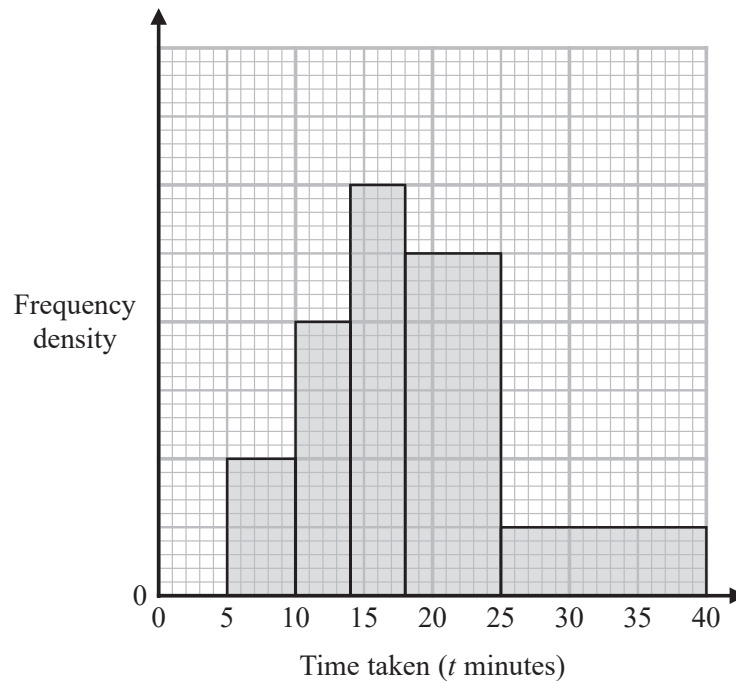
Leave
blank

(Total 10 marks)

Q3



1. The histogram shows the times taken, t minutes, by each of 100 people to swim 500 metres.



- (a) Use the histogram to complete the frequency table for the times taken by the 100 people to swim 500 metres.

Time taken (t minutes)	5 – 10	10 – 14	14 – 18	18 – 25	25 – 40
Frequency (f)	10	16	24		

(1)

- (b) Estimate the number of people who took less than 16 minutes to swim 500 metres.

(2)

- (c) Find an estimate for the mean time taken to swim 500 metres.

(2)

Given that $\sum ft^2 = 41\,033$

- (d) find an estimate for the standard deviation of the times taken to swim 500 metres.

(2)

Given that $Q_3 = 23$

- (e) use linear interpolation to estimate the interquartile range of the times taken to swim 500 metres.

(3)



Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)



1. The back-to-back stem and leaf diagram on page 3 shows information about the running times of 31 Action films and 31 Comedy films.
The running times are given to the nearest minute.
- (a) Write down the modal running time for these Action films. (1)

Some of the quartiles for these two distributions are shown in the table below.

	Action films	Comedy films
Lower quartile	121	a
Median	b	117
Upper quartile	138	c

- (b) Find the value of a , the value of b and the value of c (3)
- (c) For these Action films find, to one decimal place,
- (i) the mean running time,
 - (ii) the standard deviation of the running times.
- (You may use $\sum x = 4016$ and $\sum x^2 = 525056$ where x is the running time, in minutes, of an Action film.) (3)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

- (d) Evaluate this measure and describe the skewness for the running times of these Action films. (2)
- (e) Comment on one difference between the distribution of the running times of these Action films and the distribution of the running times of these Comedy films.
State the values of any statistics you have used to support your comment. (1)

Question 1 continued

Totals	Action films		Comedy films	Totals
(1)	0	9	2 2 3 5 8	(5)
(0)		10	3 5 6 6 8 9	(6)
(5)	9 8 6 4 2	11	0 2 4 6 7 9 9 9	(8)
(10)	9 9 8 7 6 5 4 3 1 0	12	1 2 4 6 6 7 7 7 7 8 9	(11)
(8)	8 7 7 7 5 4 2 1	13	1	(1)
(7)	7 7 6 6 4 3 1	14		(0)

Key: 0 | 9 | 2 means 90 minutes for an Action film and 92 minutes for a Comedy film

Question 1 continued

Question 1 continued

(Total for Question 1 is 10 marks)



1. The company *Seafield* requires contractors to record the number of hours they work each week. A random sample of 38 weeks is taken and the number of hours worked per week by contractor Kiana is summarised in the stem and leaf diagram below.

Stem	Leaf	
1	4 4 4 5 5 5 6 6 9 9 9	(11)
2	1 2 2 3 3 4 4 4 w 9	(10)
3	2 3 4 4 5 6 7 7 7 9	(10)
4	1 1 2 3	(4)
5	1 9	(2)
6	4	(1)

Key : 3|2 means 32

The quartiles for this distribution are summarised in the table below.

Q_1	Q_2	Q_3
x	26	y

- (a) Find the values of w , x and y

(3)

Kiana is looking for outliers in the data. She decides to classify as outliers any observations greater than

$$Q_3 + 1.0 \times (Q_3 - Q_1)$$

- (b) Showing your working clearly, identify any outliers that Kiana finds.

(2)

- (c) Draw a box plot for these data in the space provided on the grid opposite.

(3)

- (d) Use the formula

$$\text{skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

to find the skewness of these data. Give your answer to 2 significant figures.

(2)

Kiana's new employer, *Landacre*, wishes to know the average number of hours per week she worked during her employment at *Seafield* to help calculate the cost of employing her.

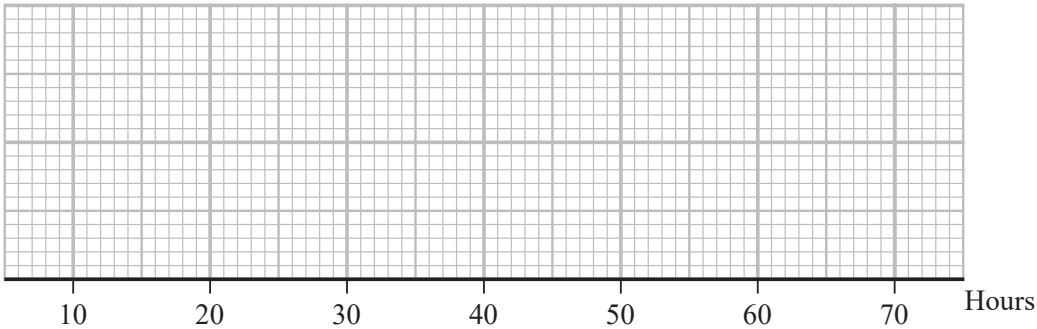
- (e) Explain why *Landacre* might prefer to know Kiana's mean, rather than median, number of hours worked per week.

(1)



Question 1 continued

Lined area for writing the answer to Question 1.



Turn over for a spare grid if you need to redraw your box plot.



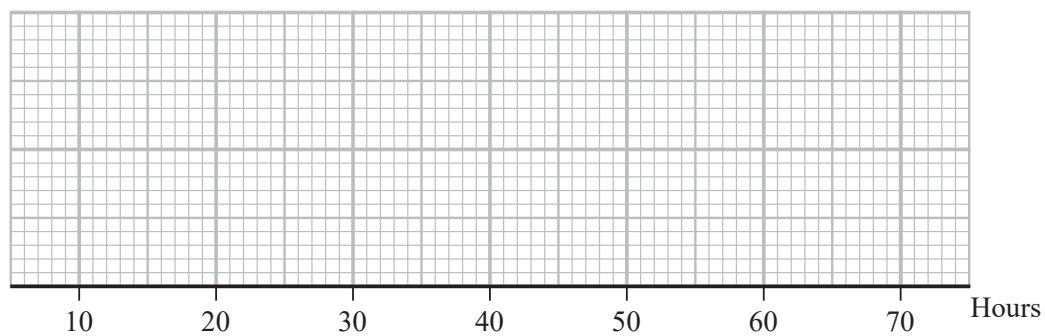
Question 1 continued

Lined area for writing the answer to Question 1 continued.



Question 1 continued

Only use this grid if you need to redraw your box plot.



(Total for Question 1 is 11 marks)

1. The stem lengths of a sample of 120 tulips are recorded in the grouped frequency table below.

Stem length (cm)	Frequency
$40 \leq x < 42$	12
$42 \leq x < 45$	18
$45 \leq x < 50$	23
$50 \leq x < 55$	35
$55 \leq x < 58$	24
$58 \leq x < 60$	8

A histogram is drawn to represent these data.

The area of the bar representing the $40 \leq x < 42$ class is 16.5 cm^2

- (a) Calculate the exact area of the bar representing the $42 \leq x < 45$ class.

(2)

The height of the tallest bar in the histogram is 10 cm.

- (b) Find the exact height of the second tallest bar.

(3)

Q_1 for these data is 45 cm.

- (c) Use linear interpolation to find an estimate for

(i) Q_2

(ii) the interquartile range.

(4)

One measure of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

- (d) By calculating this measure, describe the skewness of these data.

(2)

Question 1 continued

Question 1 continued

(Total for Question 1 is 11 marks)



2. As part of an investigation, Bobby collects a sample of 47 observations, x

The results are shown in the following stem and leaf diagram, where a is a constant.

Stem	Leaf	
2	1 2 5 7 7	(5)
3	0 2 2 3 4 5 5 5 9 9	(10)
4	0 0 1 4 4 5 7 8 8 9 9	(11)
5	3 3 5 6 7 9	(6)
6	0 2 a a a 7 8	(7)
7	1 2 3 6 8	(5)
8	0 6 7	(3)

Key: 3 | 2 means 0.32

- (a) Find the range of these observations.

(1)

- (b) Find the value of the median of these observations.

(1)

Given that the interquartile range of these observations is 0.31

- (c) find the value of a

(3)

Bobby calculates the following statistics from these observations

$$\sum x = 23.72 \quad \sum x^2 = 13.4228$$

- (d) Show that the standard deviation of these observations is 0.176 to 3 significant figures.

(2)

Bobby now collects 18 more observations, y , from the same investigation.

- (e) Using all 65 observations, the sample mean is 0.502 and the sample standard deviation is 0.204

- (i) Show that $\sum y = 8.91$

(2)

- (ii) Showing your working, calculate $\sum y^2$

(3)



Question 2 continued



Question 2 continued

Question 2 continued

(Total for Question 2 is 12 marks)



2. The weights, to the nearest kilogram, of a sample of 33 red kangaroos taken in December are summarised in the stem and leaf diagram below.

	Weight (kg)	Totals	Key: 3 2 represents 32 kg
1	6	(1)	
2	3 6	(2)	
3	2 4 6	(3)	
4	2 5 5 6 6 7 8	(7)	
5	3 4 7 7 7 8 9 9	(8)	
6	0 2 2 3 3 7 8	(7)	
7	2 8	(2)	
8	2 6	(2)	
9	4	(1)	

(a) Find

- (i) the value of the median
 (ii) the value of Q_1 and the value of Q_3
 for the weights of these red kangaroos.

(3)

For these data an outlier is defined as a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$

or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(b) Show that there are 2 outliers for these data.

(3)

Figure 1 on page 7 shows a box plot for the weights of the same 33 red kangaroos taken in February, earlier in the year.

(c) In the space on Figure 1, draw a box plot to represent the weights of these red kangaroos in December.

(4)

(d) Compare the distribution of the weights of red kangaroos taken in February with the distribution of the weights of red kangaroos taken in December of the same year. You should interpret your comparisons in the context of the question.

(3)



Question 2 continued

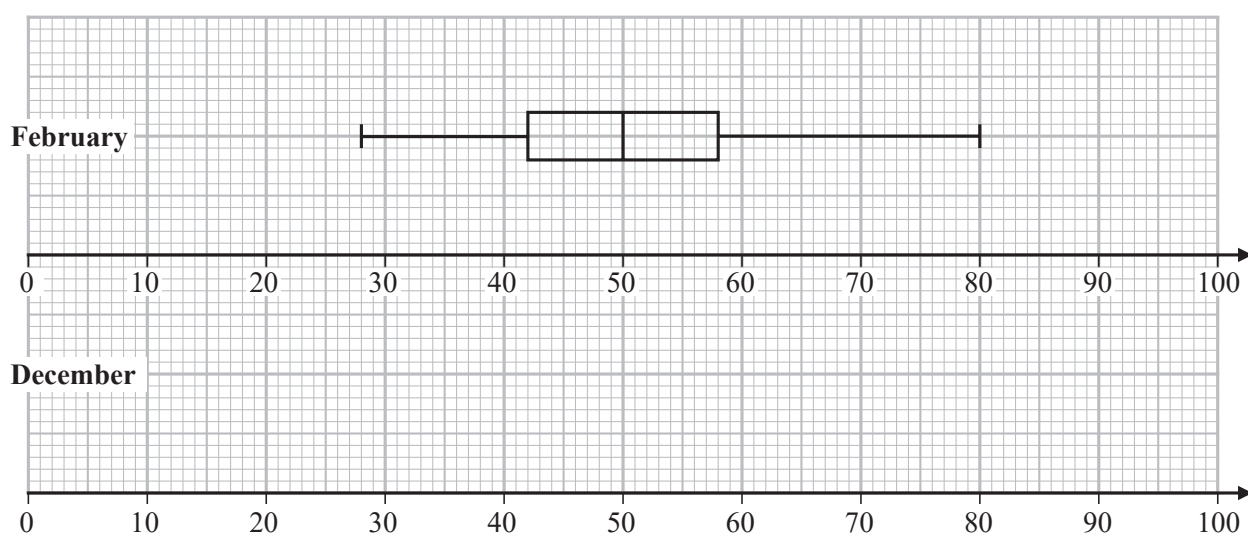


Figure 1

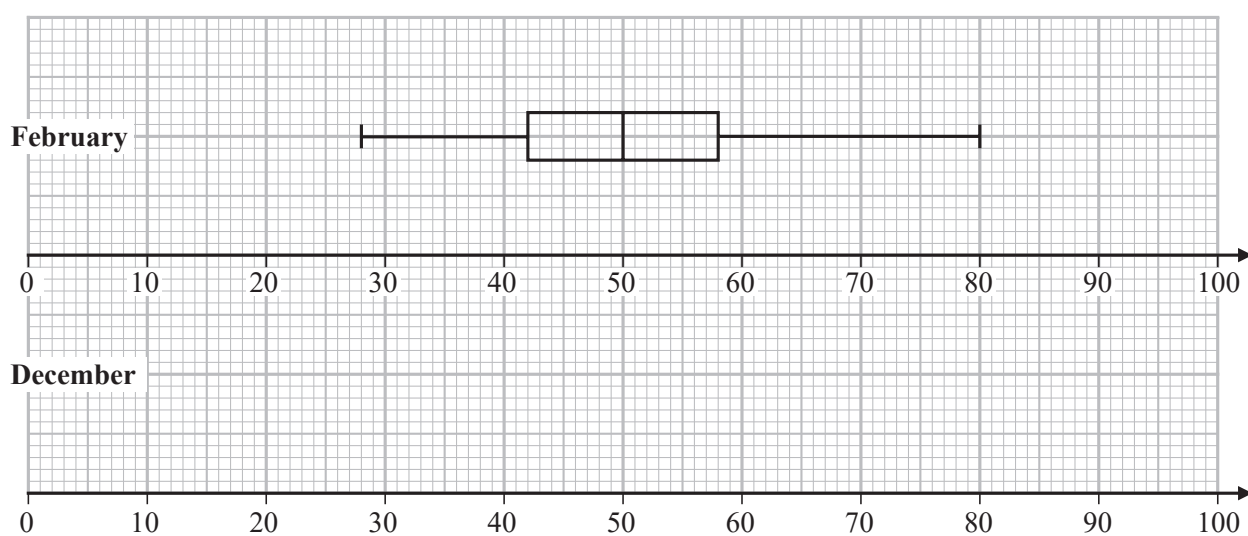
Turn over for a spare grid if you need to redraw your box plot.



Question 2 continued

Question 2 continued

Only use this grid if you need to redraw your box plot.



(Total for Question 2 is 13 marks)



1. A researcher is investigating the growth of two types of tree, Birch and Maple. The height, to the nearest cm, a seedling grows in one year is recorded for 35 Birch trees and 32 Maple trees. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Birch		Maple	Totals
(2)	9 8	2	5 7 7 8 9	(5)
(8)	9 9 9 6 5 3 1 1	3	0 2 6 6 8 9 9	(7)
(9)	9 8 8 7 6 3 1 1 1	4	1 1 1 k 7 8	(6)
(9)	7 7 7 5 4 3 2 1 0	5	0 1 2 3 4 4 4	(7)
(3)	7 6 5	6	3 4 6	(3)
(3)	6 5 4	7	0 7	(2)
(1)	5	8	0 0	(2)

Key: 5 | 6 | 3 means 65 cm for a Birch tree and 63 cm for a Maple tree

The median height that these **Maple** trees grow in one year is 45 cm.

- (a) Find the value of **k**, used in the stem and leaf diagram. (1)

- (b) Find the lower quartile and the upper quartile of the height grown in one year for these **Birch** trees. (2)

The researcher defines an outlier as an observation that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or less than $Q_1 - 1.5 \times (Q_3 - Q_1)$

- (c) Show that there is only one outlier amongst the Birch trees. (2)

The grid on page 3 shows a box plot for the heights that the Maple trees grow in one year.

- (d) On the same grid draw a box plot for the heights that the Birch trees grow in one year. (4)

- (e) Comment on any difference in the distributions of the growth of these Birch trees and the growth of these Maple trees.
State the values of any statistics you have used to support your comment. (1)

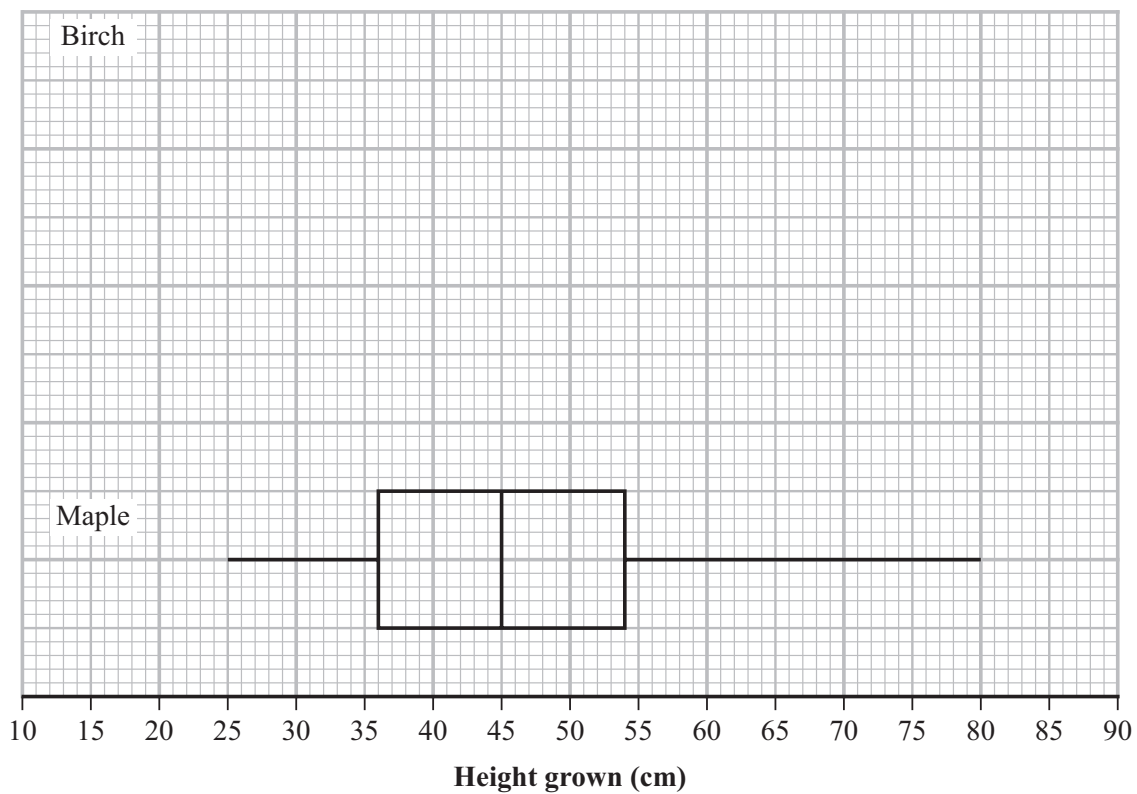
The researcher realises he has missed out 4 pieces of data for the **Maple** trees. The heights each seedling grows in one year, to the nearest cm, in ascending order, for these 4 Maple trees are 27 cm, a cm, 48 cm, $2a$ cm.

Given that there is no change to the box plot for the **Maple** trees given on page 3

- (f) find the range of possible values for a
Show your working clearly. (3)



Question 1 continued



Turn over for spare grid if you need to redraw your answer for part (d)

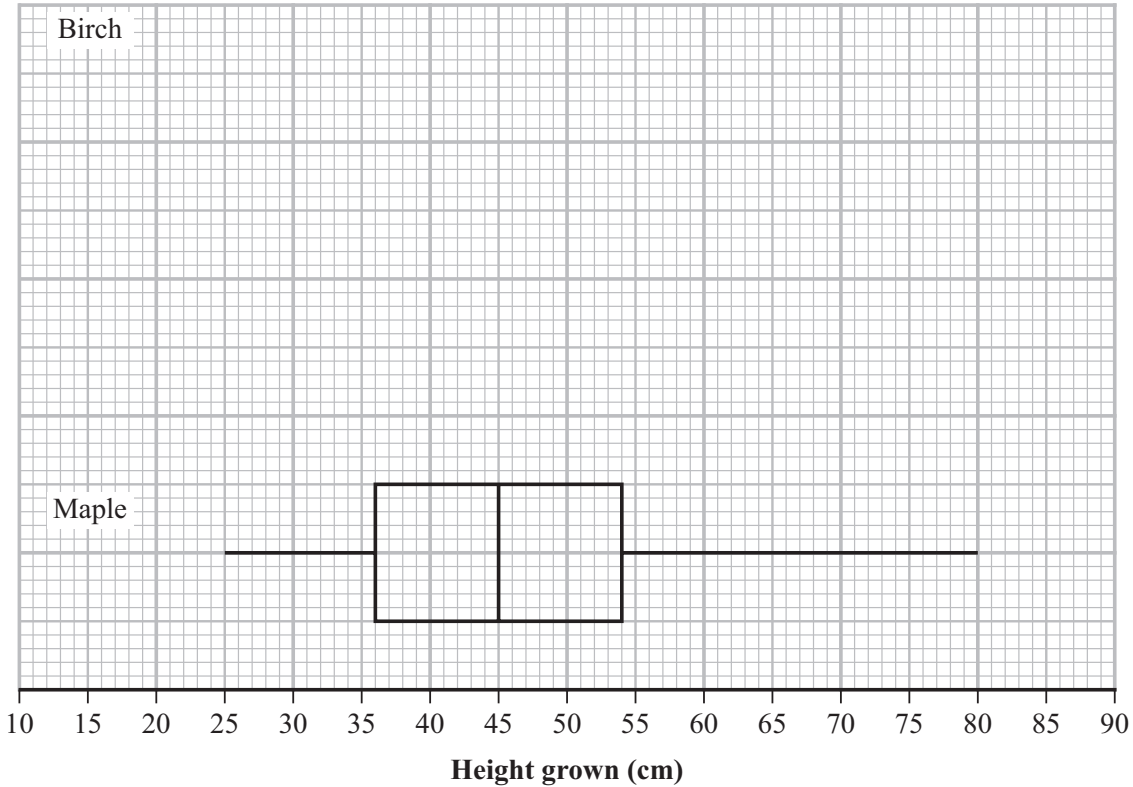


Question 1 continued

Question 1 continued

Handwriting practice area with 15 horizontal lines.

Only use this grid if you need to redraw your answer for part (d)



(Total for Question 1 is 13 marks)



Leave
blank

4. A group of students took some tests. A teacher is analysing the average mark for each student. Each student obtained a different average mark.

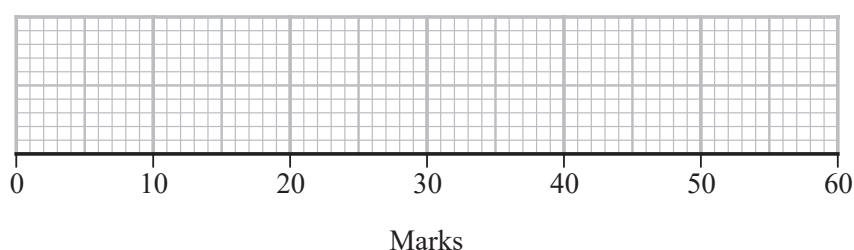
For these average marks, the lower quartile is 24, the median is 30 and the interquartile range (IQR) is 10

The three lowest average marks are 8, 10 and 15.5 and the three highest average marks are 45, 52.5 and 56

The teacher defines an outlier to be a value that is either

more than $1.5 \times \text{IQR}$ below the lower quartile or
more than $1.5 \times \text{IQR}$ above the upper quartile

- (a) Determine any outliers in these data. (4)
- (b) On the grid below draw a box plot for these data, indicating clearly any outliers. (3)

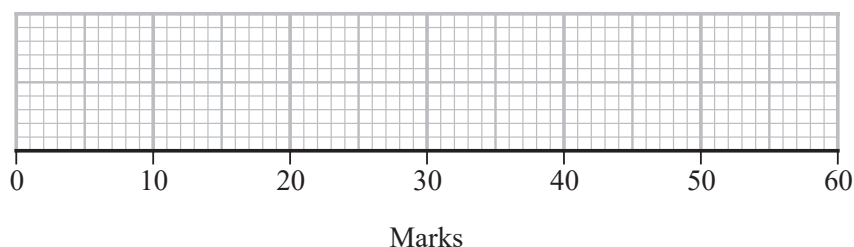


- (c) Use the quartiles to describe the skewness of these data.
Give a reason for your answer. (2)

Two more students also took the tests. Their average marks, which were both less than 45, are added to the data and the box plot redrawn.

The median and the upper quartile are the same but the lower quartile is now 26

- (d) Redraw the box plot on the grid below. (3)



- (e) Give ranges of values within which each of these students' average marks must lie. (2)

Question 4 continued

Leave
blank

Turn over for spare grids if you need to redraw your answer for part (b) or part (d).



Question 4 continued

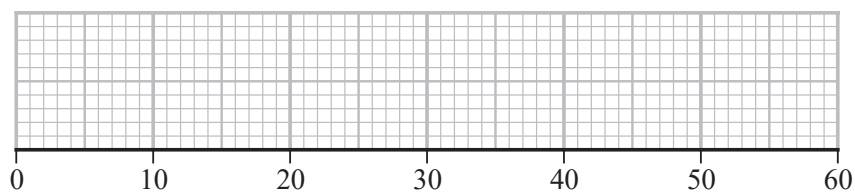
Leave
blank

Question 4 continued

Leave
blank

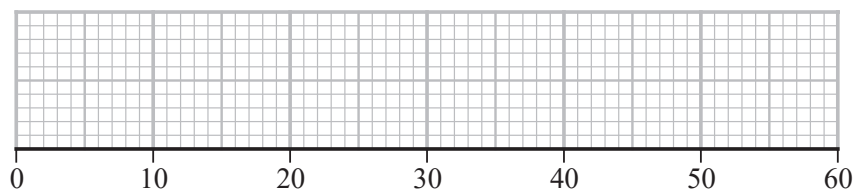
Only use these grids if you need to redraw your answer for part (b) or part (d).

Copy of grid for part (b)



Marks

Copy of grid for part (d)



Marks

(Total 14 marks)

Q4



Leave
blank

3. The stem and leaf diagram shows the ages of the 35 male passengers on a cruise.

Age		Key: 1 3 represents an age of 13 years
1	3	(1)
2	7 9	(2)
3	1 2 8 8	(4)
4	5 5 6 7 8 8 9	(7)
5	2 2 3 3 4 4 5 6 6 8	(10)
6	0 1 1 4 4 4 7	(7)
7	3 6	(2)
8	7 8	(2)

- (a) Find the median age of the male passengers. (1)
- (b) Show that the interquartile range (IQR) of these ages is 16 (2)

An outlier is defined as a value that is more than

$1.5 \times \text{IQR}$ above the upper quartile

or

$1.5 \times \text{IQR}$ below the lower quartile

- (c) Show that there are 3 outliers amongst these ages. (3)
- (d) On the grid in Figure 1 on page 9, draw a box plot for the ages of the male passengers on the cruise. (4)

Figure 1 on page 9 also shows a box plot for the ages of the female passengers on the cruise.

- (e) Comment on any difference in the distributions of ages of male and female passengers on the cruise.
State the values of any statistics you have used to support your comment. (1)

Anja, along with her 2 daughters and a granddaughter, now join the cruise.

Anja's granddaughter is younger than both of Anja's daughters.

Anja had her 23rd birthday on the day her eldest daughter was born.

When their 4 ages are included with the other female passengers on the cruise, the box plot does not change.

- (f) State, giving reasons, what you can say about
(i) the granddaughter's age
(ii) Anja's age. (3)



Question 3 continued

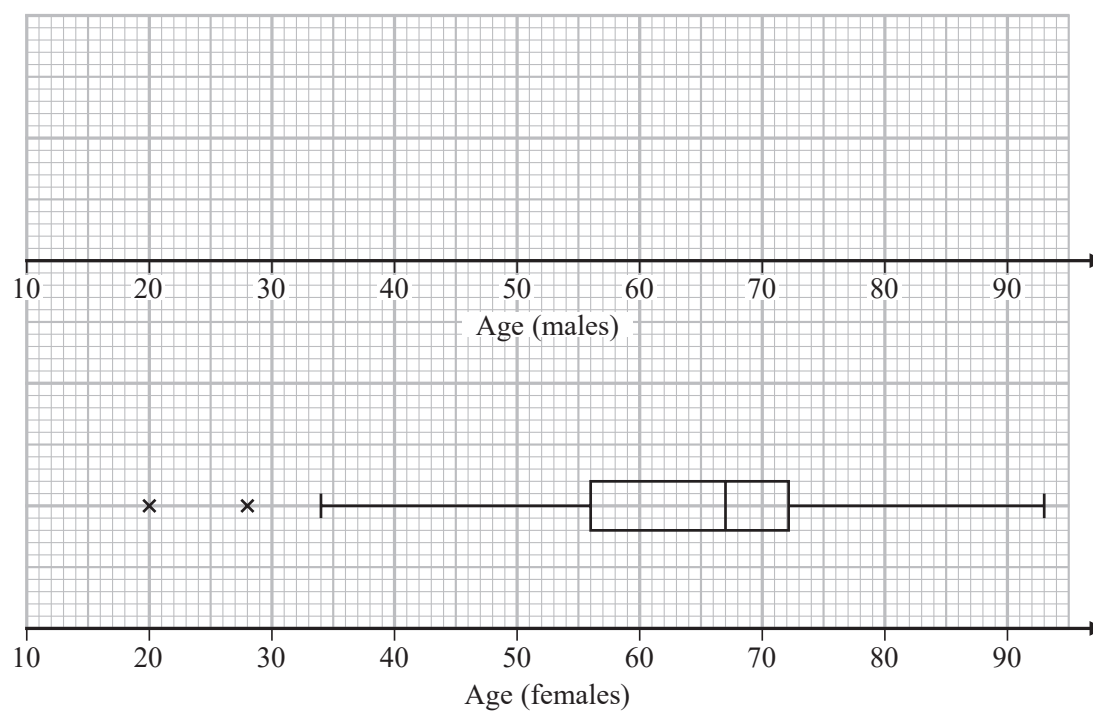
Leave
blank

Figure 1

Turn over for a spare grid if you need to redraw your box plot.



Question 3 continued

Leave
blank

Lined area for writing the answer to Question 3 continued.



Question 3 continued

Leave
blank

Only use this grid if you need to redraw your box plot.

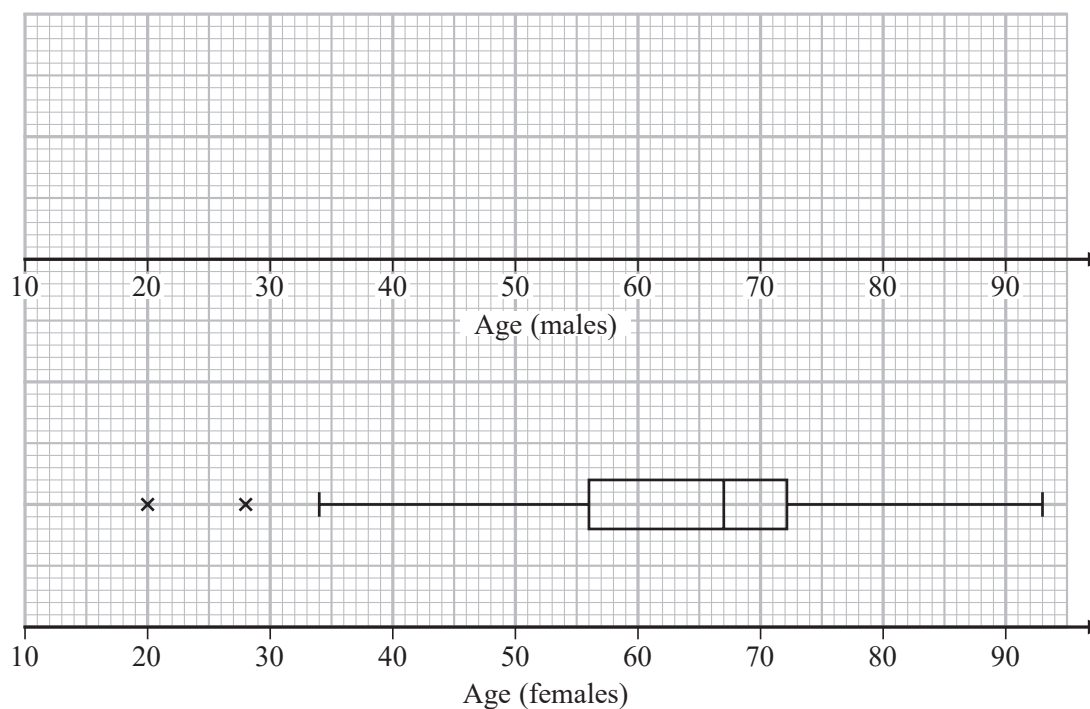


Figure 1

(Total 14 marks)

Q3



5. The following grouped frequency distribution summarises the speeds in km/h that a random sample of 225 cars were doing on a road.

Speed (x km/h)	Number of cars
$20 \leq x < 25$	54
$25 \leq x < 30$	90
$30 \leq x < 40$	60
$40 \leq x < 55$	15
$55 \leq x < 70$	6

A histogram is drawn to represent these data.

The height of the tallest bar is 10 cm.

(a) Calculate the height of the second tallest bar. (3)

(b) Estimate how many of the 225 cars were doing a speed between 35 km/h and 57 km/h.
You must show your working clearly. (2)

(c) Use linear interpolation to estimate the median speed in km/h for the 225 cars. (2)

The lower quartile for these data is 25.14 km/h to 4 significant figures.

(d) Find an estimate for (i) the upper quartile, (ii) the interquartile range. (4)

(e) Describe, giving a reason, the skewness of these data. (2)

Ali suggests that a normal distribution is a suitable model for the speed that cars were doing on this road.

(f) With reference to your answer in part (e), comment on Ali's suggestion, giving a reason for your answer. (1)

Question 5 continued

Lined area for writing the answer to Question 5.



Question 5 continued

Lined area for writing the answer to Question 5.



Question 5 continued

(Total for Question 5 is 14 marks)



MARK SCHEMES

official mark schemes, in the same order as the questions



Question	Scheme	
2(a)	$[\sum x_B =] 54 \times 45 - 24 \times 47 [=1302]$ or $\frac{24 \times 47 + 30b}{54} = 45$	M1
	$[\bar{x} =] \frac{54 \times 45 - 24 \times 47}{30} = 43.4 *$	A1*
		(2)
(b)	$\sqrt{\frac{66876 + 73826}{54} - 45^2} =$ or $\sqrt{\frac{(66876 + 73826) - \frac{2430^2}{54}}{54}}$	M1
	24.095.... awrt 24.1	A1
		(2)
(c)(i)	No effect as e.g. adding 2 to each does not change the variance	B1
(ii)	Increase as e.g. the total mark will increase	B1
(iii)	Decrease as e.g. the mean mark of class B is closer to the mean mark of class A	B1
		(3)
	Notes	Total 7
(a)	M1 for a correct method to find the total marks for class B or for a correct equation for b eg $1128 + 30b = 2430$ but not $30b = 1302$ on its own	
	A1* a fully correct method shown leading to 43.4 For the equation in b there should be 1 line of intermediate working between the equation and the given answer.	
SC:	Verification method e.g. $\frac{24 \times 47 + 30 \times 43.4}{54} = 45$ scores M1A0	
(b)	M1 for a correct method to find the standard deviation including root	
	A1 awrt 24.1 correct answer scores 2 out of 2 allow $s =$ awrt 24.3	
(c)(i)	B1 for no effect and a correct reason: - Spread/variation/range or difference (in marks) remains the same o.e. - Addition has no effect/only multiplication/division affects the variance o.e. Variance not affected by coding on its own is B0	
(ii)	B1 for increase and a correct reason: - Total/mark/score/$\sum x$ increases o.e. - Mean of class B increases (by 2 marks) o.e. - Total/mark/score/$\sum x$ for class B increases (by 60 marks) o.e. Incorrect statements score B0 e.g. "The mean (of all 54 students) increases by 2" or e.g. "Total for class B increases by 48 marks" Ignore calculation if seen, but do not allow increases by 1.11.... on its own to score the mark here.	
(iii)	B1 for decrease and correct reason: - Mean/marks of class B now closer to mean/marks of class A o.e. - Values/marks/scores are closer together o.e. - Less spread/variation between values/marks/scores (for class A and class B) o.e.	

Question Number	Scheme	Marks
3. (a)	$\left[\bar{x} = \frac{-1.2}{8} [= -0.15]\right]$ $"-0.15" = \frac{\bar{b} - 21}{2} \text{ oe}$ $\left[\bar{b} = \frac{165.6}{8}\right]$ $= \underline{\underline{20.7}} \text{ (cm)}$	M1 M1 A1 (3)
(b)	$\sigma_x = \sqrt{\frac{5.1}{8} - \left(\frac{-1.2}{8}\right)^2} [= \sqrt{0.615} = 0.784...]$ $\sigma_b = 2 \times '0.784...'$ $= \text{awrt } \underline{\underline{1.57}} \text{ (cm)}$	M1 M1 A1 (3)
(c)(i)	$x_9 = 1.2 \rightarrow b_9 = 1.2 \times 2 + 21 \text{ oe or } 9 \times 21 - 8 \times '20.7' [= 354.6]$ $= \underline{\underline{23.4}} \text{ (cm)}$	M1 A1 (2)
(ii)	$\sum x^2 = 5.1 + 1.2^2 [= 6.54] \quad \left[\Rightarrow \sigma_x = \sqrt{\frac{5.1 + 1.2^2}{9} - 0^2} \right]$ $= \text{awrt } \underline{\underline{0.852}} \text{ (cm)}$	M1 A1 (2)
Notes		Total 10
(a)	1 st M1 for correct expression for $\bar{x}$ ignore letter 2 nd M1 Using equation. " $\bar{x} = \frac{\bar{b} - 21}{2}$ " where $-1.2 < 'x' < 1.2$ Condone b rather than $\bar{b}$ A1 20.7 oe	1 st M1 for correct expression for $\sum b$ 2 nd M1 use of " $\sum b \div n$ " where ' $\sum b > 18$ '
(b)	1 st M1 correct method for σ_x or σ_x^2 or $5.1 = \frac{\sum b^2 - 42 \times "168.6" + 8 \times 441}{4}$ or $\sum b^2 = 3447.6$ 2 nd M1 for use of $2 \times$ their σ_x (or $4 \times$ their σ_x^2) (adding 21 is M0) or $\frac{"3447.6"}{8} - \left(\frac{"165.6"}{8}\right)^2 \text{ or } \sqrt{\frac{"3447.6"}{8} - \left(\frac{"165.6"}{8}\right)^2}$ A1 awrt 1.57 Allow $\frac{\sqrt{246}}{10}$ (allow $s_b = \text{awrt } 1.68$ or $\frac{4\sqrt{246}}{35}$ from an $n - 1$ method)	
(c)(i)	M1 for a correct equation using $x_9 = 1.2$ to enable b_9 to be found eg $1.2 = \frac{b - 21}{2}$ or a correct method to find $\sum x$ for the 9 squirrels. ft their 20.7 A1 23.4 oe	
(ii)	M1 for $5.1 + "(\pm 1.2)^2 [= 6.54]$ seen ft their x_9 Condone $5.1 + "(\pm 9.6)^2 [= 97.26]$ A1 awrt 0.852 Allow $\frac{\sqrt{654}}{30}$ (allow $s_x = \text{awrt } 0.904$ from an $n - 1$ method)	



Question Number	Scheme	Marks
3 (a)	Width = 2.5 (cm) 1.5 cm ² for freq of 5 so $6 \times 1.5 = 9 \text{ cm}^2$ for freq of 30 or $fd = \frac{5}{3} w \times h = 9$ So $h = 9 \div 2.5$ or $6 \div \frac{5}{3} = \mathbf{3.6 \text{ (cm)}}$	B1 M1 A1 (3)
(b)	$Q_2 = [12] + \frac{16}{25} \times 3$ allow use of $(n+1)$ giving $[12] + \frac{16.5}{25} \times 3$ $= 13.92 = \text{awrt } \mathbf{13.9}$	M1 A1 (2)
(c)(i)	$\sum fx = 5 \times 6.5 + 13 \times 9 + 16 \times 11 + 25 \times 13.5 + 30 \times 17.5 + 11 \times 24 = 1452$ $\bar{x} = 14.52 = \text{awrt } \mathbf{14.5}$	M1 A1 (2)
(ii)	$\sum fx^2 = 6.5^2 \times 5 + 9^2 \times 13 + 11^2 \times 16 + 13.5^2 \times 25 + 17.5^2 \times 30 + 24^2 \times 11 = 23\,280$ $\sigma_x = \sqrt{\frac{23\,280}{100} - ("14.52")^2}$ or $\sqrt{21.9696}$ $\sigma_x = 4.687... = \text{awrt } \mathbf{4.69}$	M1 M1 A1 (3)
(d)	$\frac{1}{2} \times 13 + 16 + 25 + 30 + \frac{1}{4} \times 11$ So proportion is 80.25 % or 0.8025 awrt 0.803	M1 A1 (2)
(e)	Profit = $2.2 \times "0.8025" + 0.8 \times \frac{0.75 \times 11}{100} - 1.2 \times \left(1 - \left[0.8025 + \frac{0.75 \times 11}{100}\right]\right)$ $= 1.6935$ awrt 1.7 (p)	M1 A1 (2)
	Notes	[14 marks]
(a)	B1 for width = 2.5 (cm) M1 for sight of 9 cm ² or $w \times h = 9$ or $fd = \frac{5}{3}$ (o.e.) A1 for height = 3.6 (cm)	
(b)	M1 for $\frac{16}{25} \times 3$ or $\frac{9}{25} \times 3$ or $\frac{m-12}{15-m} = \frac{16}{9}$ For any correct equation leading to Q_2 or correct fraction as part of Q_2	
(c)(i)	A1 for awrt 13.9 (use of $(n+1)$ giving 13.98 = awrt 14.0) M1 for attempt at $\sum fx$ with at least 3 correct terms or $900 < \sum fx < 1800$ for info $\sum fx = 32.5 + 117 + 176 + 337.5 + 525 + 264$ A1 for awrt 14.5 (correct answer only 2/2)	
(ii)	1 st M1 for attempt at $\sum fx^2$ with at least 3 correct terms or $20\,000 < \sum fx^2 < 26\,000$ for info $\sum fx^2 = 211.25 + 1053 + 1936 + 4556.25 + 9187.5 + 6336$ 2 nd M1 for a correct expression including $\sqrt{\quad}$ (ft their $\sum fx^2$ if clear it is $\sum fx^2$) Do not allow $(\sum fx)^2$ for $\sum fx^2$ A1 for awrt 4.69 (allow $s = 4.7107... \text{ awrt } 4.71$) (correct answer only 3/3)	
(d)	M1 for attempt at a correct expression (allow 1 error or omission) eg $100 - \left(5 + \frac{13}{2}\right) - \frac{33}{4}$	
(e)	A1 for awrt 80.3% or 0.803 M1 for a correct expression ft their 0.8025 o.e. eg $[2.2 \times (100 - 11.5 - 8.25) + 0.8 \times 8.25 - 1.2 \times 11.5] \div 100$ Condone $[2.2 \times "80" + 0.8 \times (8) - 1.2 \times (12)] \div 100$ A1 for awrt 1.7 Allow £0.017 (this must have units)	

Qu. No.	Scheme		Marks
3(a)	Width = 2.5 (cm)		B1
	Height = $\left(\frac{35}{15} \times 6\right) \div "2.5"$ or $\frac{4}{5} \times 7$ or $6\text{cm}^2 = 15$ (logs) or $14\text{cm}^2 = 35$ (logs) oe		M1
	= 5.6 (cm)		A1
			(3)
(b)	Median = $[20] + \frac{25}{35} \times 5$ allow $[20] + \frac{25.5}{35} \times 5$		M1
	= 23.57... allow 23.64.... awrt 23.6		A1
			(2)
(c)	19 + 35 + 8 (= 62*) or 88 – 3 – 15 – 8 (= 62*)		B1cso*
			(1)
(d)	$P(4 \text{ logs fit}) = \left(\frac{62}{88}\right) \times \left(\frac{61}{87}\right) \times \left(\frac{60}{86}\right) \times \left(\frac{59}{85}\right)$		M1
	= 0.23922.... awrt 0.239		A1
			(2)
(e)(i)	mean of $y = \frac{924}{88} [= 10.5]$		M1
	mean of $w = ("10.5" + 255) \times 2$		M1
	= 531		A1
			(3)
(ii)	variance of $y = \frac{12862}{88} - ("10.5")^2 [= 35.909...]$		M1
	variance of $w = "35.909" \times 4$ or $"35.909" \div 0.5^2$		M1
	= 143.636... awrt 144		A1
			(3)
Notes			Total 14
(a)	B1	2.5 oe	
	M1	Correct method to relate area to number of logs (may be implied by "their w" $\times$ "their h" = 14)	
	A1	5.6 oe	
(b)	M1	For a correct fraction multiplied by 5. Allow working downwards e.g. $[25] - \frac{10}{35} \times 5$	
	A1	Correct answer from correct working. Allow exact fraction.	
(c)	B1	A correct calculation seen. Allow $\frac{1}{2}(16)$ for 8. Allow equivalent methods $\frac{x-54}{70-54} = \frac{26-25}{27-25}$ Minimum working required $54 + 8$ or $70 - 8$ or $\frac{54+70}{2}$ NB: $26 + x = 88 \rightarrow x = 62$ is B0.	
(d)	M1	For $\left(\frac{n}{88}\right) \times \left(\frac{n-1}{87}\right) \times \left(\frac{n-2}{86}\right) \times \left(\frac{n-3}{85}\right)$ (allow any $n < 88$)	
	A1	awrt 0.239	
	SC	With replacement awrt 0.246 scores M1A0	
(e)(i)	M1	Correct method for finding mean of y (implied by 10.5) or for equation $0.5 \sum w - 88 \times 255 = 924$	
	M1	Correct method for finding mean of w or $\sum w = 46728$ and $\frac{46728}{88}$	
	A1	Cao	
(ii)	M1	Correct method for finding variance of y (implied by awrt 35.9) or $0.25 \sum w^2 + 255^2 \times 88 - 255 \sum w = 12862$	
	M1	Correct method for finding variance of w or $\sum w^2 = 24825208$ and $\frac{24825208}{88} - "531"{}^2$	

Question	Scheme		Marks
3(a)	Width = 1.25 [cm]		B1
	18.75 cm ² for freq of 20 so $\frac{18.75}{20} \times 16 = 15$ cm ² for a frequency of 16 or $w \times h = 15$ or $fd = 5$		M1
	$[h = 15 \div 1.25 \text{ or } h = 8 \div 5 \times 7.5 =] 12$ (cm)		A1
			(3)
(b)	$Q_2 = [32 + \frac{7}{20} \times 4 \text{ or using } n + 1 \text{ gives } Q_2 = [32 + \frac{7.5}{20} \times 4$		M1
	$= 33.4 \text{ (} n + 1 \text{ gives } 33.5)$		A1
			(2)
(c)	$\bar{y} = \frac{104}{50} [= 2.08]$	$\sum (w - 20) = 10 \times 104 [= 1040] \text{ or } \sum w = 10 \times 104 + 50 \times 20 [= 2040]$	M1
	$\bar{w} = 10 \times "2.08" + 20 = 40.8^*$	$\frac{"1040"}{50} + 20 = 40.8 \text{ or } \frac{"2040"}{50} = 40.8$	A1*
			(2)
(d)	$[\text{Variance of } y = \frac{233.54}{50} - ("2.08")^2 = \frac{861}{2500} = 0.3444] \text{ or } 10 \times \text{sd of } y = \text{sd of } w$		M1
	or $100 \times 233.54 = \sum (w^2) - 40 \times "2040" + 50 \times 400 [\Rightarrow \sum (w^2) = 84954] \text{ oe}$		
	$[\text{Variance of } w = "0.3444" \times 100 \text{ or } \frac{"84954"}{50} - 40.8^2 = \frac{861}{25} = 34.44]$		M1
	or $\text{sd of } y = \sqrt{"0.3444"} [= \frac{\sqrt{861}}{50} = 0.5868...]$		
	$\text{sd of } w = \sqrt{"0.3444" \times 100} \text{ or } \sqrt{"34.44"} \text{ or } 10 \times \frac{\sqrt{861}}{50}$		M1
	$= 5.868...$ awrt 5.87		A1
(e)(i) (ii)	The mean would not change (as 40.8 is the mean)		B1
	The standard deviation would decrease (as 40.8 is in the middle so data closer together)		B1
	Both correct with a correct reason for why the standard deviation decreases		ddB1
			(3)
	Notes		Total 14
(a)	B1	for width = 1.25 no need for units	
	M1	for sight of 15 or "their w " $\times$ "their h " = 15 or $fd = 5$ May be implied by $h = 12$	
	A1	for height = 12 no need for units	
(b)	M1	for $\frac{7}{20} \times 4$ or $\frac{13}{20} \times 4$ or $\frac{m-32}{25-18} = \frac{4}{20}$ oe or $\frac{36-m}{38-25} = \frac{4}{20}$ oe (allow 25.5 rather than 25)	
	A1	33.4 or if using $(n+1)$ 33.5	
(c)	M1	for a correct method to find the mean of y or $\sum (w - 20)$ or $\sum w$ ($10 \times 104 + k$ where $k \neq 20 \times 50$ is M0)	
	A1*	for a correct method to find mean of w which leads to 40.8	
(d)	M1	for a correct method to find the Variance of y or writing/using $10 \times \text{sd of } y = \text{sd of } w$ or a correct equation to find $\sum w^2$	
	M1	for a correct method to find the Variance of w or sd of y ft their $\text{Var}(y)$	
	M1	for a correct method to find the sd of w ft their $\text{Var}(w)$	
	A1	awrt 5.87 NB an exact answer $\frac{\sqrt{861}}{5}$ scores A0	
(e)(i)	B1	for no change (No reason needed) Allow mean = 40.8 to imply no change	
(ii)	B1	for sd decreases/be smaller/go down (condone Var decreases) (No reason needed)	
	ddB1	Both previous B1 awarded. For a correct reason for the sd decreasing Allow $(x - \bar{x}) = 0$ Allow $\sum (x - \bar{x})^2$ doesn't change and n increases. Allow the values would be more concentrated about the mean	

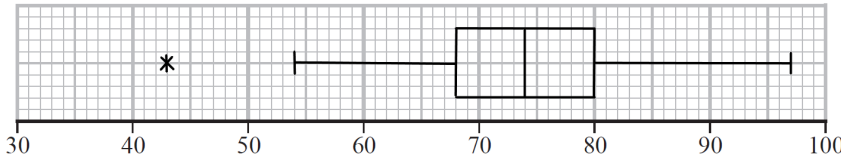
Qu	Scheme	Marks
1(a)	eg 60 people = 1.5 large squares/6 medium squares/150 small squares or eg 1 person = 0.025 large squares/0.6 medium squares/2.5 small squares or eg [1 small square =] 0.4 people/[1 medium square =] 10 people/[1 large square =] 40 people eg a correct f.d. eg 60/(20 – 10) [= 6] eg a correct frequency, 100, 70, 20, 24 associated with the appropriate bar	B1
	eg $\frac{8}{10} \times 20$ or $\frac{15}{30} \times 24$ or 8×2 or 0.8×15 or $\frac{40}{2.5}$ or $\frac{30}{2.5}$ or $8 \times 5 \times 0.4$ $2 \times 15 \times 0.4$ or 16 or 12 or 70×0.4	M1
	28 people	A1
		(3)
(b)	Median = $[5] + \frac{5}{70} \times 37$ or $[10] - \frac{5}{70} \times 33$	M1
	= 7.642...	awrt 7.64
		A1
		(2)
(c)	$\sum \text{midpoint} \times \text{freq} = 2.5 \times 100 + 7.5 \times 70 + 15 \times 60 + 25 \times 20 + 45 \times 24 [= 3255]$	M1
	Mean = $\frac{3255}{274}$	dM1
	= 11.879...	awrt 11.9
		A1
		(3)
	Notes	Total 8
(a)	B1: for finding a ratio between people and area. Allow just the numbers for 1 person or for 1 square ie 0.025, 0.6, 2.5, 0.4, 1.66 or 40 or calculating f.d. for any bar correctly, fd = 20, 14, 6, 2 or 0.8 Information may be seen on diagram. Must be clear it is a fd either by correct use, seen on axes associated with correct bar or stated as an fd. May be implied by M1	
	M1: for a correct method to find the number of people between 22 and 30 km or 30 and 45 km or between 22 and 45 km	
	A1: 28	
	NB An answer of 28 gains 3/3 unless from obvious incorrect working	
(b)	M1: Allow equivalent for $n + \frac{5}{70} \times 37$ or $\frac{Q_2 - n}{5} = \frac{137 - 100}{170 - 100}$ or $n - \frac{33}{70} \times 5$ or $\frac{n - Q_2}{5} = \frac{170 - 137}{170 - 100}$ oe Allow alternative methods eg $\frac{Q_2 - 5}{10 - Q_2} = \frac{137 - 170}{170 - 137}$ Allow 37.5 for 37, 137.5 for 137, 32.5 for 33	
	A1: awrt 7.64 or $\frac{107}{14}$ or allow awrt 7.68 or $\frac{215}{28}$ if using $n + 1$ Allow awrt 7640 m or 7680 m but must have units.	
(c)	M1: Attempt at correct expression for $\sum \text{midpoint} \times \text{freq}$ - at least 3 products with correct midpoints added with at least 1 of these products fully correct . Allow for 3255	
	M1: dep on M1 being awarded for dividing “their sum” by 274	
	A1: awrt 11.9 or $\frac{3255}{274}$ Allow awrt 11900 m but must have units	

Question Number	Scheme		Marks
1 (a)	$2 \times 36 = 72 \quad 8 \times 4 = 32$		M1 A1
			(2)
(b)	$[13] + \frac{(204 - 184)}{120} \times 2$	$[13] + \frac{(204.5 - 184)}{120} \times 2$	M1
	$= \frac{40}{3} = \text{awrt } 13.3$		A1
			(2)
(c)	Symmetrically distributed/No skew as the mean $\approx$ median		B1
			(1)
(d)	$\frac{32}{4} + 152 + \frac{120}{2} [= 220]$		M1
	$\frac{'220'}{408} \times \frac{'219'}{407}$		M1
	$\frac{365}{1258}$ or 0.2901... awrt 0.29		A1
			(3)
	Notes		Total 8
(a)	M1	For any equivalent method to find either frequency Maybe implied by either correct frequency Also maybe implied by two frequencies which add to 104 Also maybe implied by a correct scale on the fd axis , at least 3 labels	
	A1	For 72 and 32	
(b)	M1	For any equivalent method to find the median e.g. $\frac{Q_2 - 13}{15 - 13} = \frac{204 - 184}{304 - 184}$ or $\frac{15 - Q_2}{Q_2 - 13} = \frac{304 - 204}{204 - 184}$ allow working downwards $[15] - \frac{(304 - 204)}{120} \times 2$	
	A1	awrt 13.3	
(c)	B1	For a correct identification of skew [which must either be symmetric/no skew or (slight) negative skew] with a correct supporting reason. Condone mean < median so negative skew Allow use of 'their median' in the comparison provided 'their median' ... 13.2 Allow $Q_1 = \text{awrt } 10.8$ or awrt 10.9 and $Q_3 = \text{awrt } 15.1$ and $Q_2 - Q_1 > Q_3 - Q_2$ so negative skew. Comments referring only to the diagram (being symmetrical therefore no skew) send to review	
(d)	M1	For a correct method to find the number of plants between 8cm and 14cm (may be implied by sight of 220)	
	M1	For $\frac{n}{408} \times \frac{n-1}{407}$ or $\left(\frac{n}{408}\right)^2$ with 210 ,, n ,, 230	
	A1	awrt 0.29 may see $\frac{3025}{10404}$ from $\left(\frac{220}{408}\right)^2$	



Question Number	Scheme		Marks
5 (a)	[Time is] continuous		B1
			(1)
(b)	16 photographers = 160 small squares or 64 photographers = 640 small squares or $\frac{16}{160} = 0.1$ or $\frac{64}{640} = 0.1$ or $\frac{160}{16} = 10$ or $\frac{640}{64} = 10$ Frequency density = 1.6 or Correct scale on the frequency density axis		M1
	$\frac{x}{240} = \frac{16}{160}$ oe or $\frac{(20-12)}{10} \times 16 + \frac{(24-20)}{5} \times 14$		M1
	= 24		A1
			(3)
(c)	Using n $[Q_2 =] 20 + \frac{(32-21)}{35-21} \times (25-20)$ oe or $25 - \frac{35-32}{35-21} \times (25-20)$ oe = awrt 23.9	Using $n + 1$ $[Q_2 =] 20 + \frac{(32.5-21)}{35-21} \times (25-20)$ oe or $25 - \frac{35-32.5}{35-21} \times (25-20)$ oe = awrt 24.1	M1
			A1
			(2)
(d)	Mean = Median or Mean $\approx$ Median e.g. Appropriate decision. Consistent with expectation for a normal distribution.		M1
			A1ft
			(2)
	Notes		Total 8
(a)	B1	Allow not discrete	
(b)	M1	For establishing a ratio between photographers and area or calculating frequency density (may be implied by 2 nd M1)	
	M1	For a correct ratio or expression using areas for photographers from 12 to 24	
	A1	Cao	
(c)	M1	For a correct method to find median using either n or $n + 1$	
	A1	awrt 23.9 or awrt 24.1 if using $n + 1$	
(d)	M1	For a correct comment about mean and median ft their median Allow mean is close to median to imply mean $\approx$ median Ignore any comments made about the shape of the histogram	
	A1ft	For a correct compatible comment about Charlie's decision ft their median If Mean = Median or Mean $\approx$ Median, then the decision should be that a normal distribution is suitable [due to symmetry] If Mean < Median or Mean > Median or mean $\neq$ median, then the decision should be that a normal distribution is not suitable [due to the skew in the data]	

Question Number	Scheme		Marks
7 (a)	$P(C < 570) = 0.5 + \frac{570 - 550}{650 - 550} \times 0.25$ or $\frac{x - 100}{150 - 100} = \frac{570 - 550}{650 - 550}$		M1
	$= 0.55$		A1
			(2)
(b)	$Q_3 + 1.5 \times (Q_3 - Q_1) = 650 + 1.5 \times 200$		
	$= 950$		B1
			(1)
(c)	Normal distribution is supported as box plot is reasonably symmetrical oe		B1
			(1)
(d)	$z = 2$		
	$P(Z > 2) = [1 - 0.9772]$		M1
	$= 0.0228$		A1
			(2)
(e)	$1000 = 560 + 2\sigma$ or $1000 > 560 + 2\sigma$		M1
	$\sigma = 220$		A1
			(2)
Notes			Total 8
(a)	M1	Correct method $0.5 + p \times 0.25$ where $0 < p < 1$ or attempt to find the number of cabbages weighing less than 570 (implied by $x = 110$)	
	A1	0.55 condone awrt 0.55 for 2 out of 2 marks	
(b)	B1	cao	
(c)	B1	Supports Normal/Yes (supports assumption)	
		<u>and</u> reference to symmetry or no skew eg $Q_3 - Q_2 = Q_2 - Q_1$ oe in words Do not allow mean = median on its own for symmetric	
(d)	M1	Use of $P(Z > 2)$ can be implied by sight of awrt 0.977 or sight of awrt 0.0228	
	A1	awrt 0.0228 (calculator gives 0.0227501...)	
(e)	M1	Attempt to use $560 + 2\sigma$ to set up appropriate equation or inequality (allow ...)	
		Implied by sight of 220	
	A1	allow $\sigma < 220$ or $\sigma \approx 220$ condone 219.999... but eg 219 as final answer is A0	

Question Number	Scheme	Marks
2. (a)	[Median =] <u>74</u>	B1
(b)	$Q_1 = 68 \quad Q_3 = 80$ $[IQR = 80 - 68 =] \underline{12}$	M1 A1 (1) (2)
(c)	$Q_1 - 1.5 \times (IQR) = "68" - 1.5 \times "12" [= 50]$ $\text{or } Q_3 + 1.5 \times (IQR) = "80" + 1.5 \times "12" [=98]$ Outliers are < 50 or > 98 So there is just one outlier at <u>43</u>	M1 A1ft A1 (3)
(d)		M1 A1ft A1 (3) [9 marks]
Notes		
(a)	B1 for 74	
(b)	M1 for an attempt at both and at least one correct. May be in a calculation e.g. $80 - A$ (where $60 < A < 80$) <u>or</u> $B - 68$ (where $68 < B < 90$) A1 for 12	
(c)	M1 for correct attempt for at least one of the limits. Can ft their quartiles and IQR 1 st A1ft for correct attempts for both limits and with at least one correct limit or correct ft using their quartiles and IQR Sight of the two limits 50 and 98 will score M1A1 2 nd A1 for identifying only one outlier at 43 (e.g. may say " $43 < 50$ ") Must be stated in (c) Just stating the outlier is 43 (or seeing it on box plot) without sight of limits is M0A0A0	
(d)	M1 for drawing a box with only two whiskers one at each end 1 st A1ft for Q_1, Q_2 and Q_3 as a correctly drawn box (<u>or</u> ft their values for $Q_1 < Q_2 < Q_3$) 2 nd A1 for upper whisker ending at 97 and lower whisker ending at 54 or 50 and only one outlier, shown at 43 Allow ± 0.5 of a square for accuracy NB A fully correct box plot can score full marks in (d) even if other parts are missing or incorrect	

Qu	Scheme		Marks
3(a)	$[\bar{x} =] \frac{3711}{81} [= 45.814...]$	$[\sum l =] 3711 + 81 \times 600 [= 52311]$	M1
	$[\bar{l} =] "45.814..." + 600$	$[\bar{l} =] \frac{"52311"}{81}$	M1
	$[\bar{l} =] 645.81...$	awrt 646	A1
			(3)
(b)	$[\sigma_x^2 =] \frac{475181}{81} - \left(\frac{3711}{81}\right)^2 [= 3767]$	$[\text{Var}(L) =] \frac{34088381}{81} - \left(\frac{"52311"}{81}\right)^2$	M1
	$= 3767.43... \Rightarrow \sigma_l^2 = 3767.43...$	$= 3767.43... \quad \text{awrt 3770}$	A1
			(2)
(c)	40		B1cao
			(1)
(d)	IQR = 5400 – 3800 [= 1600]		M1
	5400 + 1.5 × "1600" [= 7800] or 3800 – 1.5 × "1600" [= 1400]		M1
	7800 > 7700 and 1400 < 1600 therefore there are no outliers		A1
			(3)
	Notes		Total 9
(a)	M1 for a correct method to find $\bar{x}$ or $\sum l$ Allow 45.8 or better. Ignore labels		
	M1 for a correct method to find $\bar{l}$ ft their $\bar{x}$ if it is clearly labelled or it comes from $\frac{3711}{81}$ or ft their $\sum l$ if it is clearly labelled or comes from $3711 + 81 \times 600$		
	A1 awrt 646 or $\frac{17437}{27}$ or $\frac{52311}{81}$ oe		
(b)	M1 correct method to find Var(X) implied by awrt 3770 or a correct method to find Var(L) ft their $\sum l$ or Allow calculation of $\text{sd}[\sigma_x] = \text{awrt 61.4}$ Ignore labels		
	A1 awrt 3770 labelled clearly as Var(L) or Var(L) = Var(X) or $\sigma_l = \sigma_x$ stated or variance is not changed by coding is stated or they have gained the answer from $\frac{34088381}{81} - \left(\frac{"52311"}{81}\right)^2$		
(c)	B1 cao		
(d)	M1 correct method to find IQR. May be implied by a correct limit. NB $1.5 \times (5400 - 3800) = 2400$		
	M1 for a correct method to find the upper or the lower outlier boundary.		
	A1 both 7800 and 1400 correct and 7700 and 1600 (as the minimum not IQR) seen and explicitly stating no outliers		

3. (a)	lower quartile = 116 upper quartile = 125 "125" + 1.5 × ("125" – "116") or "125" + 1.5 × (9) Outlier is greater than 138.5, so $c = 9^*$	B1 M1 A1*cs0 (3)
(b)	$\bar{x} = \frac{-96}{24} [= -4]$ $\bar{d} = '\bar{x}' + 125$	$\sum d = 125 \times 24 - 96 [= 2904]$ $\bar{d} = \frac{2904}{24}$ $\bar{d} = 121$ M1 M1 A1 (3)
(c)	$[\sigma_x = \sigma_d] = \sqrt{\frac{1306}{24}}$ $[\sigma_d] = 7.3767...$ awrt 7.38	M1 A1 (2)
(d)	$[P(D > 118 X < 0)] = \frac{P(118 < D < 125)}{P(D < 125)}$ or $\frac{P(-7 < X < 0)}{P(X < 0)}$ or $\frac{5/24}{14/24}$ $= \frac{5}{14}$	M1 A1 (2) [10]
Notes		
(a)	B1 both values correct. Both values must be seen either in the calculation or separately. They are not implied by the IQR = 9 M1 use of $Q_3 + 1.5 \times \text{IQR}$ with their values. May be implied by 138.5 if B1 awarded A1*cs0 for 138.5 and conclusion $c = 9$ (do not accept $c = 139$) with no errors. Answer is given so working must be shown.	
(b)	1 st M1 for correct expression for $\bar{x}$ 2 nd M1 use of $\bar{d} = '\bar{x}' + 125$ A1 121 NB condone no labelling or incorrect labelling throughout part(b)	1 st M1 for correct expression for $\sum d$ 2 nd M1 use of " $\sum d$ " ÷ 24 must be clear it is their sum
(c)	M1 correct expression $\sqrt{\frac{1306}{24}}$ A1 awrt 7.38 final answer	
(d)	M1 correct probability statement (allow a probability of $\frac{k}{14}$ where $0 < k < 14$ to score M1) A1 allow awrt 0.357	

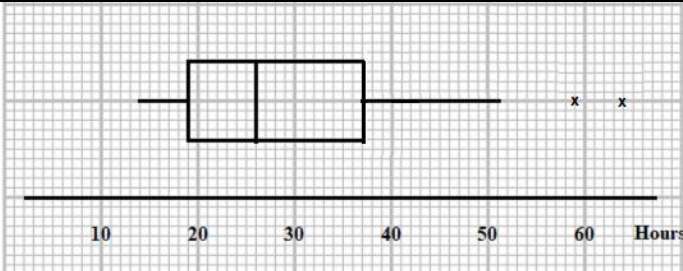
Question	Scheme						Marks	
1 (a)	Time taken (<i>t</i> minutes)		5 – 10	10 – 14	14 – 18	18 – 25	25 – 40	B1
	Frequency (<i>f</i>)		10	16	24	35	15	
(b)	$10 + 16 + (2 \times 6)$ or $10 + 16 + \frac{24}{2}$ or $\frac{x-26}{50-26} = \frac{16-14}{18-14}$							M1
	$= 38$							A1
								(2)
(c)	$[\sum ft =]7.5 \times 10 + 12 \times 16 + 16 \times 24 + 21.5 \times '35' + 32.5 \times '15' [= 1891]$							M1
	$\text{Mean} = \frac{1891}{100} = 18.91$							A1
								(2)
(d)	$\text{Standard deviation} = \sqrt{\frac{41033}{100} - \left(\frac{1891}{100}\right)^2}$ or $\sqrt{\frac{41033 - 100 \times '18.91'^2}{99}}$							M1
	$= 7.262...$ or $7.298...$ awrt 7.26 or awrt 7.3[0]							A1
								(2)
(e)	$[\text{LQ} =] 10 + \frac{15}{16}(14 - 10)[= 13.75]$			$[\text{LQ} =] 10 + \frac{15.25}{16}(14 - 10)[= 13.8125]$			M1	
	or $14 - \frac{1}{16}(14 - 10)[= 13.75]$			or $14 - \frac{0.75}{16}(14 - 10)[= 13.8125]$				
	or $\frac{Q_1 - 10}{14 - 10} = \frac{25 - 10}{26 - 10}[= 13.75]$			or $\frac{Q_1 - 10}{14 - 10} = \frac{25.25 - 10}{26 - 10}[= 13.8125]$				
	or $\frac{Q_1 - 14}{14 - 10} = \frac{25 - 26}{26 - 10}[= 13.75]$			or $\frac{Q_1 - 14}{14 - 10} = \frac{25.25 - 26}{26 - 10}[= 13.8125]$				
	$\text{IQR} = 23 - '13.75'$			$\text{IQR} = 23 - '13.8125'$			M1	
	$= 9.25$			$= \text{awrt } 9.19$			A1	
								(3)
Notes								Total 10
(a)	B1	for 35 and 15 (If answers given are in both the table and answer lines then mark the answers given in the table)						
(b)	M1	for $10 + 16 + (2 \times 6)$ or $10 + 16 + \frac{24}{2}$ or $\frac{x-26}{50-26} = \frac{16-14}{18-14}$						
	A1	Cao						
(c)	M1	A correct method for finding $\sum ft$ May be implied by 1891 Allow one error						
	A1	18.91 Allow 18.9						
(d)	M1	for a correct calculation of the standard deviation ft their mean						
	A1	awrt 7.26 or awrt 7.3 if using $n - 1$						
(e)	M1	for $10 + \frac{15}{16}(14 - 10)$ or $14 - \frac{1}{16}(14 - 10)$ or $\frac{Q_1 - 10}{14 - 10} = \frac{25 - 10}{26 - 10}$ or $\frac{Q_1 - 14}{14 - 10} = \frac{25 - 26}{26 - 10}$ or $10 + \frac{15.25}{16}(14 - 10)$ or $14 - \frac{0.75}{16}(14 - 10)$ or $\frac{Q_1 - 10}{14 - 10} = \frac{25.25 - 10}{26 - 10}$ or $\frac{Q_1 - 14}{14 - 10} = \frac{25.25 - 26}{26 - 10}$						
	M1	UQ – LQ ft their LQ provided LQ < UQ						
	A1	For 9.25 or awrt 9.19 if $n + 1$ is used						

Question Number	Scheme	Marks
1 (a)	[Mode =] 137	B1
		(1)
(b)	$a = 106$ $b = 129$ $c = 126$	B1 B1 B1
		(3)
(c) (i)	$\left[\text{mean} = \frac{4016}{31} = \right] 129.5$	B1
(ii)	[Standard deviation =] $\sqrt{\frac{525056}{31} - \left(\frac{4016}{31}\right)^2}$ or $\sqrt{\frac{31}{30} \left(\frac{525056}{31} - \left(\frac{4016}{31}\right)^2 \right)}$	M1
	= 12.4 or 12.6	A1
		(3)
(d)	$\frac{'129.5' - '137'}{'12.4'} [= -0.6]$	M1
	Negative [skew]	A1ft
		(2)
(e)	A correct difference of the average or a correct difference of the spread e.g. On average action films run for longer than comedy films as the median is greater 129 > 117	B1ft
		(1)
Notes		Total 10
(a)	B1	Cao Must be seen in part (a)
(b)	B1	$a = 106$ Must be attached to a (Condone Q_1 /lower quartile for a)
	B1	$b = 129$ Must be attached to b (Condone Q_2 /median for b)
	B1	$c = 126$ Must be attached to c (Condone Q_3 /upper quartile for c)
(c) (i)	B1	awrt 130
(ii)	M1	For a correct method (including the square root) to find the standard deviation. Ft their mean
	A1	May be implied by awrt 12.4 or awrt 12.9 or awrt 12.6 if sample standard deviation is calculated
	A1	awrt 12.4 or awrt 12.6 if sample standard deviation is calculated Correct answer only scores 2/2
(d)	M1	For substitution of the mode, the mean and the standard deviation into the expression. Ft their mode, their mean and their standard deviation. Allow awrt -0.6 if no working shown.
	A1ft	A correct interpretation ft their expression. Ignore any reference to correlation
(e)	B1ft	for a correct comment, referring to length , with reference to a correctly named statistic . Must include the correct figures compared. Ft their values. Ignore any reference to skew

The following table may be useful (but not an exhaustive list of possible answers)

Statistic	A	C	Comment
Median	129	117	On average action films run for longer as the median is greater 129 > 117
Mean	129.5	114.5	On average action films have a greater running time as the mean is greater 129.5 > 114.5
Mode	137	127	On average comedy films are shorter in length as the mode is less 127 < 137
IQR	17	20	Comedy films have a greater spread of running times as the IQR is greater 20 > 17
Range	57	39	Action films have a larger variation of running times as the range is greater 57 > 39
SD	12.4	11.9	Comedy films show less variability in the length of films as the SD is less 11.9 < 12.4



Question Number	Scheme		Marks
1(a)	$w = 8$		B1
	$x = 19$		B1
	$y = 37$		B1
			(3)
(b)	$"37" + 1 \times ("37" - "19") [= 55]$		M1
	59 and 64		A1ft
			(2)
(c)			M1 A1ft A1ft
			(3)
(d)	$\frac{("37" - 26) - (26 - "19")}{("37" - "19")}$		M1
	$= 0.22$ (to 2 sf)		A1
			(2)
(e)	E.g. 'The mean uses all the data'		B1
			(1)
			Total 11
Notes			
(a)	B1B1B1	Cao May be seen in table before part (a). $w = 28$ is first B0.	
(b)	M1	Calculation for the outliers using their lower quartile and upper quartile. Allow "their upper quartile" + "their IQR" for this mark i.e. $37 + 18$	
	A1ft	For identifying 59 and 64 as outliers from correct working. Ft the identification of outlier(s) (if any) from "their 55" from their shown calculation. Answer only is M0A0.	
(c)	M1	For a box with at least one whisker drawn	
	A1ft	14 for lowest whisker, 26 for median, "19" and "37" plotted for quartiles ft their values for quartiles	
	A1ft	Upper whisker at 51 or "their 51" plus "their outliers" plotted but there must be at least one outlier for this mark. Condone upper whisker at "their 55". NB award A0 if there is more than one whisker at either end	
(d)	M1	For substituting their values into the formula	
	A1	Allow awrt 0.22 (allow $\frac{2}{9}$ or 0.2)	
(e)	B1	A correct reason which supports Landacre's use of the mean or rejects their use of the median. Allow comment relating to (slight) positive skew so mean > median so Landacre will know the larger average they will have to pay. Comments about skewness/symmetry on their own score B0. Mean includes the outliers is B1. Condone Median is not affected by the outliers for B1. Mean is more accurate is B0.	

Question Number	Scheme	Marks
1. (a)	[Area = $k \times \text{frequency} \rightarrow 16.5 = k \times 12 \rightarrow$] Area = $\frac{16.5}{12} \times 18$ oe = <u>24.75</u> (cm ²)	M1 A1 (2)
(b)	fd method $\frac{24}{58-55} [= 8]$ and $\frac{35}{55-50} [= 7]$ or Area method $\frac{16.5}{12} \times 24 [= 33]$ and $\frac{16.5}{12} \times 35 [= 48.125]$ or $\frac{16.5}{12} \times \frac{24}{3} [= 11]$ and $\frac{16.5}{12} \times \frac{35}{5} [= 9.625]$ Let h = height of the 2 nd tallest bar [$h = \frac{10}{8} \times 7$] or [$h = \frac{48.125 \times 10 \times 3}{5 \times 33}$] or [$h = 9.625 \times \frac{10}{11}$] = <u>8.75</u> (cm)	M1 dM1 A1 (3)
(c)(i)	[$Q_2 =$] $50 + \frac{7}{35} \times 5$ or [$Q_2 =$] $55 - \frac{28}{35} \times 5$ = <u>51</u> (cm)	M1 A1
(ii)	[$Q_3 =$] $55 + \frac{2}{24} \times 3$ or [$Q_3 =$] $58 - \frac{22}{24} \times 3$ and "55.25" – 45 = <u>10.25</u> (cm)	M1 A1 (4)
(d)	$\frac{55.25 - 2("51") + 45}{55.25 - 45}$ [$= -0.17073... < 0$] negative [skew].	M1 A1ft (2)
Notes		[11]
(a)	M1 allow equivalent eg $16.5 \times \frac{3}{2}$ A1 for 24.75 allow 24.8	
(b)	M1 correct method for finding the frequency density or area for the highest and 2nd highest bars Allow if 8 and 7 seen or 33 and 48.125 seen or 9.625 rather than 48.125 and/or 11 seen rather than 33 dM1 dep on previous M mark awarded. A fully correct expression for h or a fully correct equation to enable h to be found eg $\frac{33}{10 \times 3} = \frac{48.125}{5h}$ A1 8.75 oe NB answer of 8.75 seen as final answer 3/3	
(c)(i)	M1 for $50 + \frac{7}{35} \times k$ or $55 - \frac{28}{35} \times k$ or $\frac{Q_2 - 50}{k} = \frac{60 - 53}{88 - 53}$ or $\frac{55 - Q_2}{k} = \frac{88 - 60}{88 - 53}$ where 4,, k ,, 6 oe (condone use of $n + 1$ ie 7.5 rather than 7, 27.5 rather than 28 or 60.5 rather than 60) A1 51 (condone for use of $n + 1$ awrt 51.1)	
(ii)	M1 $55 + \frac{2}{24} \times t$ or $58 - \frac{22}{24} \times t$ or $\frac{Q_3 - 55}{t} = \frac{90 - 88}{112 - 88}$ or $\frac{58 - Q_3}{t} = \frac{112 - 90}{112 - 88}$ oe where 2,, t ,, 4 and using "their Q_3 " – 45 (condone use of $n + 1$ ie 2.5 or 2.75 rather than 2, 21.5 or 21.25 rather than 22 or 90.5 or 90.75 rather than 90) A1 10.25 oe eg 41/4 allow 10.3 from correct working	



Question Number	Scheme					Marks										
1 (a)	Discrete uniform					B1										
						(1)										
(b)	$\frac{1}{2}$					B1										
						(1)										
(c)(i)	$[E(R) =] 2.5$					B1										
(ii)	$[E(B) =] 4$					B1										
						(2)										
(d)	$[E(B^2) =] \frac{1}{4}(1^2 + 3^2 + 5^2 + 7^2) = 21$					M1										
	$\text{Var}(B) = E(B^2) - (E(B))^2 = "21" - "4" = 17$					M1										
	$= 5$					A1										
						(3)										
(e)	Possible combinations $(R, B) : (1,1) (1,3) (2,1) (3,1) (4,1) (2,3)$ <u>or</u> $\frac{1}{4} \times \frac{1}{4} \times 6$					M1										
	$P(R + B = 5) = \frac{6}{16}$					A1										
						(2)										
(f)	<table><tr><td>R^2</td><td>1</td><td>4</td><td>9</td><td>16</td></tr><tr><td>B</td><td>1</td><td>3</td><td>5</td><td>7</td></tr></table>	R^2	1	4	9	16	B	1	3	5	7					
	R^2	1	4	9	16											
	B	1	3	5	7											
	Possible combinations $(R, B) : (1,3) (1,5) (1,7) (2,4) (2,5) (2,7)$ <u>or</u> $\frac{1}{4} \times \frac{1}{4} \times 5$					M1A1										
$P(R^2 < B) = \frac{5}{16}$					A1											
						(3)										
(g)	$B = 5 \text{ and } R = 1, B = 7 \text{ and } R = 3 \rightarrow D = 4$		$B = 7 \text{ and } R = 2 \rightarrow D = 5$			M1										
	$P(D = 4) = \frac{1}{4} \times \frac{1}{4} \times 2 = \frac{1}{8}$		$P(D = 5) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$			A1										
	$p = F(4) = \frac{3}{4} + \frac{1}{8} = \frac{7}{8}$		$p = F(5) - P(D = 5) = \frac{15}{16} - \frac{1}{16} = \frac{7}{8}$			A1										
						(3)										
Notes						Total 15										
(a)	B1	Must include both words (in either order). Ignore extraneous non-contradictory words.														
(b)	B1	0.5 oe														
(c) (i)	B1	cao need not be labelled unless done in wrong order (blue then red)														
(ii)	B1	cao need not be labelled unless done in wrong order (blue then red)														
(d)	M1	Correct method to find $E(B^2)$ at least 3 terms correct (implied by $E(B^2) = 21$). Ignore label. $E(B^2) = \frac{21}{4}$ is M0														
	M1	Correct method to find $\text{Var}(B)$ fit their $E(B^2)$ and their $E(B)$														
	A1	cao an answer of 5 without working send to review														
(e)	M1	At least 4 correct combinations identified with no incorrect ones given. Ignore duplicates, but do not accept eg (1, 4) as a duplicate of (4, 1). If not labelled, combinations must be consistently ordered. <u>or</u> if no combinations given, correct probability calculation (implied by correct answer with no obvious incorrect working)														
	A1	0.375 oe must come from correct combinations or correct working														

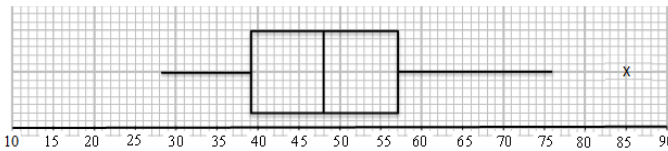
(f)	M1	At least 3 correct combinations identified with no incorrect ones given. Ignore duplicates, but do not accept eg (1, 4) as a duplicate of (4, 1). If not labelled, combinations must be consistently ordered. (allow 4, 5 and 4, 7 stated as combinations instead of 2, 5 and 2, 7)
	A1	All 5 correct combinations with no extras or duplicates. <u>or</u> if no combinations given, correct probability calculation (M1A1 implied by correct answer with no obvious incorrect working)
	A1	$\frac{5}{16}$ oe (accept awrt 0.313) must come from correct combinations or correct working
(g)	M1	Correct combinations identified for $D = 4$ or $D = 5$ (may be implied by correct working for 1 st A1).
	A1	$P(D = 4) = \frac{1}{8}$ or $P(D = 5) = \frac{1}{16}$ need not be labelled but working/combinations must imply the correct label do not award this mark for $F(4) = \frac{1}{8}$ on its own These may be seen as part of the probability distribution of D
	A1	0.875 oe Correct answer does not automatically imply 3 out of 3. Need to see correct combinations identified <u>or</u> see correct probability calculation/distribution. Answer only is 0 out of 3.

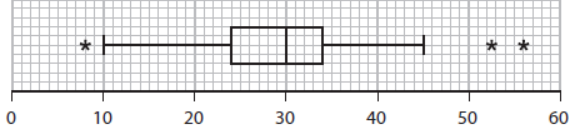
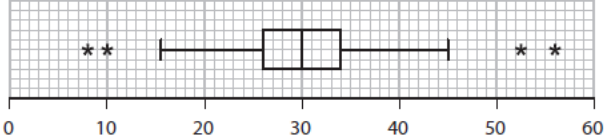
Question Number	Scheme		Marks
2 (a)	[Range = 0.87 – 0.21]	= 0.66	B1 (1)
(b)	Median (24 th value) =	= 0.48	B1 (1)
(c)	LQ (12 th value) = 0.35 0.31 = UQ – “0.35” [UQ = 0.66]	$a = \underline{6}$	B1 M1 A1 (3)
(d)	$sd = \sqrt{\frac{13.4228}{47} - \left(\frac{23.72}{47}\right)^2}$ or $S_{xx} = 13.4228 - \frac{23.72^2}{47} [= 1.4517...]$ and $sd = \sqrt{\frac{S_{xx}}{47}}$	$= 0.17575... = \underline{0.176^*}$	M1 A1* (2)
(e)(i)	$\frac{\sum y + 23.72}{65} = 0.502$ or $0.502 \times 65 = 32.63$ $32.63 = \sum y + 23.72$	8.91*	M1 A1* (2)
(ii)	$0.204 = \sqrt{\frac{[13.4228 + \sum y^2]}{65} - 0.502^2}$ or $0.204^2 = \frac{[13.4228 + \sum y^2]}{65} - 0.502^2$ $\sum y^2 = (0.204^2 + 0.502^2) \times 65 - 13.4228$ $= 5.6625$ awrt 5.66		M1 M1 A1 (3)
Notes			Total 12
(a)	B1	0.66 oe	
(b)	B1	0.48 oe (do not accept 4 8)	
(c)	B1	LQ = 0.35 stated or implied allow 35 for this mark (but 35.25 or the UQ = 35 th value is B0).	
	M1	0.31 = UQ – “0.35” allow any rearrangement of this for M1 Condone eg 0.6a for UQ Also allow use of 35 + 31 for this mark (implied by 66 or 0.66)	
	A1	cao May come from poor notation eg 0.6a = 0.66 Do not isw. Do not award for 66 or 0.66	
(d)	M1	Use of correct formula (need $\sqrt{\quad}$) $\sqrt{\frac{13.4228}{47} - \frac{23.72^2}{47}}$ is M0	
	A1*	awrt 0.176 with correct exact working seen allow awrt 0.176 coming from 0.1757... or better allow awrt 0.176 with the mean to 0.5046... or better seen Note: Inaccurate working on its own scores M1A0 eg $\sqrt{\frac{13.4228}{47} - 0.505^2} = 0.176$ (note $s = 0.17765...$ send to review)	
(e)(i)	M1	For a correct equation for sample mean or for 0.502×65 (implied by 32.63 seen)	
	A1*	8.91 cso Correctly rearranging $\sum y$ leading to given answer with 1 line of intermediate working. $\sum y = 0.502 \times 65 - 23.72 = 8.91$ is M1A1 Condone poor notation for A1 but working must be correct.	
(ii)	M1	Use of correct standard deviation or variance formula seen For this mark we only require 0.204, 65 and 0.502 in the correct places – allow any numerator	

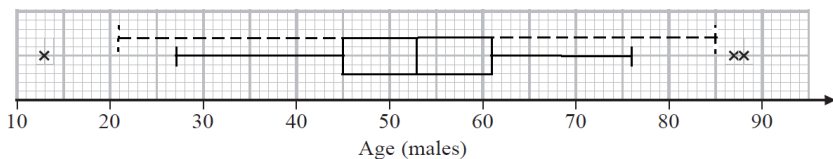
	M1	Find $\sum y^2$ using correct order of operations on $a = \sqrt{\frac{b + \sum y^2}{65}} - c$ or $m = \frac{n + \sum y^2}{65} - p$ $b \neq 0, n \neq 0$ At least 1 line of rearrangement from variance must be shown to score this mark. Condone poor notation for $\sum y^2$ (may be implied by awrt 5.66)
	A1	awrt 5.66 (SC attempting to use s gives 5.62 send to review)



Question Number	Scheme		Marks
2 (a)(i)	$Q_2 = 57$		B1
(ii)	$Q_1 = 45 \quad Q_3 = 63$		B1 B1
			(3)
(b)	'63'+1.5('63'-'45')[= 90] or '45'-1.5('63'-'45')[=18]		M1
	= 90 or = 18		A1ft
	16 and 94 [are outliers]*		A1*
			(3)
(c)	A boxplot drawn with 2 whiskers		M1
	Q_1 , Q_2 and Q_3 plotted correctly		A1ft
	Whiskers drawn correctly		A1ft
	Outliers marked at 16 and 94		A1
			(4)
(d)	The median / Q_2 for February is less/lower than the median / Q_2 for December oe		B1ft
	The IQR / range for February is less/lower than December (allow similar) oe		B1ft
	For a correct interpretation of either average or spread e.g. - on average February weigh less than December oe - the weights of February are less varied/little change in variability than the weights of December oe - They weighed more later in the year oe - Most of the distribution has shifted right, implying that most kangaroos have gained weight but some appear to have lost weight.		B1ft
			(3)
	Notes		Total 13
(a)(i)	B1	Cao	
(ii)	B1	Cao	
	B1	Cao	
(b)	M1	For use of either $Q_3 + 1.5(Q_3 - Q_1)$ or $Q_1 - 1.5(Q_3 - Q_1)$ ft part (a)	
	A1ft	For either 90 or 18 ft part (a)	
	A1*	For identifying both outliers with no incorrect/missing working (This can ft part (a))	
(c)	M1	A boxplot drawn with 2 whiskers	
	A1ft	For Q_1 , Q_2 and Q_3 plotted correctly ft part (a)	
	A1ft	Whiskers drawn at 18 and 90 ft part (b) or 23 and 86	
	A1	Outliers marked at 16 and 94	
(d)	B1ft	A correct comparison of medians ft their boxplot drawn or part (a) (No figures are required but if quoted then they need to be correct ft) Must mention the word median/ Q_2	
	B1ft	A correct comparison of range/IQR ft their boxplot drawn or part (a) (No figures are required but if quoted then they need to be correct ft) Must mention either IQR or range	
	B1ft	A correct interpretation of either the average or the spread ft their boxplot drawn or part (a)	
		NB Ignore any reference to skew or outliers	

Question	Scheme	Marks
1 (a)	$k = 3$	B1
		(1)
(b)	$Q_1 = 39$ $Q_3 = 57$	B1 B1
		(2)
(c)	"57" + 1.5 × ("57" – "39") or "39" – 1.5 × ("57" – "39")	M1
	84 and 12 therefore only 1 outlier [85]	A1
		(2)
(d)		M1 M1 M1 A1
		(4)
(e)	A correct difference of the medians with supporting figures e.g. On average Birch [trees grow slightly] taller as the median is larger 48 > 45 oe or A correct difference of the spread with supporting figures e.g. Maple has a greater spread/variation of heights as the range is larger 55 > 48 (excluding outlier) oe e.g. Birch has a greater spread/variation of heights as the range is larger 57 > 55 (with outlier) oe	B1ft
		(1)
(f)	36,, $a < x$ where 43,, x ,, 45 or 54,, $2a$,, 80	M1
	36,, a ,, "43" and 54,, $2a$,, 80	A1ft
	36,, a ,, 40	A1
		(3)
Notes		Total 13
(a)	B1	Cao
(b)	B1	for Q_1 correct
	B1	for Q_3 correct
(c)	M1	for either method correct or a correct value (ft their Q_1 and their Q_3)
	A1	Both limits for outliers correct and statement about the outlier or the outlier given.
(d)	M1	for a box drawn with only 2 whiskers, only one at each end (condone median line missing)
	M1	for upper whisker ending at 76 (or 84 ft their upper outlier limit) and lower whisker ending at 28
	M1	for Q_1 , Q_2 and Q_3 plotted, with $Q_2 = 48$ and ft their Q_1 and Q_3
	A1	for a fully correct box plot with the outlier correctly shown – must be only 1
(e)	B1ft	for a correct comment, referring to heights , with reference to a correctly named statistic . Must include the figures compared. Allow 'grow more/bigger' to imply taller Ignore any reference to skew SC If Q_1/Q_3 are incorrect then allow a ft comment about spread referring to the difference in IQR if compared to 18
(f)	M1	for either range correct. Allow $72 \leq 2a \leq 80$ or $27 \leq a \leq 40$ for 54,, $2a$,, 80 Condone < rather than ,, May be seen as separate inequalities e.g. $2a \leq 80$, $2a \geq 54$ is allowed for 54,, $2a$,, 80 A final answer of 36,, a ,, 40 or $36 < a < 40$ implies M1
	A1ft	for both ranges correct ft their k . Allow $72 \leq 2a \leq 80$ or $27 \leq a \leq 40$ for 54,, $2a$,, 80 Condone < rather than ,, May be seen as separate inequalities
	A1	Allow 36 to 40 or 36, 37, 38, 39, 40 NB It is possible to get M1A0A1

Question Number	Scheme	Marks
4. (a)	Upper quartile = 34 Lower limit = $24 - 15 = 9$ or upper limit is $"34" + 15 = 49$ So outliers are: 8, 52.5 and 56	B1 M1 A1ft, A1ft (4)
(b)		B1 B1 B1 (3)
(c)	$Q_2 - Q_1 (=6) > ("4" =) Q_3 - Q_2$ or e.g. in words e.g. " Q_3 closer to Q_2 than Q_1 is" So <u>negative</u> (skew)	M1 A1ft (2)
(d)	IQR now $"34" - 26 = 8$ so new outlier limits are $26 - 1.5 \times "8" = 14$ and $"34" + 1.5 \times "8" = 46$	M1
		A1ft A1 (3)
(e)	$[Q_1$ has increased so both above 24 Median same so either side of or on median] So one between 26 and 30 inc $[Q_3$ unchanged so must be either side of $Q_3]$ so one between "34" and 45 inc	B1 B1 (2)
		[14]
	Notes	
(a)	B1 for $Q_3 = 34$ either stated or used/implied (score if seen on box plot) M1 for one correct calculation (ft their 34 for upper limit) [May be implied by correct outliers] 2 nd A1ft for the lower outlier at 8 (ft their limit provided limit ≤ 12) 3 rd A1ft for upper outliers at 52.5 and 56 (ft their limit provided it is > 45) NB These accuracy marks are for the outliers not the limits	Award if their outliers are seen on box plot
(b)	1 st B1 for a box with $Q_1 = 24, Q_2 = 30 Q_3 =$ their 34 and two whiskers one on each side. 2 nd B1 for one lower whisker ending at 10 (or their 9) and outlier at 8 only 3 rd B1 for one upper whisker ending at 45 (or their 49 to match "9") and outliers at 52.5 and 56 only	
SC	Extra whiskers. If one set of whiskers gives a correct box plot award B1B0B0 Usual accuracy for plots – to within 0.5 of a square.	
(c)	M1 for correct comparison of $Q_2 - Q_1$ and $Q_3 - Q_2$ (ft their Q_3) (if no values seen <u>must</u> see comparison otherwise accept correctly assigned 6 and 4 without $>$) A1ft for correct deduction based on their Q_3 (+ve (skew) if their $Q_3 > 36$, <u>no skew</u> if their $Q_3 = 36$)	
(d)	M1 for recognising new IQR and at least one correct new limit (ft their 34, implied by correct plot) 1 st A1ft for a correct lower whisker ending at 15.5 (or their 14) and 2 correct outliers at 8 and 10 2 nd A1 for a <u>fully</u> correct box plot with upper whisker to 45 (or could go to 46 [to match their 14])	
SC	Extra whiskers. If one set of whiskers gives a correct box plot award M1A0A1	
(e)	1 st B1 for a range [26, 30] allow that (...) (o.e. eg between 26 and 30) 2 nd B1 for a range [34, 45] condone [...] or (...) (ft their 34 and allow o.e. e.g. between 34 and 45)	

Question Number	Scheme	Marks
3. (a)	[Median =] <u>53</u>	B1 (1)
(b)	$Q_1 = 45$ $Q_3 = 61$ [IQR =] $61 - 45 = \underline{16}$ (*)	M1 A1cso (2)
(c)	$Q_1 - 1.5 \times (\text{IQR}) = 45 - 1.5 \times 16 [= 21]$ or $Q_3 + 1.5 \times (\text{IQR}) = 61 + 1.5 \times 16 [= 85]$ Outliers are < 21 or > 85 So there are three outliers at <u>13, 87 and 88</u>	M1 A1ft A1 (3)
(d)	 Age (males)	M1 A1ft A1 A1 (4)
(e)	e.g. the females are generally older than the men as median is higher ($67 > 53$)	B1 (1)
(f)(i)	No change to box plot means one in each section so granddaughter <u>34~56</u>	B1
(ii)	Eldest daughter in range $[67 \sim 72]$ or Anja's age $[72 \sim 93]$ Since Anja 23 years older than eldest daughter Anja in range <u>90~93</u>	M1 A1 (3)
Notes		[14]
(a)	B1 for 53	
(b)	M1 for an attempt at both and at least one correct. No need to be labelled. A1cso for both correct quartiles seen and $61 - 45$ leading to 16	
(c)	M1 for attempting at least one of the limits. Can fit their quartiles. May be implied by 85 or 21 1 st A1ft for both outlier limits correct or correct ft using their quartiles 2 nd A1 for identifying the three outliers at 13, 87, 88 (dep on seeing both correct limits)	
(d)	M1 for drawing a box with only two whiskers one at each end 1 st A1ft for Q_1, Q_2 and Q_3 as a correctly drawn box (ft their values for Q_1, Q_2 and Q_3) 2 nd A1 for upper whisker ending at 76 (or 85) and lower whisker ending at 27 (or 21) NB Must be correctly paired. 3 rd A1 for the 3 outliers correctly shown (accuracy - half a small square throughout) SC fully correct but with both whiskers correct on each side. M0A1A0A1	
(e)	B1 for a correct comment, referring to ages , with reference to a correctly named statistic . Must include the figures compared. eg Females older than men and comparison of median, upper quartile or lower quartile, allow Q_1, Q_2 and Q_3 with their figures which must agree with the statement. eg Males ages more spread out than female and comparison of ranges with males = 75 and females = 73 eg Females older than males since Males are symmetrical $[Q_3 - Q_2 : Q_2 - Q_1] 8 : 8$ Female are negative skew 5 : 11 NB use of mean/ IQR/ minimum/ maximum is B0. Ignore incorrect comments.	
(f)(i)	B1 for deducing granddaughter is at or below lower quartile but not below 34 Allow any reasonable adjustment for her mother's age, $\{34 \text{ to } x\}$ where $35 \leq x \leq 56$	
(ii)	M1 Suitable range for eldest daughter or Anja above upper quartile. Ignore any incorrect upper limit. May be implied by a correct range. A1 for a range of $[90 \sim 93]$ for Anja's age	

Question	Scheme			
5 (a)	Speed (km/h)	Number of cars	fd	M1
	20,, $x < 25$	54	10.8	
	25,, $x < 30$	90	18	
	30,, $x < 40$	60	6	
	$90x = 54 \times 10$ or $18x = 10.8 \times 10$			M1
	height of 2 nd tallest bar = 6 (cm)			A1
				(3)
(b)	$30 + 15 + \frac{2}{15} \times 6 = 45.8$, therefore 46 cars.			M1, A1
				(2)
(c)	$[Q_2 = 25 +] \frac{58.5}{90} \times 5$			M1
	$= 28.25$ awrt 28.3			A1
				(2)
(d)(i)	$[Q_3 = 30 +] \frac{24.75}{60} \times 10$			M1
	$= 34.125$ awrt 34.1–34.3			A1
(ii)	“34.125” – 25.14			M1
	$= 8.99$ awrt 9.0 – 9.1			A1
				(4)
(e)	$Q_2 - Q_1(3.11) < Q_3 - Q_2(5.875)$			M1
	therefore, positively skewed			A1
				(2)
(f)	The suggestion is not suitable as part (e) shows the data is skewed/not symmetric .			B1ft
				(1)
				Total 14
(a)	M1 for evidence of $fd = f/w$ or for identifying/using 25,, $x < 30$ as the tallest bar or 20,, $x < 25$ as the 2 nd tallest bar or an answer of awrt 3.33			
	M1 identifying/using 25,, $x < 30$ as the tallest bar and 20,, $x < 25$ as the 2 nd tallest bar. Implied by correct working seen or correct answer. Allow equivalent ratio equation.			
	A1 6 cao Be careful the f.d. of the 3 rd bar is 6, so $60 \div 10 = 6$ on its own is M1M0A0. Answer only scores M1M0A0			
(b)	M1 for $30 + 15 + k$ where $0.5 \leq k \leq 1$ (implied by 45.8) or $k + 219 - (54 + 90 + 30)$			
	A1cao 45 or 46 (must be an integer). Correct answer only scores 2 out of 2.			
(c)	M1 for $\frac{58.5}{90} \times 5$ or $\frac{m-25}{112.5-54} = \frac{30-25}{144-54}$ or $\frac{31.5}{90} \times 5$ or $\frac{30-m}{144-112.5} = \frac{5}{90}$			
	Allow use of 113 instead of 112.5			
	A1 for awrt 28.3			
(d)(i)	M1 for $\frac{24.75}{60} \times 10$ or $\frac{Q_3-30}{168.75-144} = \frac{40-30}{204-144}$ or $\frac{35.25}{60} \times 10$ or $\frac{40-m}{204-168.75} = \frac{10}{60}$			
	Allow use of any value from 168.75 to 169.5			
	A1 awrt 34.1 to awrt 34.3 an answer in this range scores 2 out of 2 unless obvious incorrect working			
(ii)	M1 for “their (d)(i)” substituted into “ Q_3 ” – awrt 25.1			
	A1 awrt 9.0 or awrt 9.1 Allow 9 if correct working seen.			
(e)	M1 A correct comparison which must follow from their figures. Must have $25.14 < “Q_2” < “Q_3”$			
	A1 dep on M1 and dep on their figures being consistent with positive skew. Negative skew is A0. Allow calculation of estimated mean awrt $30.6 > “28.3”$ so +ve for M1A1			
(f)	B1ft For not suitable and a correct supporting reason which is consistent with “their (e)”			