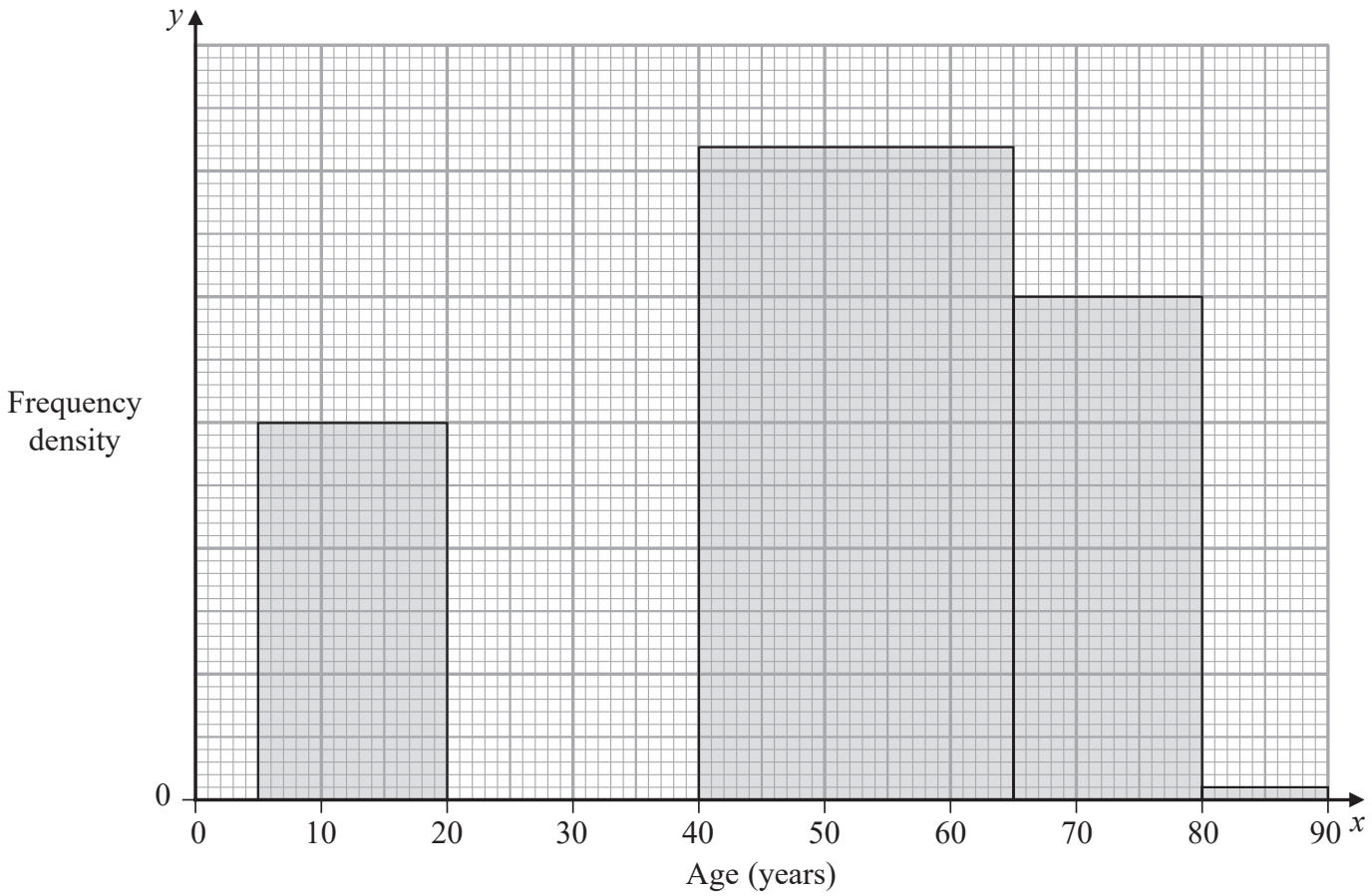
2. The partially completed table and partially completed histogram give information about the ages of passengers on an airline.

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)

4. 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)

4. 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)

1.

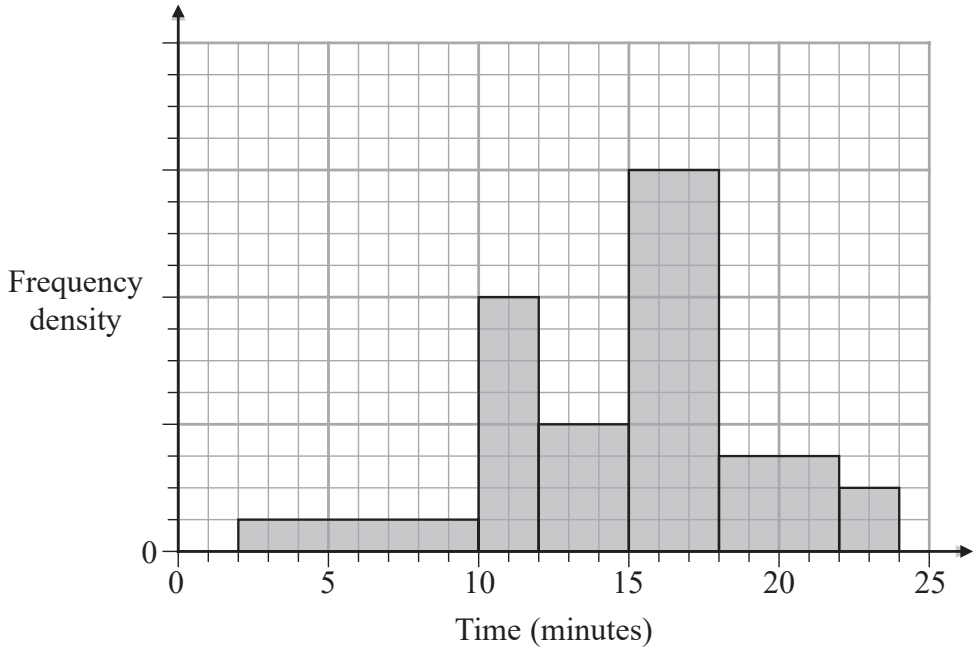


Figure 1

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)

4. 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.

(a) Give a reason why stratified random sampling cannot be used.

(1)

(b) Explain how a sample of size 160 could be taken to ensure that the estimated populations of each type of carp are fairly represented.

You should state the name of the sampling method used.

(2)

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 $\sum fm^2 = 3053$)

(c) 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

(d) 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)

SECTION A: STATISTICS

Answer ALL questions. Write your answers in the spaces provided.

1. 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.

(a) State the sampling technique Sara used.

(1)

(b) From your knowledge of the large data set explain why this process may not generate a sample of size 20.

(1)

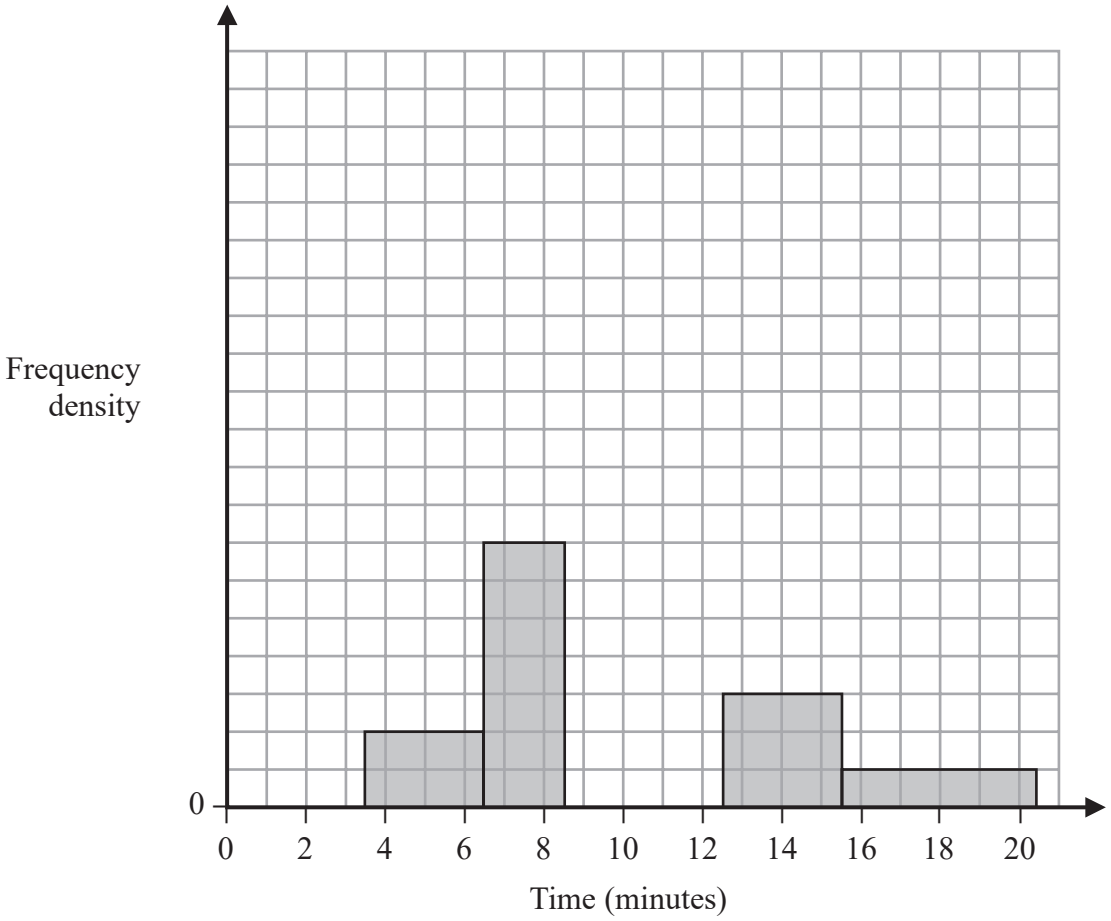
The data Sara collected are summarised as follows

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

(c) Calculate the standard deviation.

(2)

2. 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.



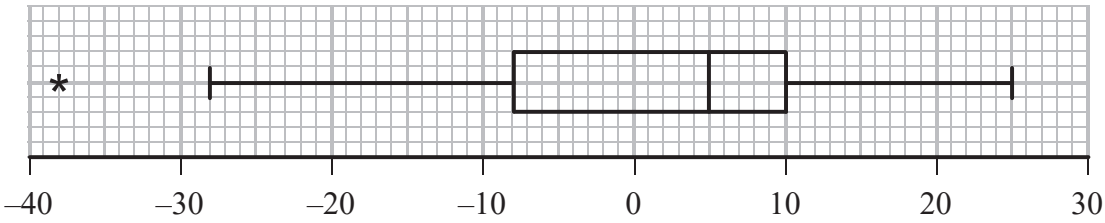
Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	17
10 – 12	45
13 – 15	9
16 – 20	

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

(5)

Leave blank

1. At the start of a course, an instructor asked a group of 80 apprentices to estimate the length of a piece of pipe. The error (true length – estimated length) was recorded in centimetres. 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)

One month later, the instructor asked the 80 apprentices to estimate the length of a different piece of pipe and recorded their errors. The results are summarised in the table below.

Error (e cm)	Number of apprentices
$-40 < e \leqslant -16$	2
$-16 < e \leqslant -8$	18
$-8 < e \leqslant 0$	33
$0 < e \leqslant 8$	14
$8 < e \leqslant 16$	10
$16 < e \leqslant 40$	3

(c) Use linear interpolation to estimate the median error for these data.

(2)

(d) Show that the upper quartile for these data, to the nearest centimetre, is 4.

(1)

For these data, the lower quartile is –8 and the five worst errors were –25, –21, 18, 23, 28

An outlier is a value 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 only 2 outliers for these data.

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

(6)

(f) State, giving reasons, whether or not the apprentices' ability to estimate the length of a piece of pipe has improved over the first month of the course.

(3)

Leave blank

1. Nina weighed a random sample of 50 carrots from her shop and recorded the weight, in grams to the nearest gram, for each carrot. The results are summarised below.

Weight of carrot	Frequency (f)	Weight midpoint (x grams)
45–54	5	49.5
55–59	10	57
60–64	22	62
65–74	13	69.5

(You may use $\sum fx^2 = 192\,102.5$)

(a) Use linear interpolation to estimate the median weight of these carrots.

(2)

(b) Find an estimate for the mean weight of these carrots.

(2)

(c) Find an estimate for the standard deviation of the weights of these carrots.

(2)

A carrot is selected at random from Nina's shop.

(d) Estimate the probability that the weight of this carrot is more than 70 grams.

(2)

Leave blank

2. The time taken to complete a puzzle, in minutes, is recorded for each person in a club. The times are summarised in a grouped frequency distribution and represented by a histogram.

One of the class intervals has a frequency of 20 and is shown by a bar of width 1.5 cm and height 12 cm on the histogram. The total area under the histogram is 94.5 cm^2

Find the number of people in the club.

(3)

Leave blank

8.

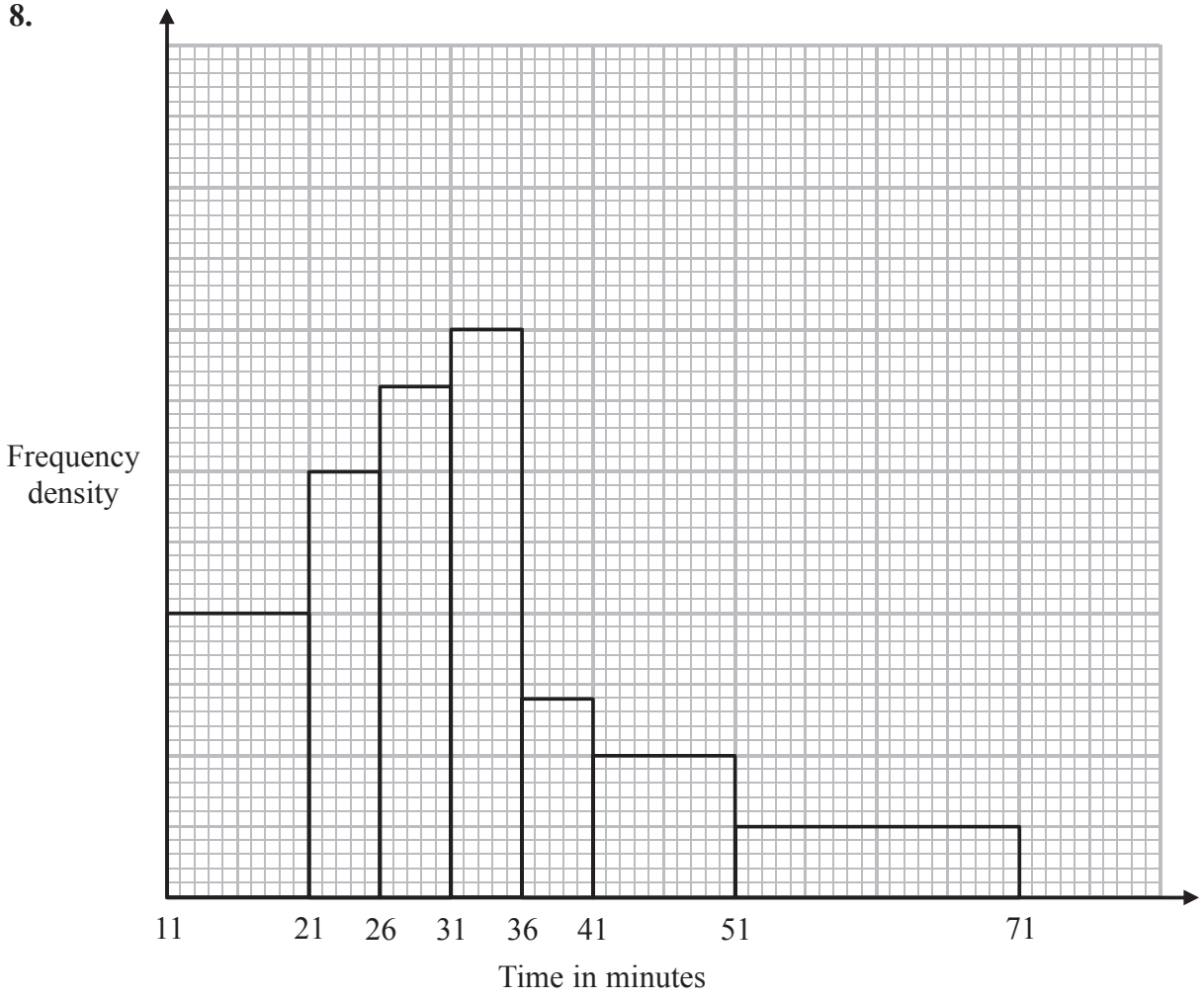


Figure 1

The histogram in Figure 1 summarises the times, in minutes, that 200 people spent shopping in a supermarket.

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

(1)

Given that 40 people spent between 11 and 21 minutes shopping in the supermarket, estimate

(b) the number of people that spent between 18 and 25 minutes shopping in the supermarket,

(3)

(c) the median time spent shopping in the supermarket by these 200 people.

(2)

The mid-point of each bar is represented by x and the corresponding frequency by f.

(d) Show that $\sum fx = 6390$

(2)

Leave blank

2. The table below shows the distances (to the nearest km) travelled to work by the 50 employees in an office.

Distance (km)	Frequency (f)	Distance midpoint (x)
0 – 2	16	1.25
3 – 5	12	4
6 – 10	10	8
11 – 20	8	15.5
21 – 40	4	30.5

[You may use $\sum fx = 394$, $\sum fx^2 = 6500$]

A histogram has been drawn to represent these data.

The bar representing the distance of 3 – 5 has a width of 1.5 cm and a height of 6 cm.

(a) Calculate the width and height of the bar representing the distance of 6 – 10

(3)

(b) Use linear interpolation to estimate the median distance travelled to work.

(2)

(c) (i) Show that an estimate of the mean distance travelled to work is 7.88 km.

(ii) Estimate the standard deviation of the distances travelled to work.

(4)

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

(2)

Peng starts to work in this office as the 51st employee.

She travels a distance of 7.88 km to work.

(e) Without carrying out any further calculations, state, giving a reason, what effect Peng's addition to the workforce would have on your estimates of the

(i) mean,

(ii) median,

(iii) standard deviation

of the distances travelled to work.

(3)

Leave blank

1. Two classes of students, class A and class B, sat a test.

Class A has 10 students. Class B has 15 students.

Each student achieved a score, x, on the test and their scores are summarised in the table below.

	n	$\sum x$	$\sum x^2$
Class A	10	770	59610
Class B	15	t	58035

The mean score for Class A is 77 and the mean score for Class B is 61

(a) Find the value of t

(1)

(b) Calculate the variance of the test scores for each class.

(3)

The highest score on the test was 95 and the lowest score was 45

These were each scored by students from the same class.

(c) State, with a reason, which class you believe they were from.

(1)

The two classes are combined into one group of 25 students.

(d) (i) Find the mean test score for all 25 students.

(ii) Find the variance of the test scores for all 25 students.

(4)

The teacher of class A later realises that he added up the test scores for his class incorrectly. Each student's test score in class A should be increased by 3

(e) Without further calculations, state, with a reason, the effect this will have on

(i) the variance of the test scores for class A

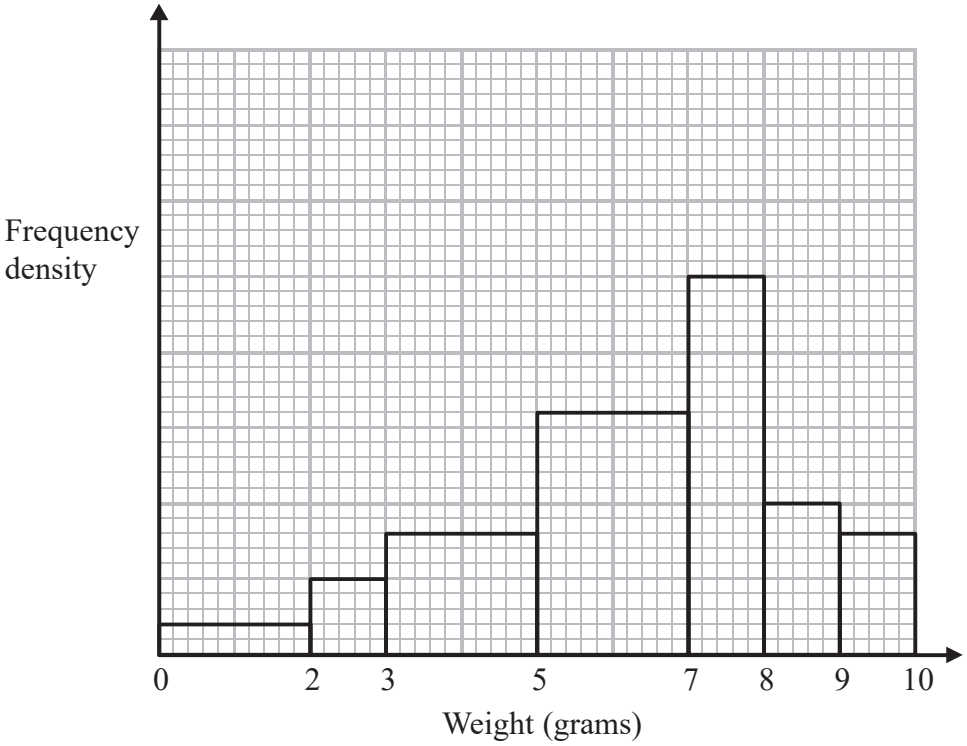
(ii) the mean test score for all 25 students

(iii) the variance of the test scores for all 25 students.

(3)

Leave blank

1. Ralph records the weights, in grams, of 100 tomatoes. This information is displayed in the histogram below.



Given that 5 of the tomatoes have a weight between 2 and 3 grams,

(a) find the number of tomatoes with a weight between 0 and 2 grams.

(2)

One of the tomatoes is selected at random.

(b) Find the probability that it weighs more than 3 grams.

(2)

(c) Estimate the proportion of the tomatoes with a weight greater than 6.25 grams.

(2)

(d) Using your answer to part (c), explain whether or not the median is greater than 6.25 grams.

(1)

Given that the mean weight of these tomatoes is 6.25 grams and using your answer to part (d),

(e) describe the skewness of the distribution of the weights of these tomatoes. Give a reason for your answer.

(1)

Two of these 100 tomatoes are selected at random.

(f) Estimate the probability that both tomatoes weigh within 0.75 grams of the mean.

(4)

Leave blank

2.

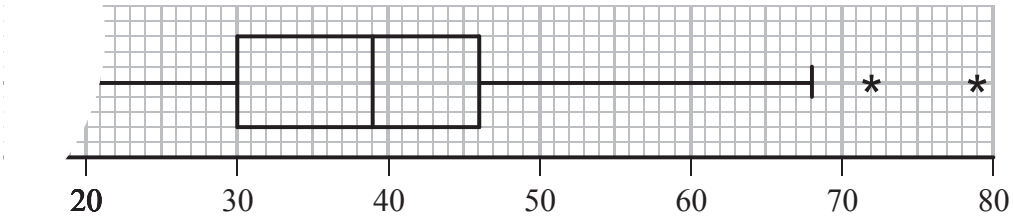


Figure 1

Figure 1 shows part of a box and whisker plot for the marks in an examination with a large number of candidates. Part of the lower whisker has been torn off.

(a) Given that 75% of the candidates passed the examination, state the lowest mark for the award of a pass.

(1)

(b) Given that the top 25% of the candidates achieved a merit grade, state the lowest mark for the award of a merit grade.

(1)

An outlier is defined as any value greater than c or any value less than d where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(c) Find the value of c and the value of d.

(2)

(d) Write down the 3 highest marks scored in the examination.

(2)

The 3 lowest marks in the examination were 5, 10 and 15

(e) On the diagram on page 7, complete the box and whisker plot.

(3)

Three candidates are selected at random from those who took this examination.

(f) Find the probability that all 3 of these candidates passed the examination but only 2 achieved a merit grade.

(3)

Leave blank

8. A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

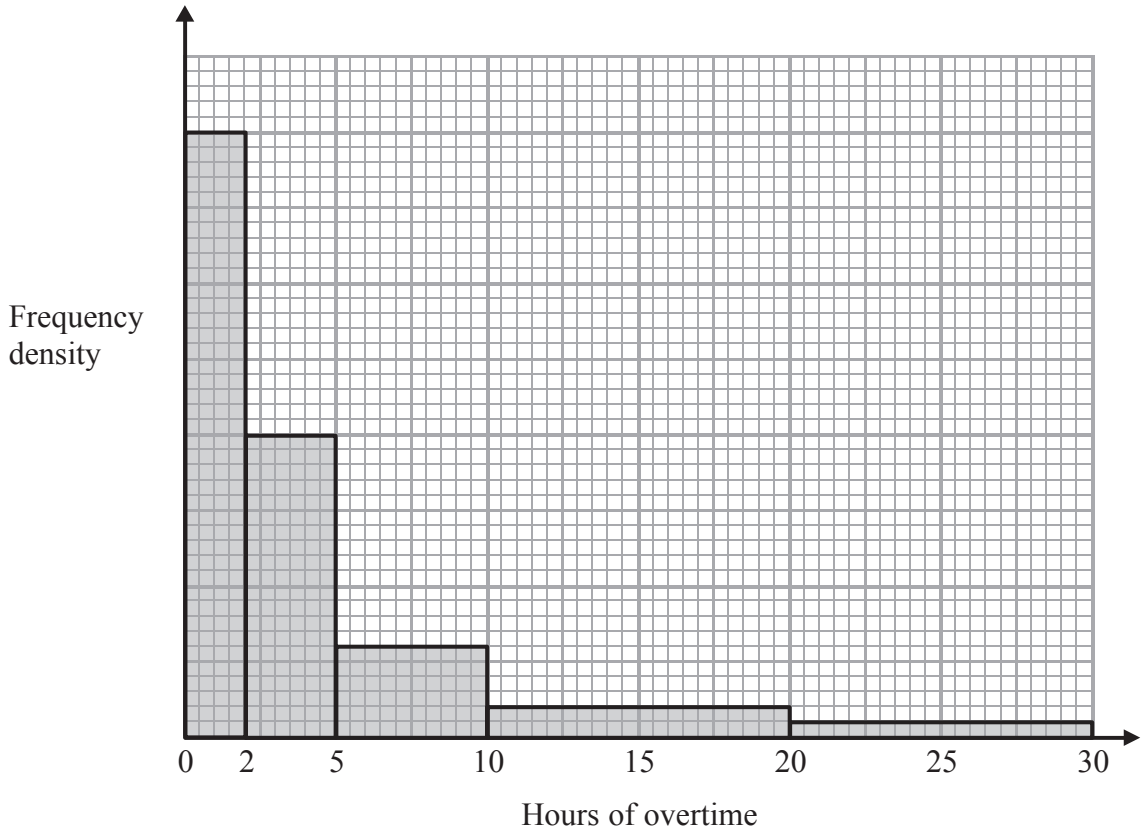


Figure 1

(a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month.

(4)

(b) Estimate the median number of hours of overtime claimed by these 40 staff in the month.

(2)

(c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month.

(2)

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

(d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff.

(2)