EXPERT TUITION

Maths Questions By Topic:

Data Presentation & Interpretation

A-Level Edexcel

0207 060 4494
www.expert-tuition.co.uk
online.expert-tuition.co.uk
enquiries@expert-tuition.co.uk
The Foundry, 77 Fulham Palace Road, W6 8JA

Table Of Contents

1. The relationship between two variables p and t is modelled by the regression line with equation

$$p = 22 - 1.1\,t$$

The model is based on observations of the independent variable, t, between 1 and 10

(a) Describe the correlation between p and t implied by this model.

(1)

Given that p is measured in centimetres and t is measured in days,

(b) state the units of the gradient of the regression line.

(1)

Using the model,

(c) calculate the change in p over a 3-day period.

(2)

Tisam uses this model to estimate the value of p when $t = 19$

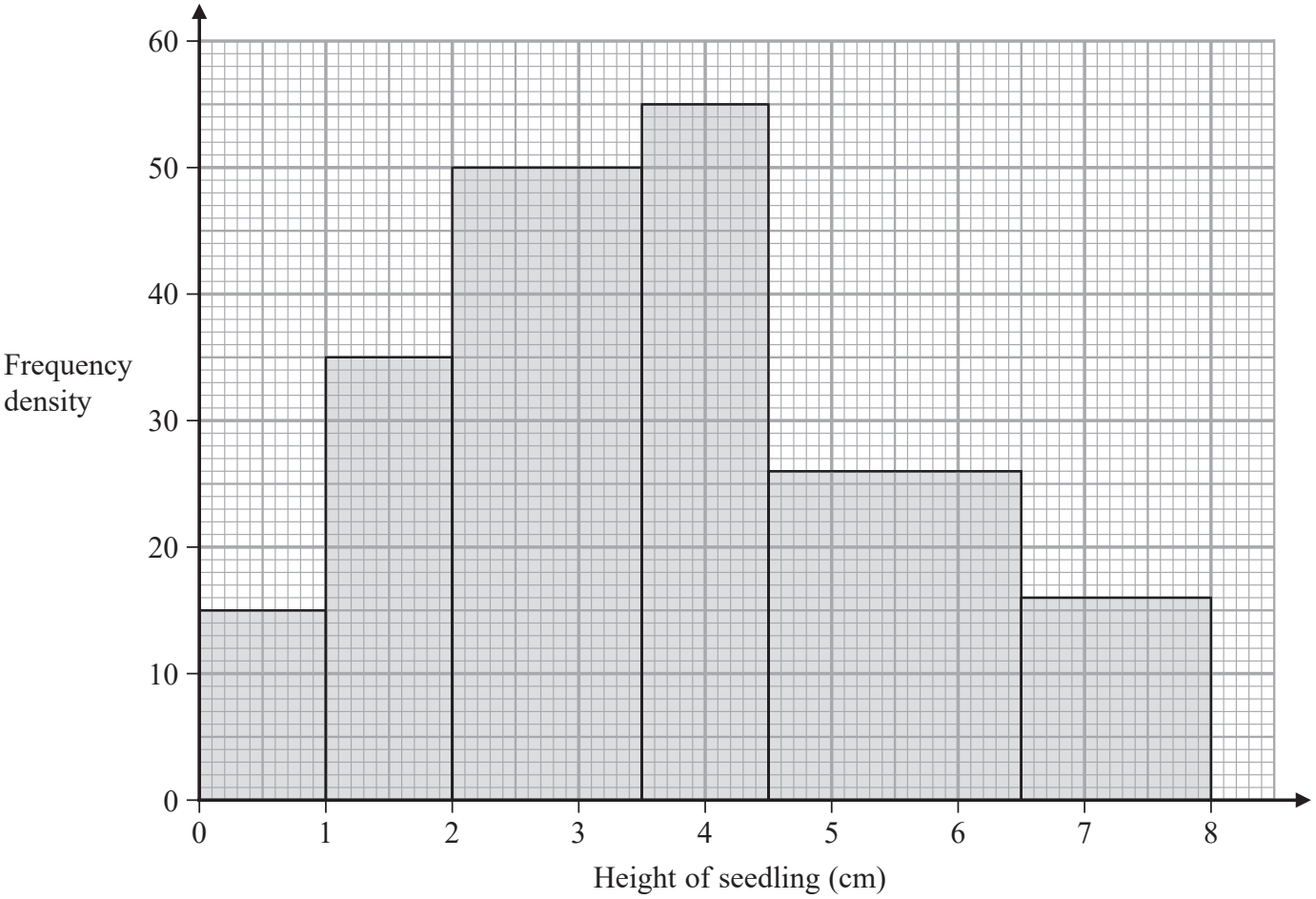
(d) Comment, giving a reason, on the reliability of this estimate.

(1)

Question 1 continued

(Total for Question 1 is 5 marks)

2. The histogram summarises the heights of 256 seedlings two weeks after they were planted.



(a) Use linear interpolation to estimate the median height of the seedlings.

(4)

Chris decides to model the **frequency density** for these 256 seedlings by a curve with equation

$$y = kx(8 - x) \qquad 0 \leqslant x \leqslant 8$$

where k is a constant.

(b) Find the value of k

(3)

Using this model,

(c) write down the median height of the seedlings.

(1)

Question 2 continued

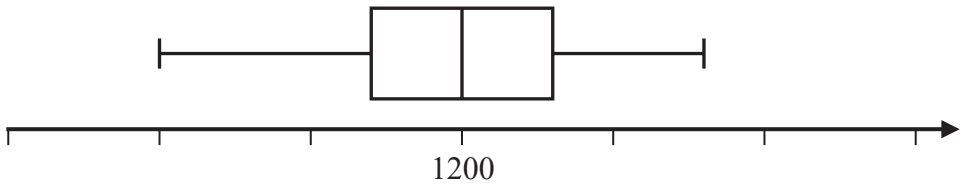
Question 2 continued

(Total for Question 2 is 8 marks)

3. Jiang is studying the variable Daily Mean Pressure from the large data set.

He drew the following box and whisker plot for these data for one of the months for one location using a linear scale but

- he failed to label all the values on the scale
- he gave an incorrect value for the median



Daily Mean Pressure (hPa)

Using your knowledge of the large data set, suggest a suitable value for

(a) the median,

(1)

(b) the range.

(1)

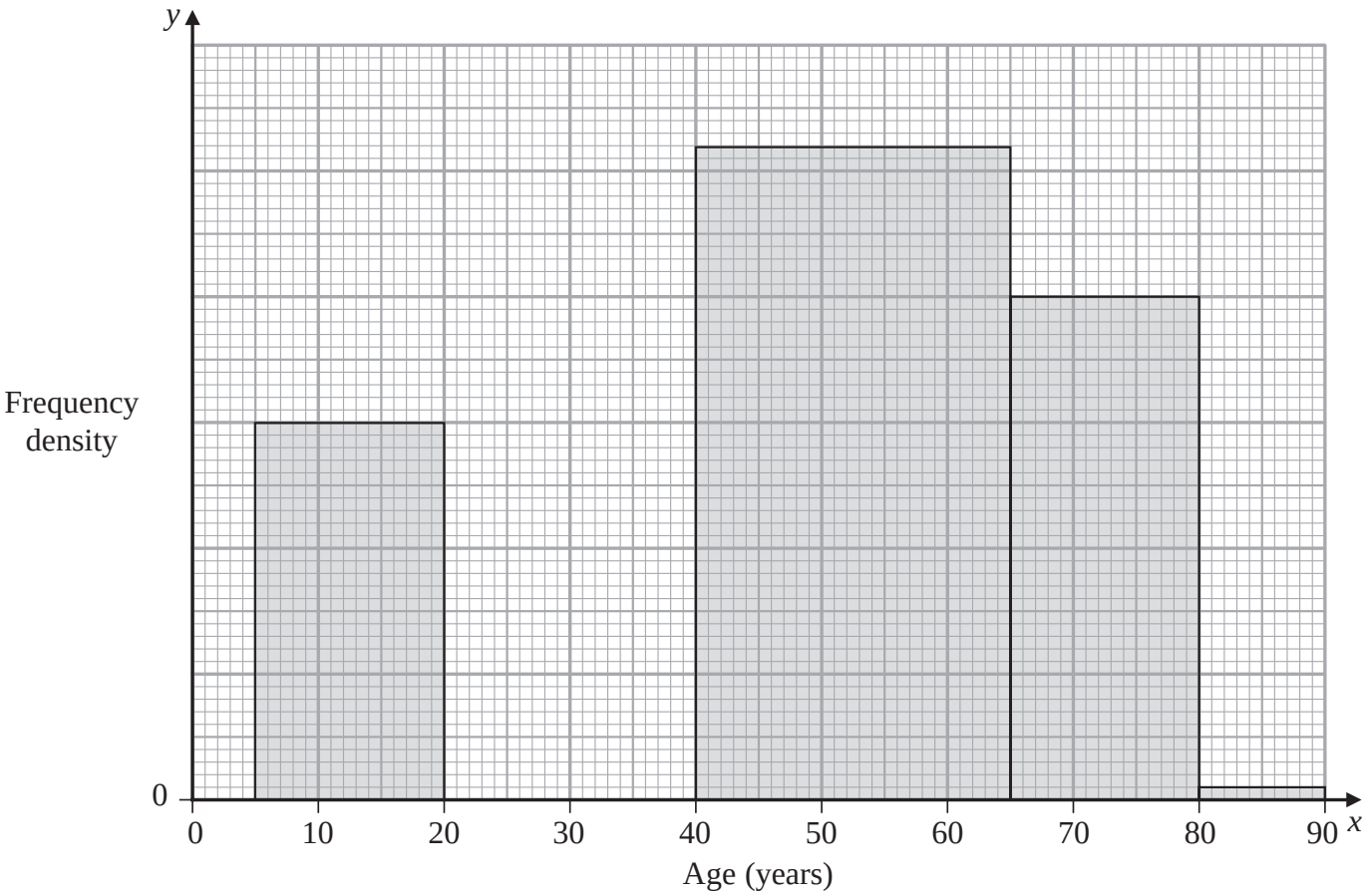
(*You are not expected to have memorised values from the large data set. The question is simply looking for sensible answers*.)

(Total for Question 3 is 2 marks)

4. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

There were no passengers aged 90 or over.

Age (x years)	$0 \leqslant x < 5$	$5 \leqslant x < 20$	$20 \leqslant x < 40$	$40 \leqslant x < 65$	$65 \leqslant x < 80$	$80 \leqslant x < 90$
Frequency	5	45	90			1



(a) Complete the histogram.

(3)

(b) Use linear interpolation to estimate the median age.

(4)

An outlier is defined as a value greater than $Q_3 + 1.5 \times$ interquartile range.

Given that $Q_1 = 27.3$ and $Q_3 = 58.9$

(c) determine, giving a reason, whether or not the oldest passenger could be considered as an outlier.

(2)

Question 4 continued

Question 4 continued

Question 4 continued

(Total for Question 4 is 9 marks)

5.

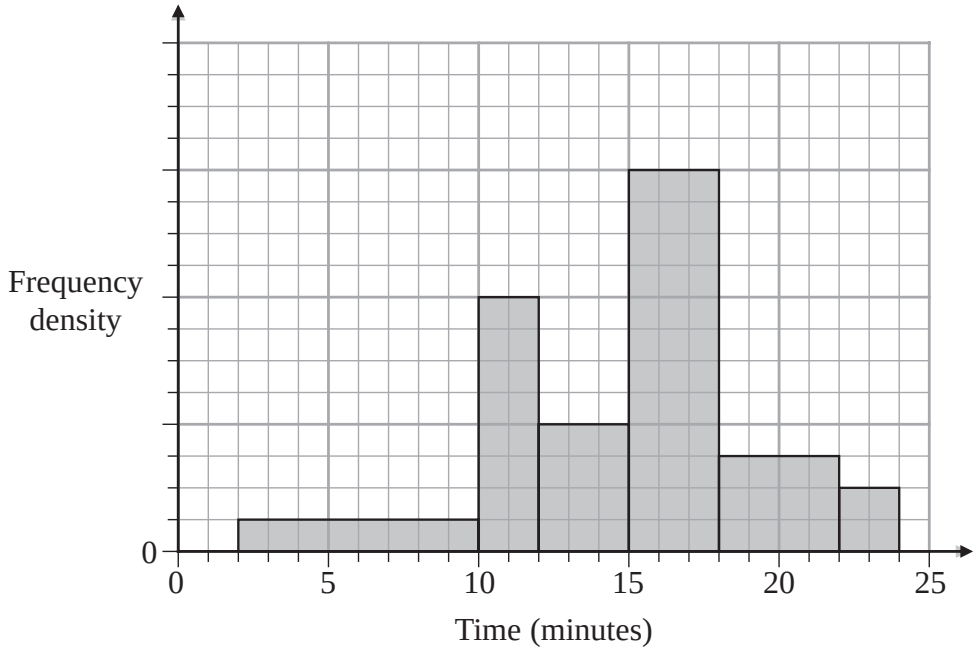


Figure 1

The histogram in Figure 1 shows the times taken to complete a crossword by a random sample of students.

The number of students who completed the crossword in more than 15 minutes is 78

Estimate the percentage of students who took less than 11 minutes to complete the crossword.

(4)

Question 5 continued

(Total for Question 5 is 4 marks)

6. Jerry is studying visibility for Camborne using the large data set June 1987.

The table below contains two extracts from the large data set.

It shows the daily maximum relative humidity and the daily mean visibility.

Date	Daily Maximum Relative Humidity	Daily Mean Visibility
Units	**%**	
10/06/1987	90	5300
28/06/1987	100	0

(The units for Daily Mean Visibility are deliberately omitted.)

Given that daily mean visibility is given to the nearest 100,

(a) write down the range of distances in metres that corresponds to the recorded value 0 for the daily mean visibility.

(1)

Jerry drew the following scatter diagram, Figure 2, and calculated some statistics using the June 1987 data for Camborne from the large data set.

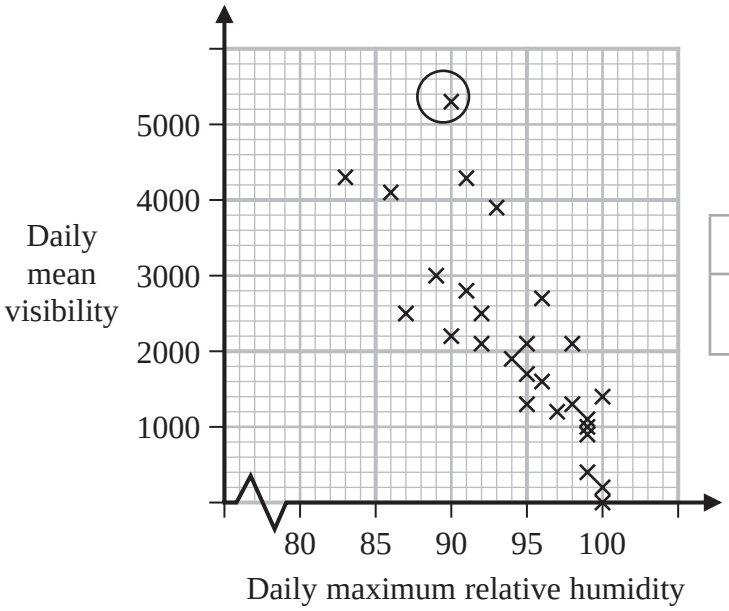


Figure 2

	Q_1	IQR
Daily mean visibility	1100	1600
Daily maximum relative humidity (%)	92	8

Jerry defines an outlier as a value that is more than 1.5 times the interquartile range above Q_3 or more than 1.5 times the interquartile range below Q_1.

(b) Show that the point circled on the scatter diagram is an outlier for visibility.

(2)

(c) Interpret the correlation between the daily mean visibility and the daily maximum relative humidity.

(1)

Jerry drew the following scatter diagram, Figure 3, using the June 1987 data for Camborne from the large data set, but forgot to label the *x*–axis.

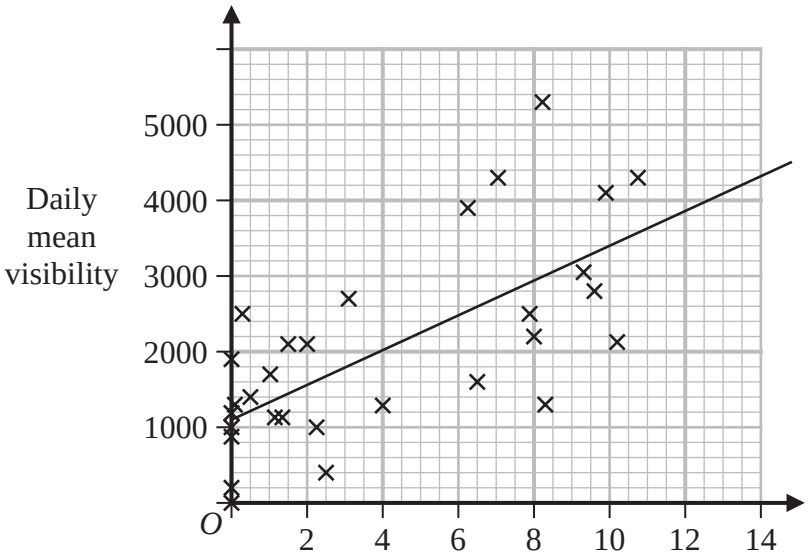


Figure 3

(d) Using your knowledge of the large data set, suggest which variable the *x*-axis on this scatter diagram represents.

(1)

Question 6 continued

Question 6 continued

(Total for Question 6 is 5 marks)

7. A lake contains three different types of carp.

There are an estimated 450 mirror carp, 300 leather carp and 850 common carp.

Tim wishes to investigate the health of the fish in the lake.

He decides to take a sample of 160 fish.

As part of the health check, Tim weighed the fish.

His results are given in the table below.

Weight (w kg)	**Frequency (f)**	**Midpoint (m kg)**
$2 \leqslant w < 3.5$	8	2.75
$3.5 \leqslant w < 4$	32	3.75
$4 \leqslant w < 4.5$	64	4.25
$4.5 \leqslant w < 5$	40	4.75
$5 \leqslant w < 6$	16	5.5

(You may use $\sum fm$ 692 and $\sum fm^2$ 3053)

(a) Calculate an estimate for the standard deviation of the weight of the carp.

(2)

Tim realised that he had transposed the figures for 2 of the weights of the fish.

He had recorded in the table 2.3 instead of 3.2 and 4.6 instead of 6.4

(b) Without calculating a new estimate for the standard deviation, state what effect

(i) using the correct figure of 3.2 instead of 2.3

(ii) using the correct figure of 6.4 instead of 4.6

would have on your estimated standard deviation.

Give a reason for each of your answers.

(2)

Question 7 continued

Question 7 continued

Question 7 continued

(Total for Question 7 is 4 marks)

8. A sixth form college has 84 students in Year 12 and 56 students in Year 13

The head teacher selects a stratified sample of 40 students, stratified by year group.

The head teacher is investigating the relationship between the amount of sleep, s hours, that each student had the night before they took an aptitude test and their performance in the test, p marks.
For the sample of 40 students, he finds the equation of the regression line of p on s to be

$$p = 26.1 + 5.60s$$

(a) With reference to this equation, describe the effect that an extra 0.5 hours of sleep may have, on average, on a student's performance in the aptitude test.

(1)

(b) Describe one limitation of this regression model.

(1)

(Total for Question 8 is 2 marks)

9. Joshua is investigating the daily total rainfall in Hurn for May to October 2015

Using the information from the large data set, Joshua wishes to calculate the mean of the daily total rainfall in Hurn for May to October 2015

(a) Using your knowledge of the large data set, explain why Joshua needs to clean the data before calculating the mean.

(1)

Using the information from the large data set, he produces the grouped frequency table below.

Daily total rainfall (r mm)	**Frequency**	**Midpoint (x mm)**
$0 \leqslant r < 0.5$	121	0.25
$0.5 \leqslant r < 1.0$	10	0.75
$1.0 \leqslant r < 5.0$	24	3.0
$5.0 \leqslant r < 10.0$	12	7.5
$10.0 \leqslant r < 30.0$	17	20.0

You may use $\sum fx = 539.75$ and $\sum fx^2 = 7704.1875$

(b) Use linear interpolation to calculate an estimate for the upper quartile of the daily total rainfall.

(2)

(c) Calculate an estimate for the standard deviation of the daily total rainfall in Hurn for May to October 2015

(2)

(d) (i) State the assumption involved with using class midpoints to calculate an estimate of a mean from a grouped frequency table.

(ii) Using your knowledge of the large data set, explain why this assumption does not hold in this case.

(iii) State, giving a reason, whether you would expect the actual mean daily total rainfall in Hurn for May to October 2015 to be larger than, smaller than or the same as an estimate based on the grouped frequency table.

(3)

Question 9 continued

Question 9 continued

Question 9 continued

(Total for Question 9 is 8 marks)

10. A company is introducing a job evaluation scheme. Points (x) will be awarded to each job based on the qualifications and skills needed and the level of responsibility. Pay (£y) will then be allocated to each job according to the number of points awarded.

Before the scheme is introduced, a random sample of 8 employees was taken and the linear regression equation of pay on points was $y = 4.5x - 47$

(a) Describe the correlation between points and pay.

(1)

(b) Give an interpretation of the gradient of this regression line.

(1)

(c) Explain why this model might not be appropriate for all jobs in the company.

(1)

Question 10 continued

(Total for Question 10 is 3 marks)

11. Helen is studying the daily mean wind speed for Camborne using the large data set from 1987. The data for one month are summarised in Table 1 below.

Windspeed	n/a	6	7	8	9	11	12	13	14	16
Frequency	13	2	3	2	2	3	1	2	1	2

Table 1

(a) Calculate the mean for these data.

(1)

(b) Calculate the standard deviation for these data and state the units.

(2)

The means and standard deviations of the daily mean wind speed for the other months from the large data set for Camborne in 1987 are given in Table 2 below. The data are not in month order.

Month	*A*	*B*	*C*	*D*	*E*
Mean	7.58	8.26	8.57	8.57	11.57
Standard Deviation	2.93	3.89	3.46	3.87	4.64

Table 2

(c) Using your knowledge of the large data set, suggest, giving a reason, which month had a mean of 11.57

(2)

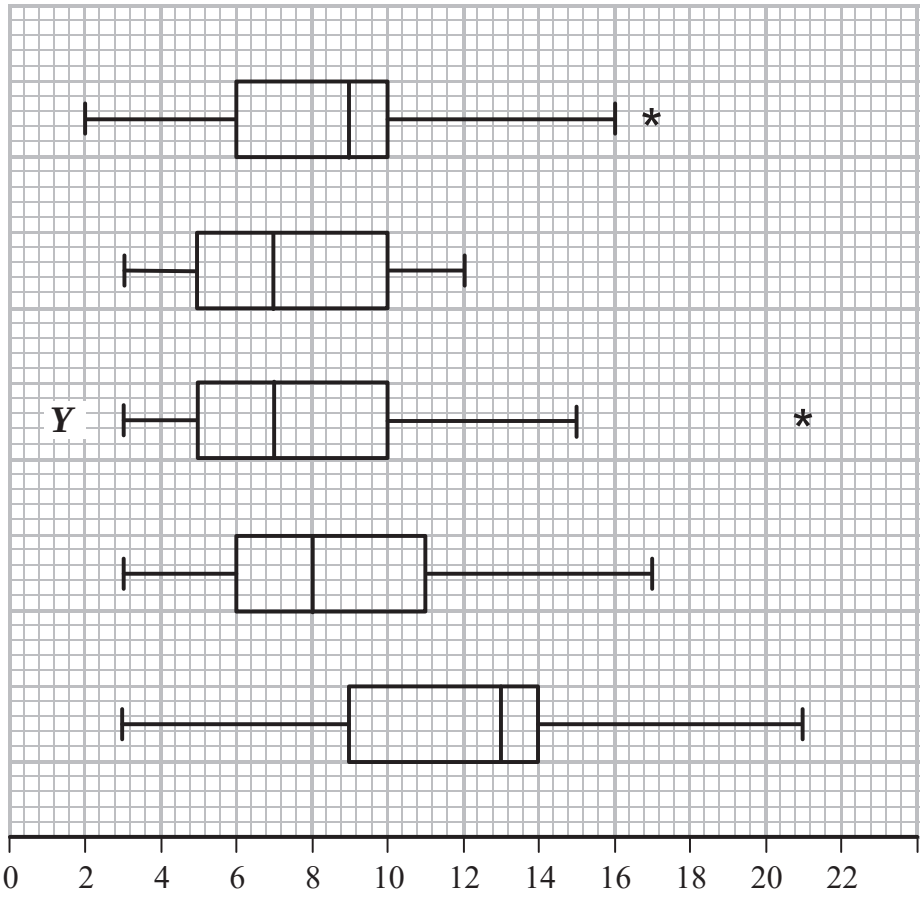
The data for these months are summarised in the box plots on the opposite page.
They are not in month order or the same order as in Table 2.

(d) (i) State the meaning of the * symbol on some of the box plots.

(ii) Suggest, giving your reasons, which of the months in Table 2 is most likely to be summarised in the box plot marked *Y*.

(3)

Question 11 continued



(Total for Question 11 is 8 marks)

12. A company manager is investigating the time taken, t minutes, to complete an aptitude test. The human resources manager produced the table below of coded times, x minutes, for a random sample of 30 applicants.

Coded time (x minutes)	**Frequency (f)**	**Coded time midpoint (y minutes)**
$0 \leq x < 5$	3	2.5
$5 \leq x < 10$	15	7.5
$10 \leq x < 15$	2	12.5
$15 \leq x < 25$	9	20
$25 \leq x < 35$	1	30

(You may use $\sum fy = 355$ and $\sum f y^2 = 5675$)

(a) Use linear interpolation to estimate the median of the coded times.

(2)

(b) Estimate the standard deviation of the coded times.

(2)

The company manager is told by the human resources manager that he subtracted 15 from each of the times and then divided by 2, to calculate the coded times.

(c) Calculate an estimate for the median and the standard deviation of t.

(3)

The following year, the company has 25 positions available. The company manager decides not to offer a position to any applicant who takes 35 minutes or more to complete the aptitude test.

The company has 60 applicants.

(d) Comment on whether or not the company manager's decision will result in the company being able to fill the 25 positions available from these 60 applicants. Give a reason for your answer.

(2)

Question 12 continued

(Total for Question 12 is 9 marks)

13. Sara is investigating the variation in daily maximum gust, t kn, for Camborne in June and July 1987.
She used the large data set to select a sample of size 20 from the June and July data for 1987. Sara selected the first value using a random number from 1 to 4 and then selected every third value after that.

The data Sara collected are summarised as follows

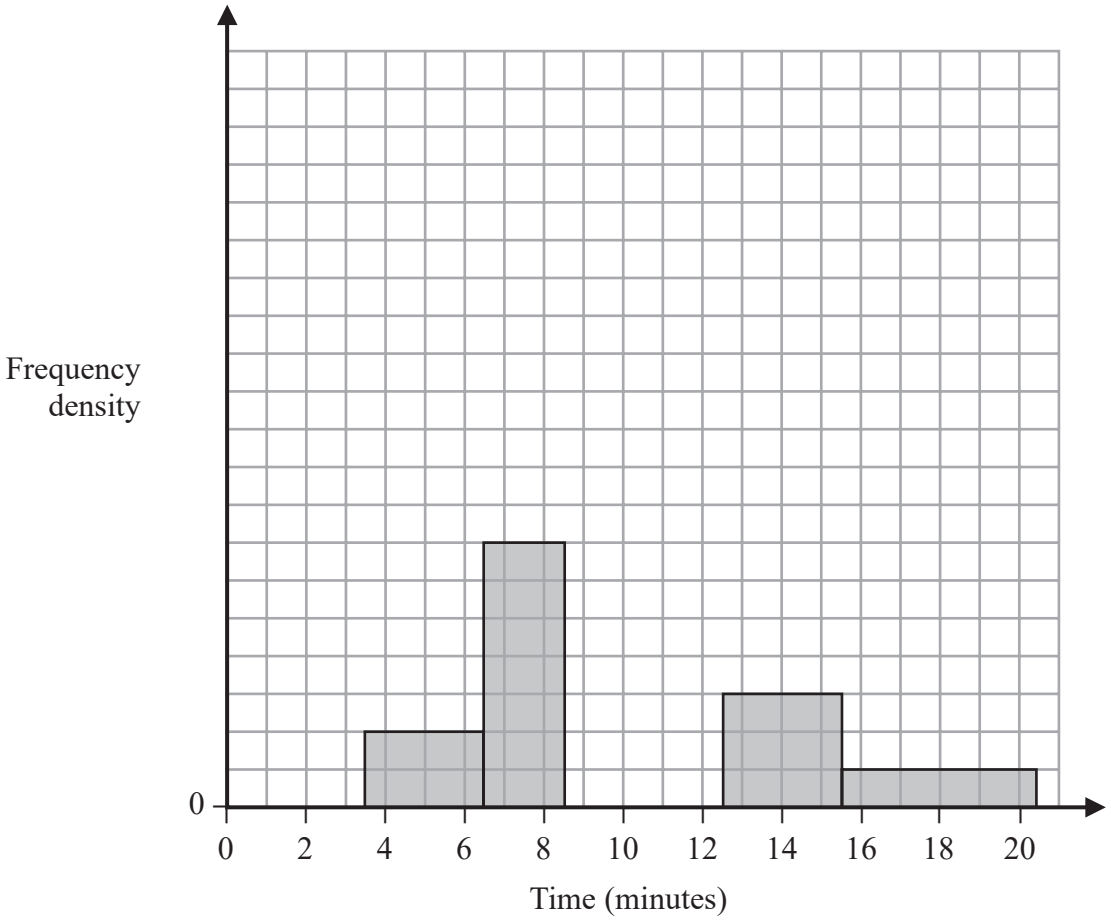
$$n = 20 \qquad \sum t = 374 \qquad \sum t^2 = 7600$$

Calculate the standard deviation.

(2)

(Total for Question 13 is 2 marks)

14. The partially completed histogram and the partially completed table show the time, to the nearest minute, that a random sample of motorists was delayed by roadworks on a stretch of motorway.



Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	17
10 – 12	45
13 – 15	9
16 – 20	

Estimate the percentage of these motorists who were delayed by the roadworks for between 8.5 and 13.5 minutes.

(5)

Question 14 continued

(Total for Question 14 is 5 marks)

15. Sara was studying the relationship between rainfall, r mm, and humidity, h%, in the UK. She takes a random sample of 11 days from May 1987 for Leuchars from the large data set.

She obtained the following results.

h	93	86	95	97	86	94	97	97	87	97	86
r	1.1	0.3	3.7	20.6	0	0	2.4	1.1	0.1	0.9	0.1

Sara examined the rainfall figures and found

$$Q_1 = 0.1 \qquad Q_2 = 0.9 \qquad Q_3 = 2.4$$

A value that is more than 1.5 times the interquartile range (IQR) above Q_3 is called an outlier.

(a) Show that $r = 20.6$ is an outlier.

(1)

(b) Give a reason why Sara might:

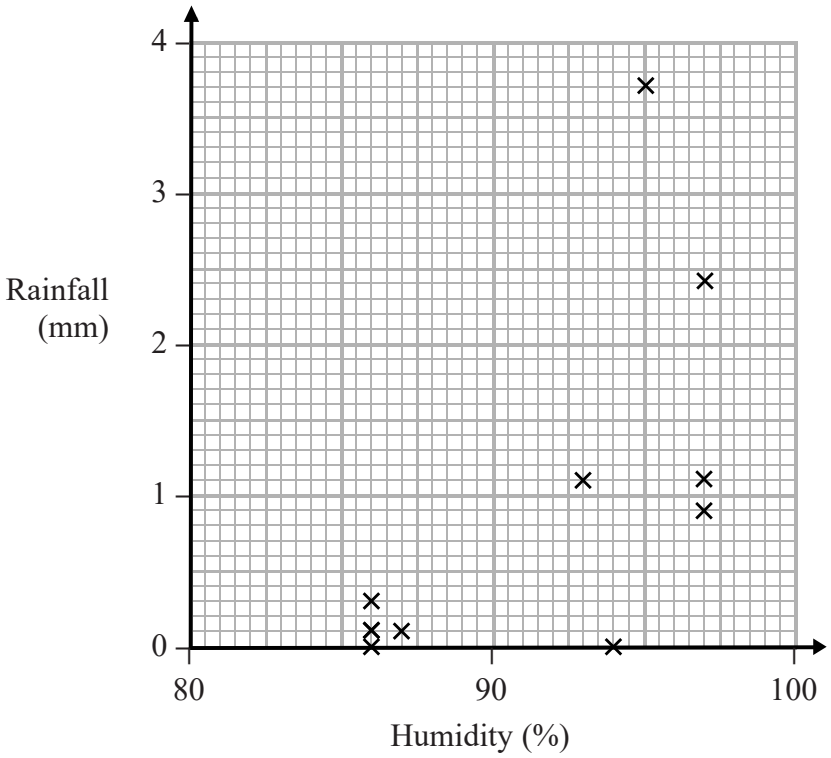
(i) include

(ii) exclude

this day's reading.

(2)

Sara decided to exclude this day's reading and drew the following scatter diagram for the remaining 10 days' values of r and h.



(c) Give an interpretation of the correlation between rainfall and humidity.

(1)

Question 15 continued

The equation of the regression line of r on h for these 10 days is $r = -12.8 + 0.15h$

(d) Give an interpretation of the gradient of this regression line.

(1)

(Total for Question 15 is 5 marks)

16. Dian uses the large data set to investigate the Daily Total Rainfall, r mm, for Camborne.

(a) Write down how a value of $0 < r \leqslant 0.05$ is recorded in the large data set.

(1)

Dian uses the data for the 31 days of August 2015 for Camborne and calculates the following statistics

$$n = 31 \qquad \sum r = 174.9 \qquad \sum r^2 = 3523.283$$

(b) Use these statistics to calculate

(i) the mean of the Daily Total Rainfall in Camborne for August 2015,

(ii) the standard deviation of the Daily Total Rainfall in Camborne for August 2015.

(3)

Dian believes that the mean Daily Total Rainfall in August is less in the South of the UK than in the North of the UK.
The mean Daily Total Rainfall in Leuchars for August 2015 is 1.72 mm to 2 decimal places.

(c) State, giving a reason, whether this provides evidence to support Dian's belief.

(2)

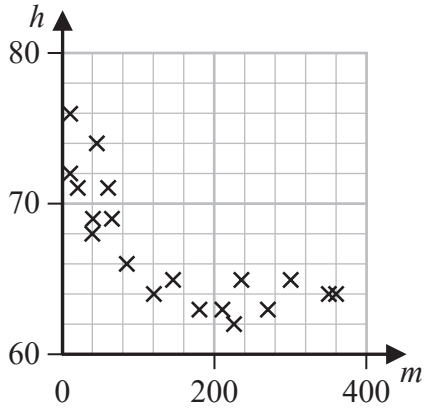
Question 16 continued

(Total for Question 16 is 6 marks)

17. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram.



(a) Interpret the nature of the relationship between h and m

(1)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

(b) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

Question 17 continued

Question 17 continued

(Total for Question 17 is 6 marks)

18. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x, in their last name
- the number of letters, y, in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y.

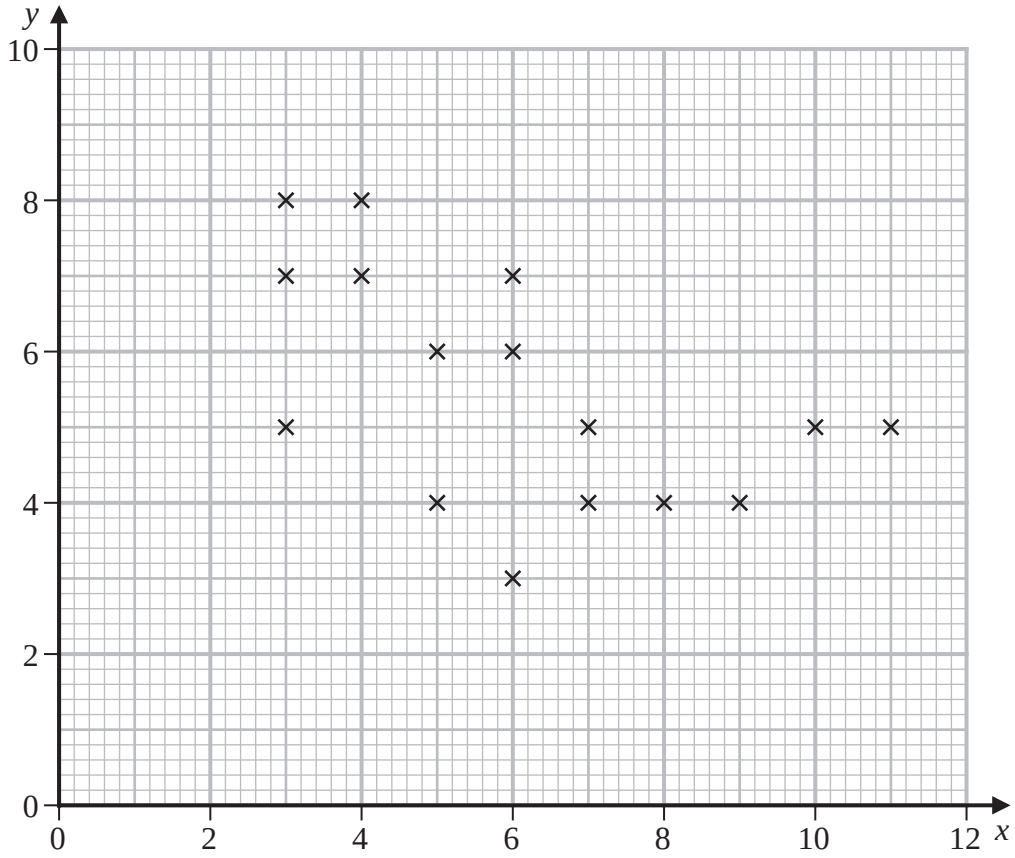
(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

Question 18 continued.



Question 18 continued

Question 18 continued.

(Total for Question 18 is 2 marks)

19. Stav is studying the large data set for September 2015

He codes the variable Daily Mean Pressure, x, using the formula $y = x - 1010$

The data for all 30 days from Hurn are summarised by

$$\sum y = 214 \qquad \sum y^2 = 5912$$

(a) State the units of the variable x

(1)

(b) Find the mean Daily Mean Pressure for these 30 days.

(2)

(c) Find the standard deviation of Daily Mean Pressure for these 30 days.

(3)

Stav knows that, in the UK, winds circulate

- in a **clockwise** direction around a region of **high** pressure
- in an **anticlockwise** direction around a region of **low** pressure

The table gives the Daily Mean Pressure for 3 locations from the large data set on 26/09/2015

Location	Heathrow	Hurn	Leuchars
Daily Mean Pressure	1029	1028	1028
Cardinal Wind Direction			

The Cardinal Wind Directions for these 3 locations on 26/09/2015 were, in random order,

W NE E

You may assume that these 3 locations were under a single region of pressure.

(d) Using your knowledge of the large data set, place each of these Cardinal Wind Directions in the correct location in the table.
Give a reason for your answer.

(2)

Question 19 continued

(Total for Question 19 is 8 marks)

20. A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

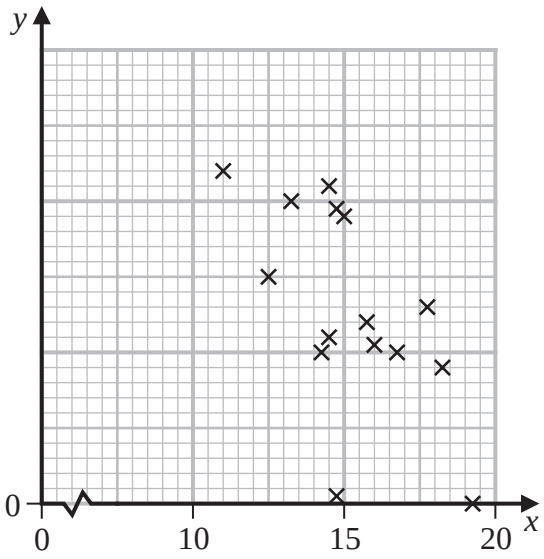


Figure 1

(a) Describe the correlation.

(1)

The variable on the *x*-axis is Daily Mean Temperature measured in °C.

(b) Using your knowledge of the large data set,

(i) suggest which variable is on the *y*-axis,

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(c) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

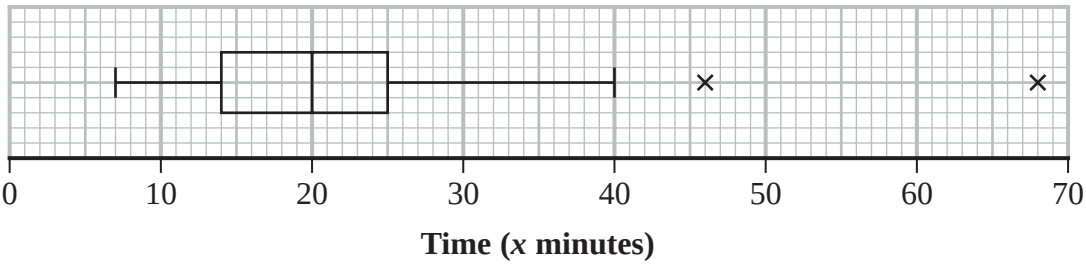
Question 20 continued

(Total for Question 20 is 4 marks)

21. Each member of a group of 27 people was timed when completing a puzzle.

The time taken, x minutes, for each member of the group was recorded.

These times are summarised in the following box and whisker plot.



Time (x minutes)

(a) Find the range of the times.

(1)

(b) Find the interquartile range of the times.

(1)

For these 27 people $\sum x = 607.5$ and $\sum x^2 = 17\,623.25$

(c) calculate the mean time taken to complete the puzzle,

(1)

(d) calculate the standard deviation of the times taken to complete the puzzle.

(2)

Taruni defines an outlier as a value more than 3 standard deviations above the mean.

(e) State how many outliers Taruni would say there are in these data, giving a reason for your answer.

(1)

Adam and Beth also completed the puzzle in a minutes and b minutes respectively, where $a > b$.
When their times are included with the data of the other 27 people

- the median time increases
- the mean time does not change

(f) Suggest a possible value for a and a possible value for b, explaining how your values satisfy the above conditions.

(3)

(g) Without carrying out any further calculations, explain why the standard deviation of all 29 times will be lower than your answer to part (d).

(1)

Question 21 continued

Question 21 continued

Question 21 continued

(Total for Question 21 is 10 marks)

22.

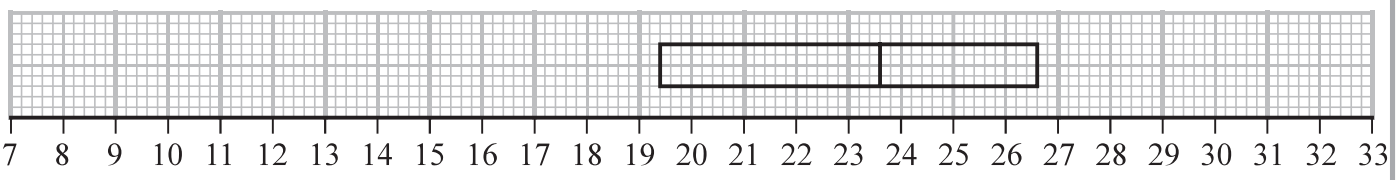


Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value
more than 1.5 × IQR below Q_1 or
more than 1.5 × IQR above Q_3

The three lowest air temperatures in the data set are 7.6 °C, 8.1 °C and 9.1 °C
The highest air temperature in the data set is 32.5 °C

(a) Complete the box plot in Figure 1 showing clearly any outliers.

(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.

(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, x °C, for Beijing in 2015

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19 °C

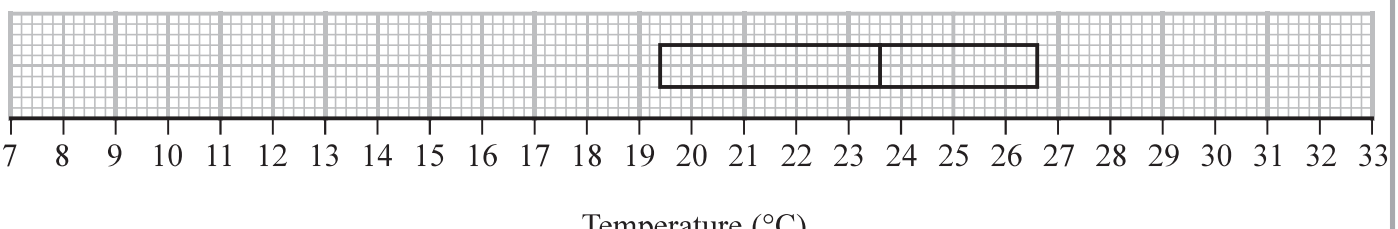
(1)

Question 22 continued

Question 22 continued

Question 22 continued

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

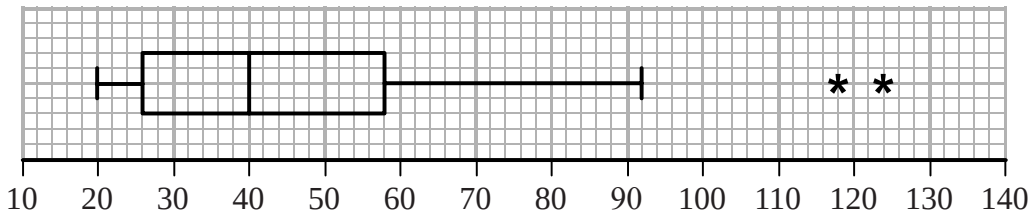


(Total for Question 22 is 6 marks)

23. Charlie is studying the time it takes members of his company to travel to the office. He stands by the door to the office from 08 40 to 08 50 one morning and asks workers, as they arrive, how long their journey was.

Taruni decided to ask every member of the company the time, x minutes, it takes them to travel to the office.

Taruni's results are summarised by the box plot and summary statistics below.



Journey time (minutes)

$n = 95 \quad \sum x = 4133 \quad \sum x^2 = 202\,294$

(a) Write down the interquartile range for these data.

(1)

(b) Calculate the mean and the standard deviation for these data.

(3)

(c) State, giving a reason, whether you would recommend using the mean and standard deviation or the median and interquartile range to describe these data.

(2)

Rana and David both work for the company and have both moved house since Taruni collected her data.

Rana's journey to work has changed from 75 minutes to 35 minutes and David's journey to work has changed from 60 minutes to 33 minutes.

Taruni drew her box plot again and only had to change two values.

(d) Explain which two values Taruni must have changed and whether each of these values has increased or decreased.

(3)

Question 23 continued

Question 23 continued

Question 23 continued

(Total for Question 23 is 9 marks)

24. *Kaff* coffee is sold in packets. A seller measures the masses of the contents of a random sample of 90 packets of *Kaff* coffee from her stock. The results are shown in the table below.

Mass w (g)	Midpoint y (g)	Frequency f
$240 \leq w < 245$	242.5	8
$245 \leq w < 248$	246.5	15
$248 \leq w < 252$	250.0	35
$252 \leq w < 255$	253.5	23
$255 \leq w < 260$	257.5	9

(You may use $\sum fy^2 = 5\ 644\ 171.75$)

A histogram is drawn and the class $245 \leq w < 248$ is represented by a rectangle of width 1.2 cm and height 10 cm.

(a) Calculate the width and the height of the rectangle representing the class $255 \leq w < 260$.

(3)

(b) Use linear interpolation to estimate the median mass of the contents of a packet of *Kaff* coffee to 1 decimal place.

(2)

(c) Estimate the mean and the standard deviation of the mass of the contents of a packet of *Kaff* coffee to 1 decimal place.

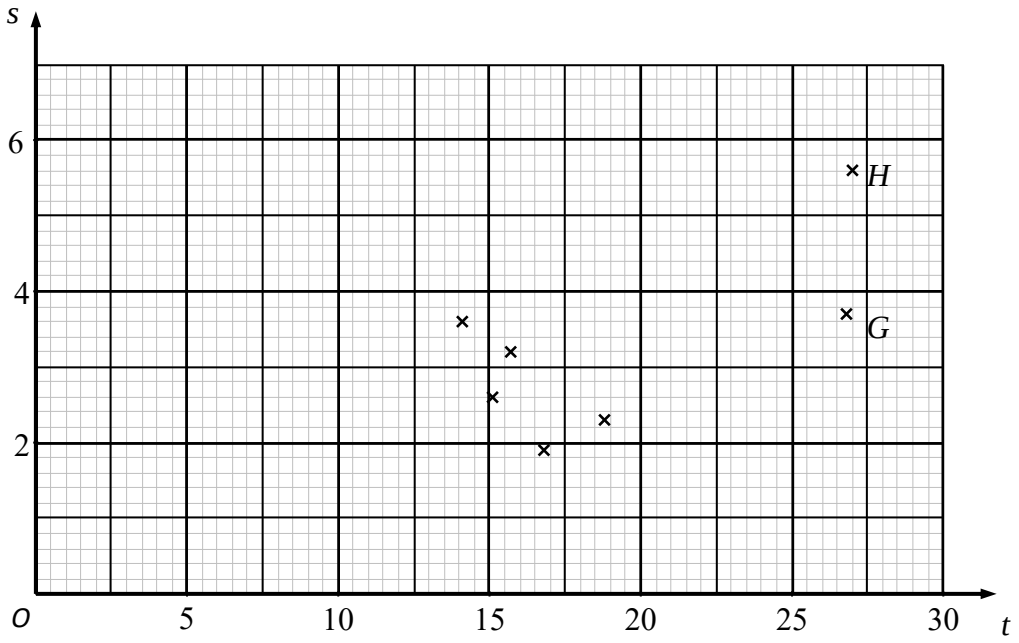
(3)

Question 24 continued

Question 24 continued

(Total for Question 24 is 8 marks)

25. A researcher believes that there is a linear relationship between daily mean temperature and daily total rainfall. The 7 places in the northern hemisphere from the large data set are used. The mean of the daily mean temperatures, t °C, and the mean of the daily total rainfall, s mm, for the month of July in 2015 are shown on the scatter diagram below.



(a) With reference to the scatter diagram, explain why a linear regression model may not be suitable for the relationship between t and s.

(1)

(b) Using your knowledge of the large data set, suggest the names of the 2 places labelled G and H.

(1)

(c) Using your knowledge from the large data set, and with reference to the locations of the two places labelled G and H, give a reason why these places have the highest temperatures in July.

(1)

(d) Suggest how you could make better use of the large data set to investigate the relationship between daily mean temperature and daily total rainfall.

(1)

Question 25 continued

(Total for Question 25 is 4 marks)

26. The number of hours of sunshine each day, y, for the month of July at Heathrow are summarised in the table below.

Hours	$0 \leqslant y < 5$	$5 \leqslant y < 8$	$8 \leqslant y < 11$	$11 \leqslant y < 12$	$12 \leqslant y < 14$
Frequency	12	6	8	3	2

A histogram was drawn to represent these data. The $8 \leqslant y < 11$ group was represented by a bar of width 1.5 cm and height 8 cm.

(a) Find the width and the height of the $0 \leqslant y < 5$ group.

(3)

(b) Use your calculator to estimate the mean and the standard deviation of the number of hours of sunshine each day, for the month of July at Heathrow.
Give your answers to 3 significant figures.

(3)

The mean and standard deviation for the number of hours of daily sunshine for the same month in Hurn are 5.98 hours and 4.12 hours respectably.
Thomas believes that the further south you are the more consistent should be the number of hours of daily sunshine.

(c) State, giving a reason, whether or not the calculations in part (b) support Thomas' belief.

(2)

(d) Estimate the number of days in July at Heathrow where the number of hours of sunshine is more than 1 standard deviation above the mean.

(2)

Question 26 continued

(Total for Question 26 is 10 marks)

Leave blank

27. The following grouped frequency distribution summarises the number of minutes, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway one Monday.

Delay (minutes)	**Number of motorists (f)**	**Delay midpoint (x)**
3–6	38	4.5
7–8	25	7.5
9–10	18	9.5
11–15	12	13
16–20	7	18

(You may use $\sum fx^2 = 8096.25$)

A histogram has been drawn to represent these data.

The bar representing a delay of (3–6) minutes has a width of 2 cm and a height of 9.5 cm.

(a) Calculate the width and the height of the bar representing a delay of (11–15) minutes.

(3)

(b) Use linear interpolation to estimate the median delay.

(2)

(c) Calculate an estimate of the mean delay.

(2)

(d) Calculate an estimate of the standard deviation of the delays.

(2)

Leave blank

Question 27 continued

Leave blank

Question 27 continued

Leave blank

Question 27 continued

(Total 9 marks)

Leave blank

28. An estate agent is studying the cost of office space in London. He takes a random sample of 90 offices and calculates the cost, £x per square foot. His results are given in the table below.

Cost (£x)	Frequency (f)	Midpoint (£y)
$20 \leqslant x < 40$	12	30
$40 \leqslant x < 45$	13	42.5
$45 \leqslant x < 50$	25	47.5
$50 \leqslant x < 60$	32	55
$60 \leqslant x < 80$	8	70

(You may use $\sum f y^2 = 226\,687.5$)

A histogram is drawn for these data and the bar representing $50 \leqslant x < 60$ is 2 cm wide and 8 cm high.

(a) Calculate the width and height of the bar representing $20 \leqslant x < 40$

(3)

(b) Use linear interpolation to estimate the median cost.

(2)

(c) Estimate the mean cost of office space for these data.

(2)

(d) Estimate the standard deviation for these data.

(2)

Leave blank

Question 28 continued

Leave blank

Question 28 continued

Leave blank

Question 28 continued

(Total 9 marks)

Leave blank

29. A midwife records the weights, in kg, of a sample of 50 babies born at a hospital. Her results are given in the table below.

Weight (w kg)	Frequency (f)	Weight midpoint (x)
$0 \leqslant w < 2$	1	1
$2 \leqslant w < 3$	8	2.5
$3 \leqslant w < 3.5$	17	3.25
$3.5 \leqslant w < 4$	17	3.75
$4 \leqslant w < 5$	7	4.5

[You may use $\sum fx^2 = 611.375$]

A histogram has been drawn to represent these data.

The bar representing the weight $2 \leqslant w < 3$ has a width of 1 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing a weight of $3 \leqslant w < 3.5$

(3)

(b) Use linear interpolation to estimate the median weight of these babies.

(2)

(c) (i) Show that an estimate of the mean weight of these babies is 3.43 kg.

(ii) Find an estimate of the standard deviation of the weights of these babies.

(3)

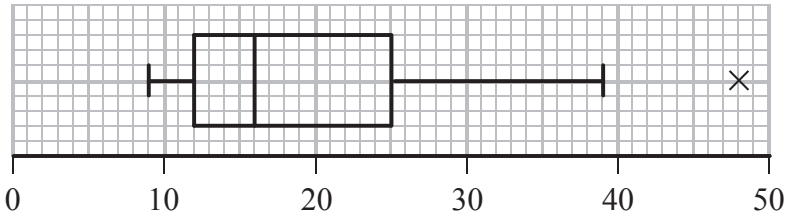
Leave blank

Question 29 continued

(Total 8 marks)

Leave blank

30. Each of 60 students was asked to draw a 20° angle without using a protractor. The size of each angle drawn was measured. The results are summarised in the box plot below.



(a) Find the range for these data.

(1)

(b) Find the interquartile range for these data.

(1)

The students were then asked to draw a 70° angle.
The results are summarised in the table below.

Angle, a, (degrees)	Number of students
$55 \leqslant a < 60$	6
$60 \leqslant a < 65$	15
$65 \leqslant a < 70$	13
$70 \leqslant a < 75$	11
$75 \leqslant a < 80$	8
$80 \leqslant a < 85$	7

(c) Use linear interpolation to estimate the size of the median angle drawn. Give your answer to 1 decimal place.

(2)

(d) Show that the lower quartile is 63°

(2)

For these data, the upper quartile is 75°, the minimum is 55° and the maximum is 84°

An outlier is an observation that falls either
more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

(e) (i) Show that there are no outliers for these data.

(ii) Draw a box plot for these data on the grid on page 3.

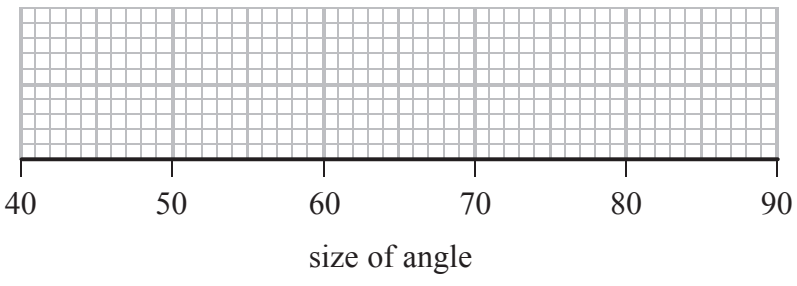
(5)

(f) State which angle the students were more accurate at drawing. Give reasons for your answer.

(3)

Leave blank

Question 30 continued



(Total 14 marks)

Leave blank

31. The mark, x, scored by each student who sat a statistics examination is coded using

$$y = 1.4x - 20$$

The coded marks have mean 60.8 and standard deviation 6.60

Find the mean and the standard deviation of x.

(4)

(Total 4 marks)

Leave blank

32. The times, in seconds, spent in a queue at a supermarket by 85 randomly selected customers, are summarised in the table below.

Time (seconds)	Number of customers, f
0 − 30	2
30 − 60	10
60 − 70	17
70 − 80	25
80 − 100	25
100 − 150	6

A histogram was drawn to represent these data. The 30 − 60 group was represented by a bar of width 1.5 cm and height 1 cm.

(a) Find the width and the height of the 70 − 80 group.

(3)

(b) Use linear interpolation to estimate the median of this distribution.

(2)

Given that x denotes the midpoint of each group in the table and

$$\sum fx = 6460 \qquad \sum fx^2 = 529\,400$$

(c) calculate an estimate for

(i) the mean,

(ii) the standard deviation,

for the above data.

(3)

Leave blank

Question 32 continued

(Total 8 marks)

Leave blank

33. The table shows the time, to the nearest minute, spent waiting for a taxi by each of 80 people one Sunday afternoon.

Waiting time (in minutes)	Frequency
2–4	15
5–6	9
7	6
8	24
9–10	14
11–15	12

(a) Write down the upper class boundary for the 2–4 minute interval.

(1)

A histogram is drawn to represent these data. The height of the tallest bar is 6 cm.

(b) Calculate the height of the second tallest bar.

(3)

(c) Estimate the number of people with a waiting time between 3.5 minutes and 7 minutes.

(2)

(d) Use linear interpolation to estimate the median, the lower quartile and the upper quartile of the waiting times.

(4)

Leave blank

Question 33 continued

(Total 10 marks)

Leave blank

34. The following table summarises the times, t minutes to the nearest minute, recorded for a group of students to complete an exam.

Time (minutes) t	11 – 20	21 – 25	26 – 30	31 – 35	36 – 45	46 – 60
Number of students f	62	88	16	13	11	10

[You may use $\sum ft^2 = 134281.25$]

(a) Estimate the mean and standard deviation of these data.

(5)

(b) Use linear interpolation to estimate the value of the median.

(2)

(c) Show that the estimated value of the lower quartile is 18.6 to 3 significant figures.

(1)

(d) Estimate the interquartile range of this distribution.

(2)

(e) Give a reason why the mean and standard deviation are not the most appropriate summary statistics to use with these data.

(1)

The person timing the exam made an error and each student actually took 5 minutes less than the times recorded above. The table below summarises the actual times.

Time (minutes) t	6 – 15	16 – 20	21 – 25	26 – 30	31 – 40	41 – 55
Number of students f	62	88	16	13	11	10

(f) Without further calculations, explain the effect this would have on each of the estimates found in parts (a), (b), (c) and (d).

(3)

Leave blank

Question 34 continued

(Total 14 marks)

Leave blank

35. An agriculturalist is studying the yields, y kg, from tomato plants. The data from a random sample of 70 tomato plants are summarised below.

Yield (y kg)	**Frequency** (f)	**Yield midpoint** (x kg)
$0 \leqslant y < 5$	16	2.5
$5 \leqslant y < 10$	24	7.5
$10 \leqslant y < 15$	14	12.5
$15 \leqslant y < 25$	12	20
$25 \leqslant y < 35$	4	30

(You may use $\sum fx = 755$ and $\sum fx^2 = 12037.5$)

A histogram has been drawn to represent these data.

The bar representing the yield $5 \leqslant y < 10$ has a width of 1.5 cm and a height of 8 cm.

(a) Calculate the width and the height of the bar representing the yield $15 \leqslant y < 25$

(3)

(b) Use linear interpolation to estimate the median yield of the tomato plants.

(2)

(c) Estimate the mean and the standard deviation of the yields of the tomato plants.

(4)

Leave blank

Question 35 continued

(Total 9 marks)

Leave blank

36. A survey of 100 households gave the following results for weekly income £y.

Income y (£)	Mid-point	Frequency f
$0 \leqslant y < 200$	100	12
$200 \leqslant y < 240$	220	28
$240 \leqslant y < 320$	280	22
$320 \leqslant y < 400$	360	18
$400 \leqslant y < 600$	500	12
$600 \leqslant y < 800$	700	8

(You may use $\sum fy^2 = 12\ 452\ 800$)

A histogram was drawn and the class $200 \leqslant y < 240$ was represented by a rectangle of width 2 cm and height 7 cm.

(a) Calculate the width and the height of the rectangle representing the class $320 \leqslant y < 400$

(3)

(b) Use linear interpolation to estimate the median weekly income to the nearest pound.

(2)

(c) Estimate the mean and the standard deviation of the weekly income for these data.

(4)

Leave blank

Question 36 continued

(Total 9 marks)

Leave blank

37. A scientist is researching whether or not birds of prey exposed to pollutants lay eggs with thinner shells. He collects a random sample of egg shells from each of 6 different nests and tests for pollutant level, p, and measures the thinning of the shell, t. The results are shown in the table below.

p	3	8	30	25	15	12
t	1	3	9	10	5	6

[You may use $\sum p^2 = 1967$ and $\sum pt = 694$]

(a) Draw a scatter diagram on the axes on page 7 to represent these data.

(2)

(b) Explain why a linear regression model may be appropriate to describe the relationship between p and t.

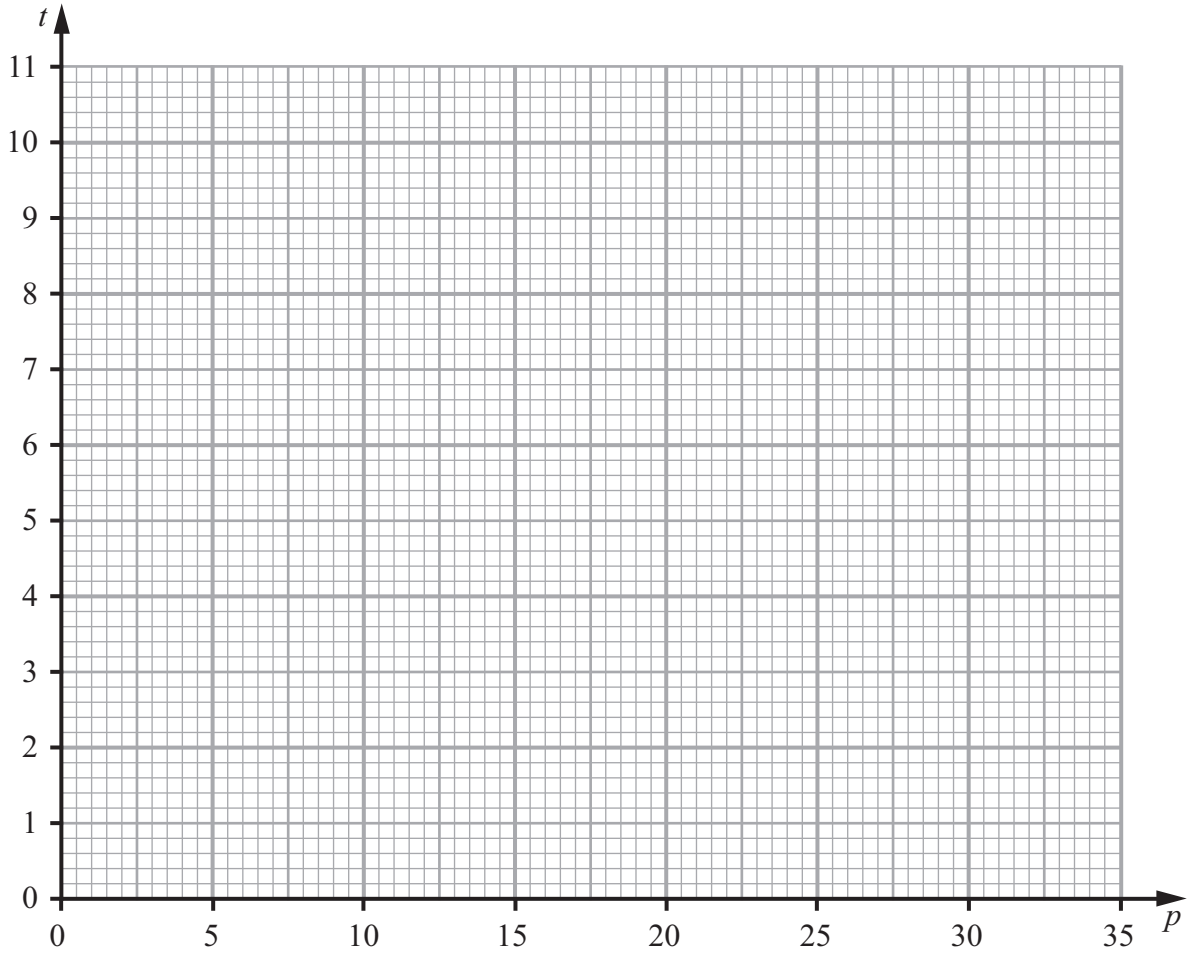
(1)

(c) Calculate the value of S_{pt} and the value of S_{pp}.

(4)

Leave blank

Question 37 continued



(Total 7 marks)

Leave blank

38.

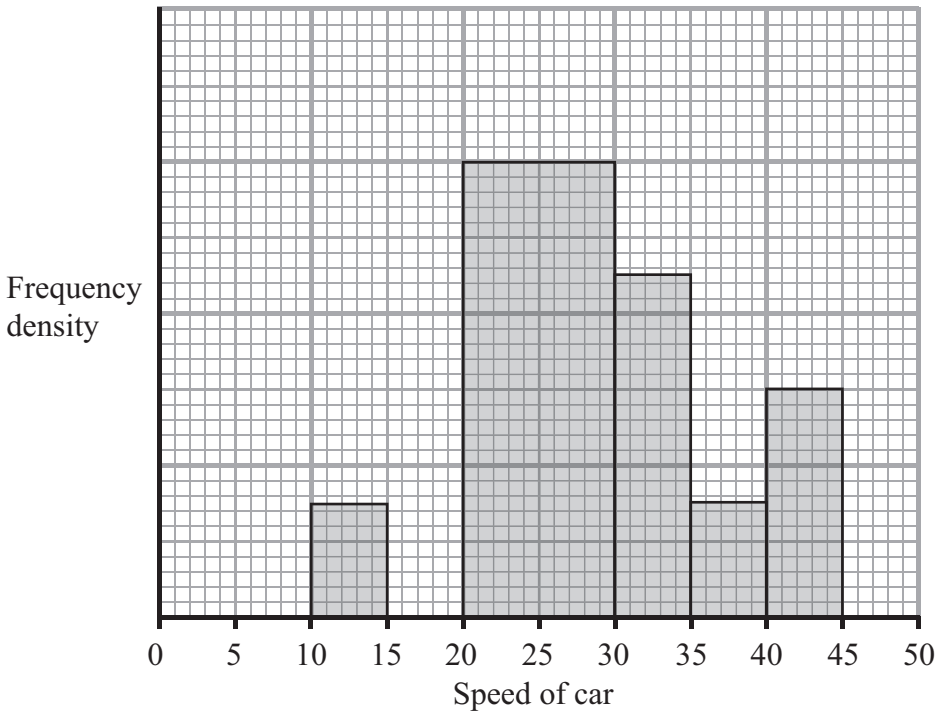


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample.

(4)

(b) Estimate the value of the mean speed of the cars in the sample.

(3)

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample.

(2)

(d) Comment on the shape of the distribution. Give a reason for your answer.

(2)

(e) State, with a reason, whether the estimate of the mean or the median is a better representation of the average speed of the traffic on the road.

(2)

Leave blank

Question 38 continued

(Total 13 marks)

Leave blank

39. The histogram in Figure 1 shows the time, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway.

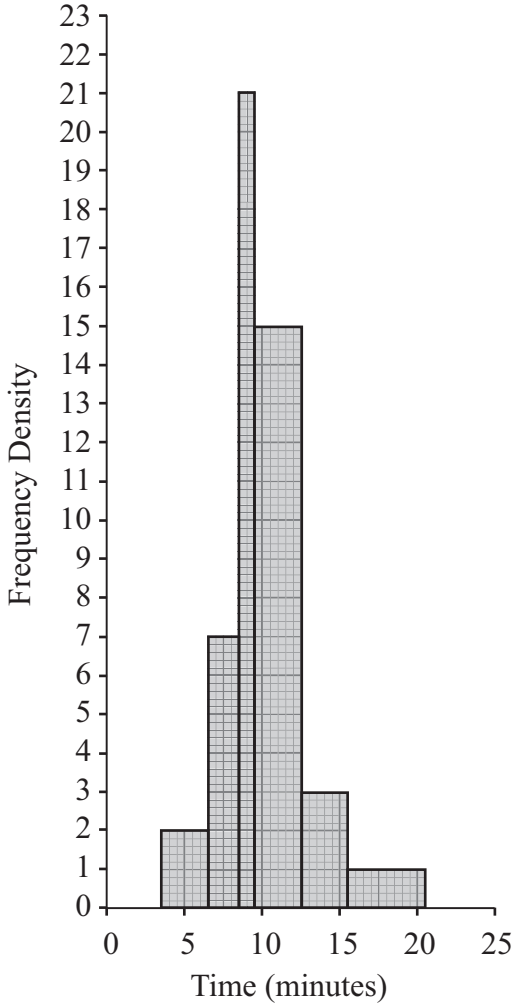


Figure 1

(a) Complete the table.

Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	21
10 – 12	45
13 – 15	9
16 – 20	

(2)

(b) Estimate the number of motorists who were delayed between 8.5 and 13.5 minutes by the roadworks.

(2)

Leave blank

Question 39 continued

(Total 4 marks)

Leave blank

40. A class of students had a sudoku competition. The time taken for each student to complete the sudoku was recorded to the nearest minute and the results are summarised in the table below.

Time	Mid-point, x	Frequency, f
2 - 8	5	2
9 - 12		7
13 - 15	14	5
16 - 18	17	8
19 - 22	20.5	4
23 - 30	26.5	4

(You may use $\sum fx^2 = 8603.75$)

(a) Write down the mid-point for the 9 - 12 interval.

(1)

(b) Use linear interpolation to estimate the median time taken by the students.

(2)

(c) Estimate the mean and standard deviation of the times taken by the students.

(5)

Leave blank

Question 40 continued

(Total 8 marks)

Leave blank

41. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1 – 10	11 – 20	21 – 25	26 – 30	31 – 40	41 – 59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the $21 - 25$ hour and $31 - 40$ hour groups.

(2)

A histogram was drawn to represent these data. The $11 - 20$ group was represented by a bar of width $4\,\text{cm}$ and height $6\,\text{cm}$.

(b) Find the width and height of the $26 - 30$ group.

(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by these students.

(2)

Leave blank

Question 41 continued

(Total 12 marks)

Leave blank

42. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, x kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum fx = 4841$ and $\sum fx^2 = 15\,889.5$]

(a) Write down the missing midpoints in the table above.

(2)

(b) Calculate an estimate of the mean birth weight.

(2)

(c) Calculate an estimate of the standard deviation of the birth weight.

(3)

(d) Use interpolation to estimate the median birth weight.

(2)

Leave blank

Question 42 continued

(Total 9 marks)

Leave blank

43. The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

(a) the width,

(1)

(b) the height

(2)

of the bar representing this class.

(Total 3 marks)

Leave blank

44. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

Foot length, *l*, (cm)	**Number of children**
$10 \leqslant l < 12$	5
$12 \leqslant l < 17$	53
$17 \leqslant l < 19$	29
$19 \leqslant l < 21$	15
$21 \leqslant l < 23$	11
$23 \leqslant l < 25$	7

(a) Use interpolation to estimate the median of this distribution.

(2)

(b) Calculate estimates for the mean and the standard deviation of these data.

(6)

Leave blank

Question 44 continued

(Total 8 marks)

Leave blank

45. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

(a) Find the median and quartiles for these data.

(3)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(b) Show that 110 is the only outlier.

(2)

(c) Using the graph paper on page 15 draw a box plot for these data indicating clearly the position of the outlier.

(3)

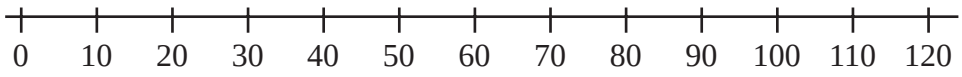
The value of 110 is omitted.

(d) Show that S_{yy} for the remaining 10 students is 2966.9

(3)

Leave blank

Question 45 continued



y minutes

(Total 11 marks)

Leave blank

46. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(a) Calculate the width and height of the rectangle representing the group (16 – 25) hours.

(3)

(b) Use linear interpolation to estimate the median and interquartile range.

(5)

(c) Estimate the mean and standard deviation of the number of hours spent shopping.

(4)

Leave blank

Question 46 continued

(Total 12 marks)

Leave blank

47. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724\ 961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.

(4)

(b) Find the median, upper and lower quartiles of these data.

(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3+1.5(Q_3-Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.

(4)

Leave blank

Question 47 continued

(Total 11 marks)

Leave blank

48. The histogram in Figure 1 shows the time taken, to the nearest minute, for 140 runners to complete a fun run.

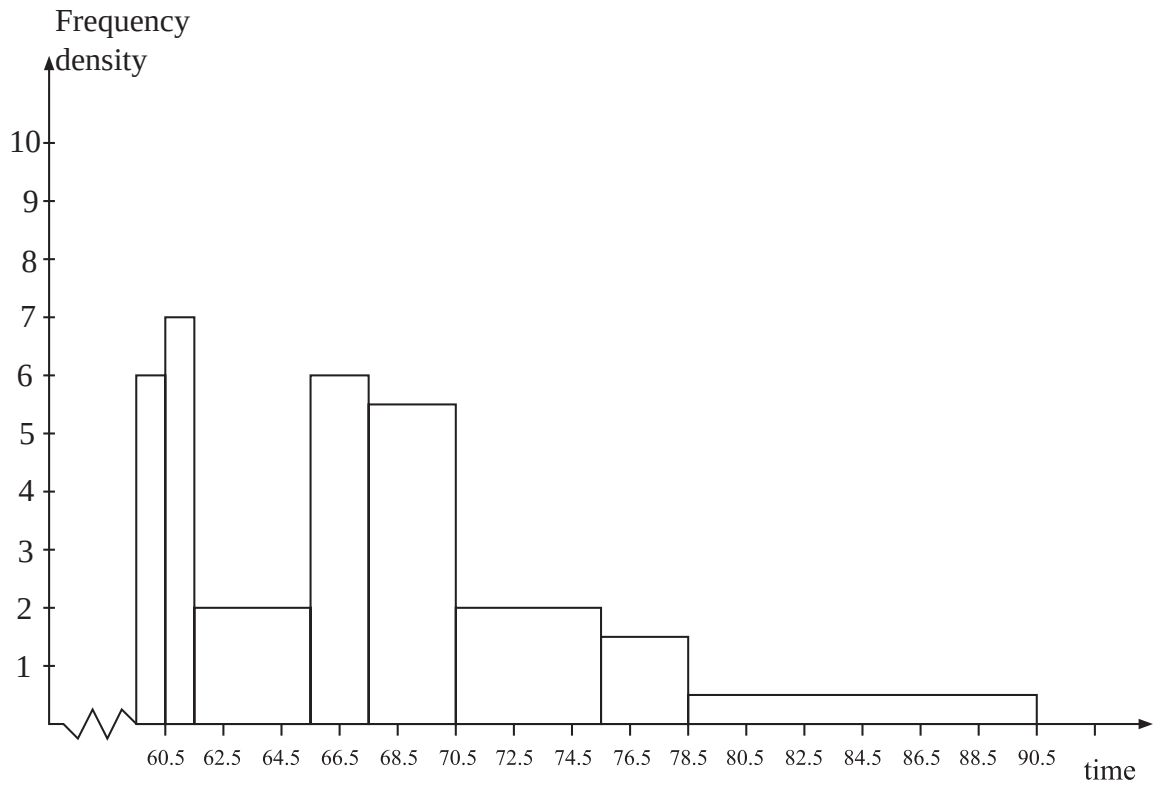


Figure 1

Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(5)

(Total 5 marks)