Statistical Diagrams and Measures

Q1, (Edexcel 6683, Jan 2010, Q2)

The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2|6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

(a) the median score,

(1)

(b) the interquartile range.

(3)

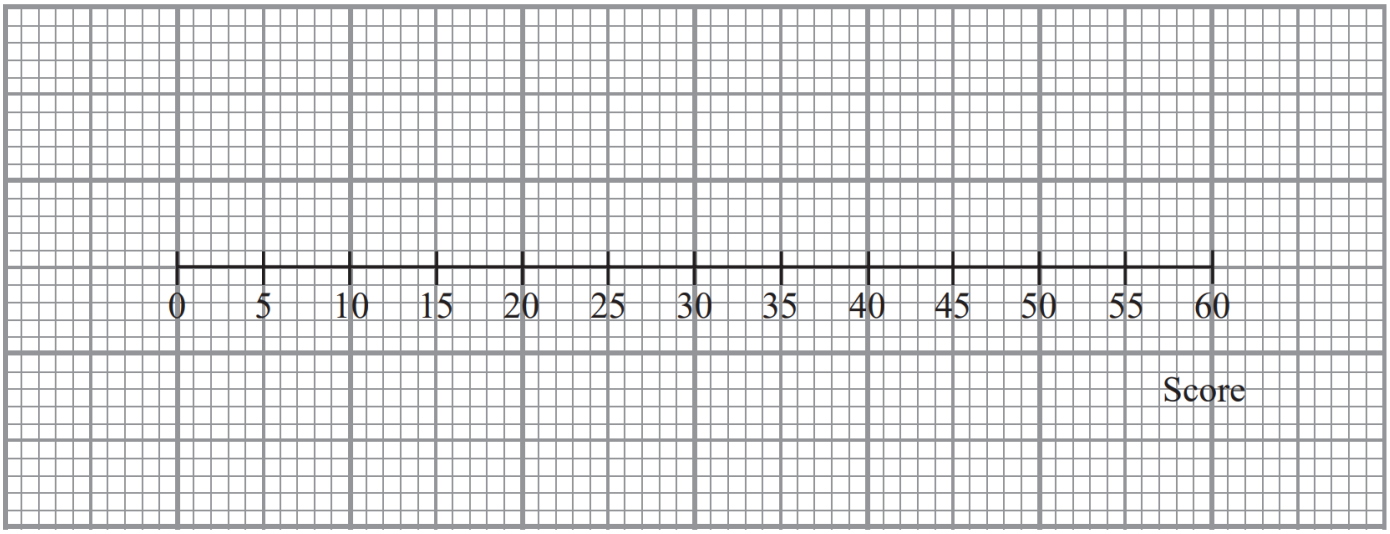
The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

(c) Explain why there is only one employee who will undergo retraining.

(2)

(d) Draw a box plot on the grid below to represent the data. (2)



Q2, (Edexcel 6683, Jun 2007, Q5a-e)

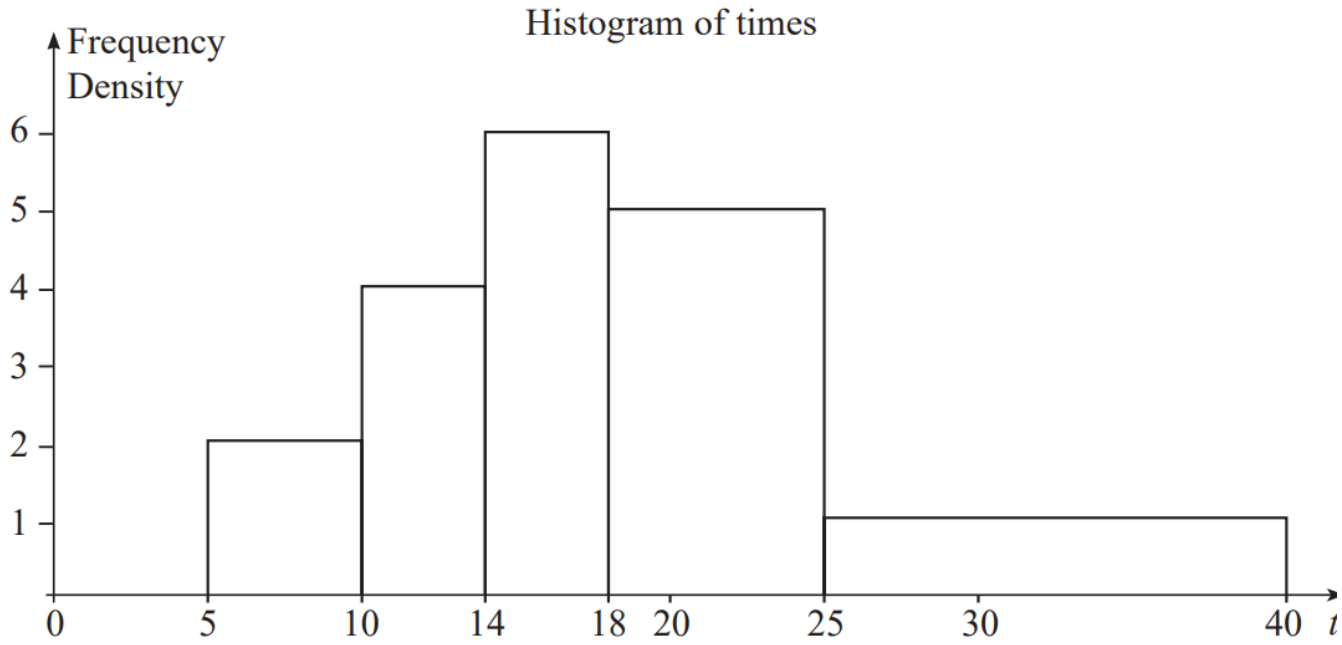


Figure 2

Figure 2 shows a histogram for the variable t which represents the time taken, in minutes, by a group of people to swim 500m.

(a) Complete the frequency table for t.

t	5–10	10–14	14–18	18–25	25–40
Frequency	10	16	24		

(2)

(b) Estimate the number of people who took longer than 20 minutes to swim 500m.

(2)

(c) Find an estimate of the mean time taken.

(4)

(d) Find an estimate for the standard deviation of t.

(3)

(e) Find the median and quartiles for t.

(4)

Q3, (Edexcel 6683, Jun 2012 Q5)

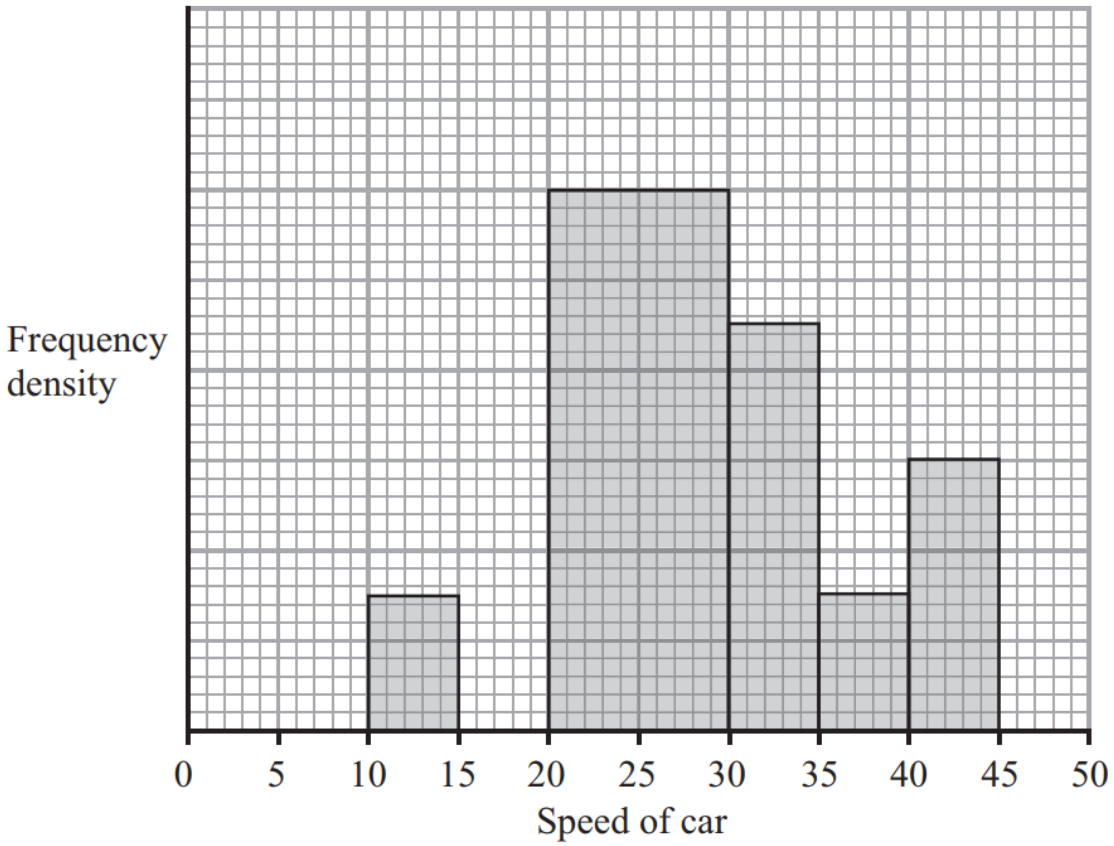


Figure 2

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.

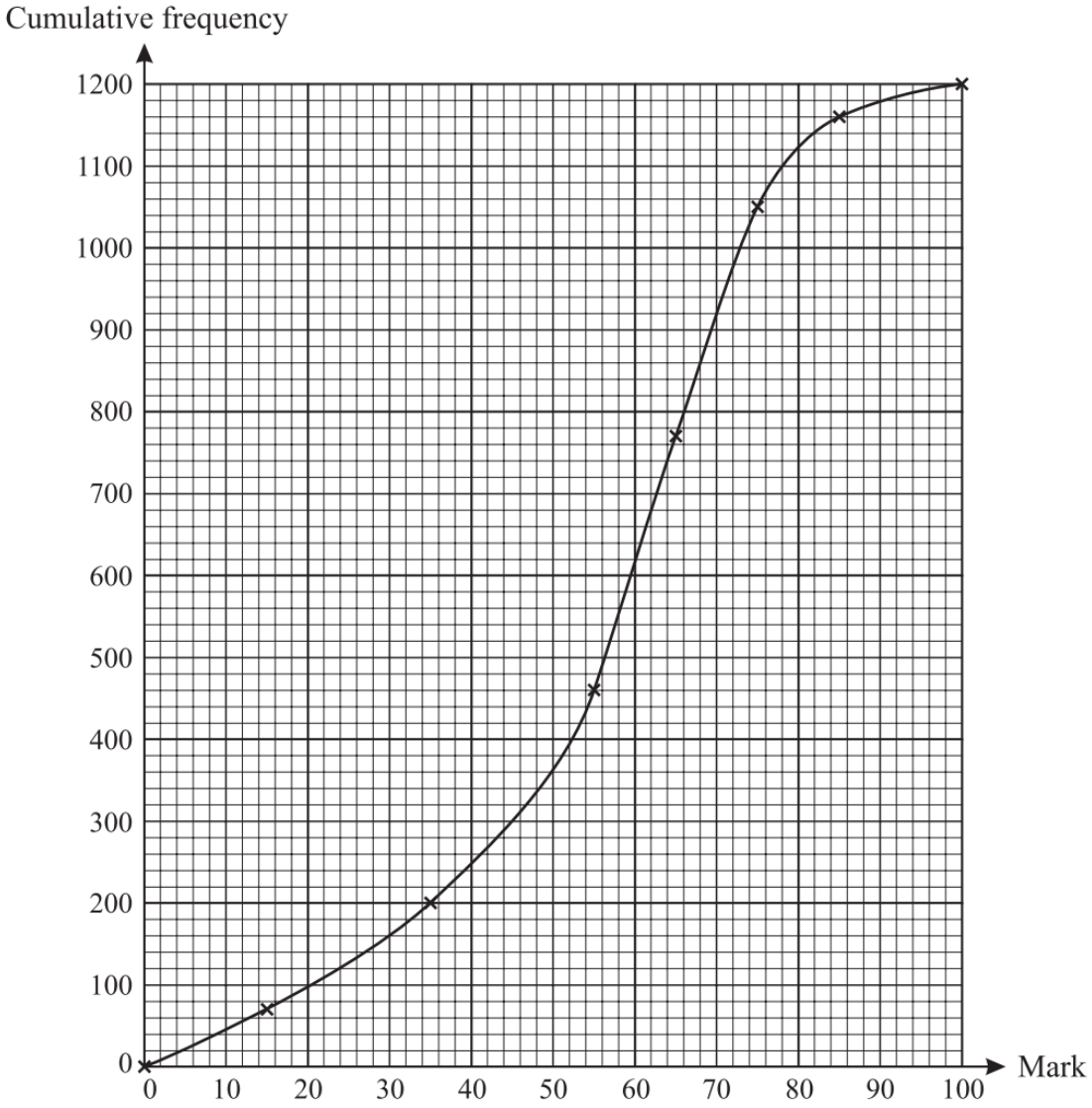
(2)

(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)

Q4, (OCR 4732, Jun 2005, Q5)

The examination marks obtained by 1200 candidates are illustrated on the cumulative frequency graph, where the data points are joined by a smooth curve.



Use the curve to estimate

(i) the interquartile range of the marks, [3]

(ii) x, if 40% of the candidates scored more than x marks, [3]

(iii) the number of candidates who scored more than 68 marks. [2]

Five of the candidates are selected at random, with replacement.

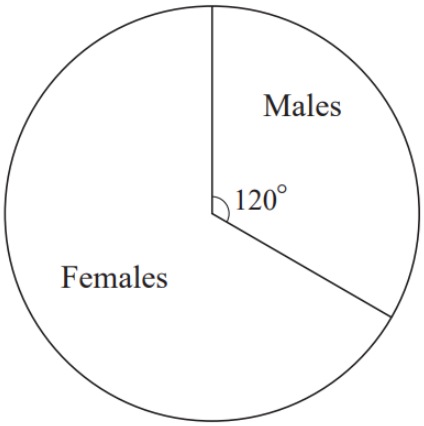
(iv) Estimate the probability that all five scored more than 68 marks. [3]

It is subsequently discovered that the candidates' marks in the range 35 to 55 were evenly distributed — that is, roughly equal numbers of candidates scored 35, 36, 37, …, 55.

(v) What does this information suggest about the estimate of the interquartile range found in part **(i)**? [2]

Q5, (OCR 4732, Jun 2008, Q6)

(i) The numbers of males and females in Year 12 at a school are illustrated in the pie chart. The number of males in Year 12 is 128.

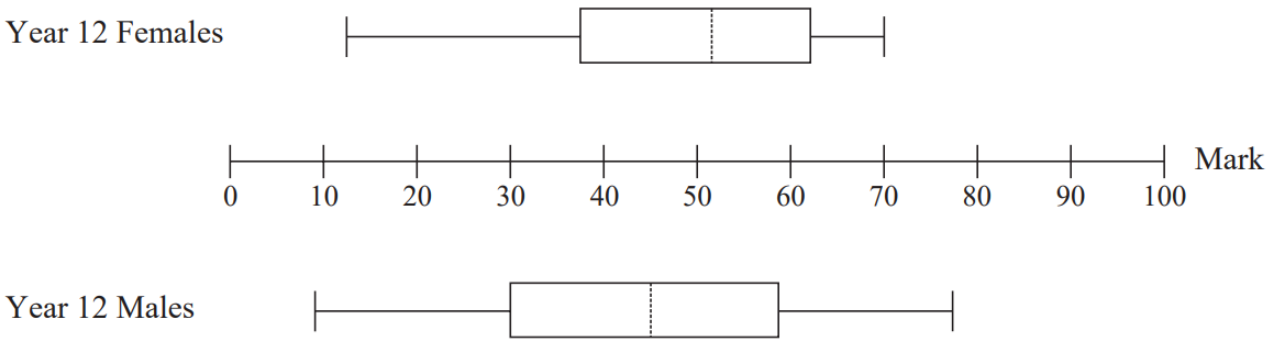


Year 12

(a) Find the number of females in Year 12. [1]

(b) On a corresponding pie chart for Year 13, the angle of the sector representing males is 150°. Explain why this does not necessarily mean that the number of males in Year 13 is more than 128. [1]

(ii) All the Year 12 students took a General Studies examination. The results are illustrated in the box-and-whisker plots.



(a) One student said "The Year 12 pie chart shows that there are more females than males, but the box-and-whisker plots show that there are more males than females."

Comment on this statement. [1]

(b) Give two comparisons between the overall performance of the females and the males in the General Studies examination. [2]

(c) Give one advantage and one disadvantage of using box-and-whisker plots rather than histograms to display the results. [2]

(iii) The mean mark for 102 of the male students was 51. The mean mark for the remaining 26 male students was 59. Calculate the mean mark for all 128 male students. [3]

Q6, (OCE 4732, Jan 2009, Q5i-iii)

The stem-and-leaf diagram shows the masses, in grams, of 23 plums, measured correct to the nearest gram.

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

Key : 6 | 2 means 62

(i) Find the median and interquartile range of these masses. **[3]**

(ii) State one advantage of using the interquartile range rather than the standard deviation as a measure of the variation in these masses. **[1]**

(iii) State one advantage and one disadvantage of using a stem-and-leaf diagram rather than a box-and-whisker plot to represent data. **[2]**

Q7, (OCR 4732, Jun 2014, Q1)

The stem-and-leaf diagram shows the heights, in metres to the nearest 0.1 m, of a random sample of trees of species *A*.

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

Key: 6 | 4 means 6.4 m

(i) Find the median and interquartile range of the heights. **[3]**

(ii) The heights, in metres to the nearest 0.1 m, of a random sample of trees of species *B* are given below.

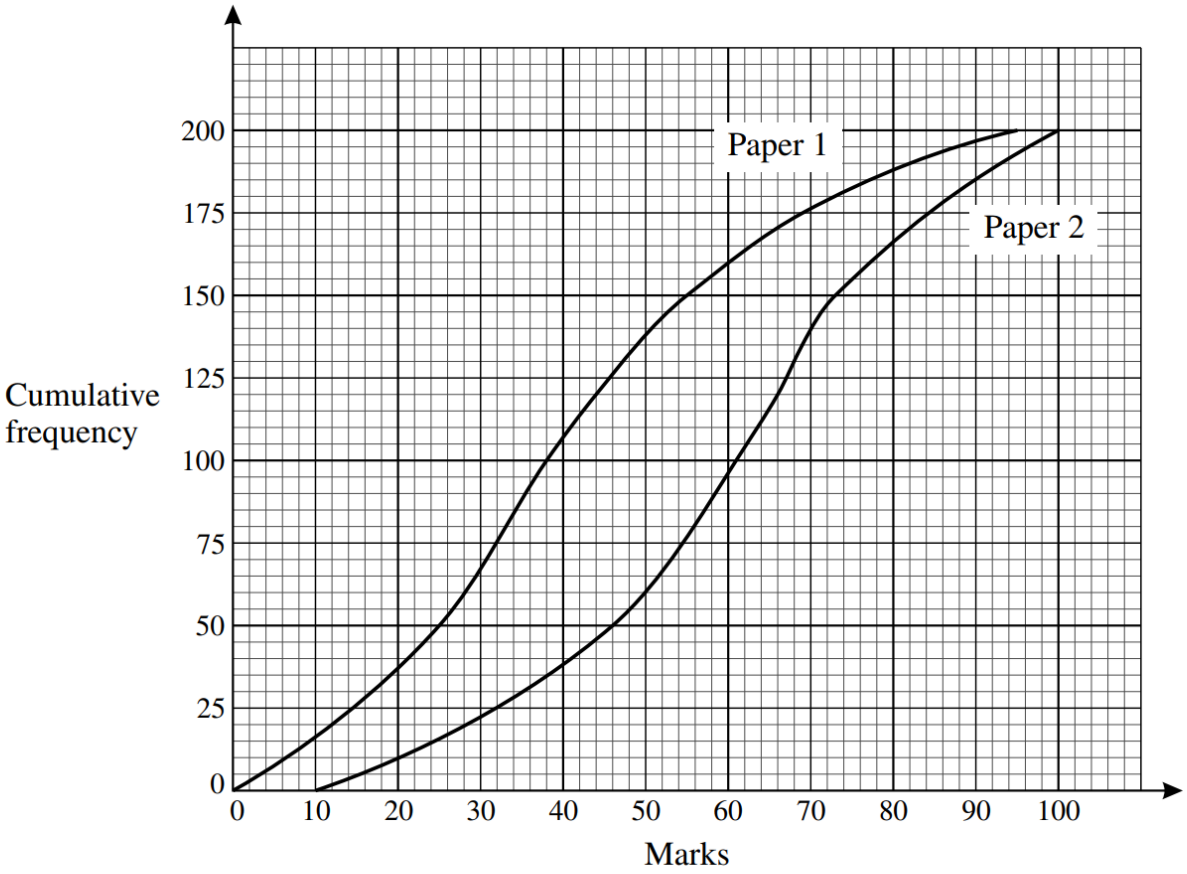
7.6 5.2 8.5 5.2 6.3 6.3 6.8 7.2 6.7 7.3 5.4 7.5 7.4 6.0 6.7

In the answer book, complete the back-to-back stem-and-leaf diagram. **[2]**

(iii) Make two comparisons between the heights of the two species of tree. **[2]**

Q8, (OCR 4732, Jan 2011, Q1)

200 candidates took each of two examination papers. The diagram shows the cumulative frequency graphs for their marks.



(i) Estimate the median mark for each of the papers. **[2]**

(ii) State, with a reason, which of the two papers was the easier one. **[2]**

(iii) It is suggested that the marks on Paper 2 were less varied than those on Paper 1. Use interquartile ranges to comment on this suggestion. **[4]**

(iv) The minimum mark for grade A, the top grade, on Paper 1 was 10 marks lower than the minimum mark for grade A on Paper 2. Given that 25 candidates gained grade A in Paper 1, find the number of candidates who gained grade A in Paper 2. **[2]**

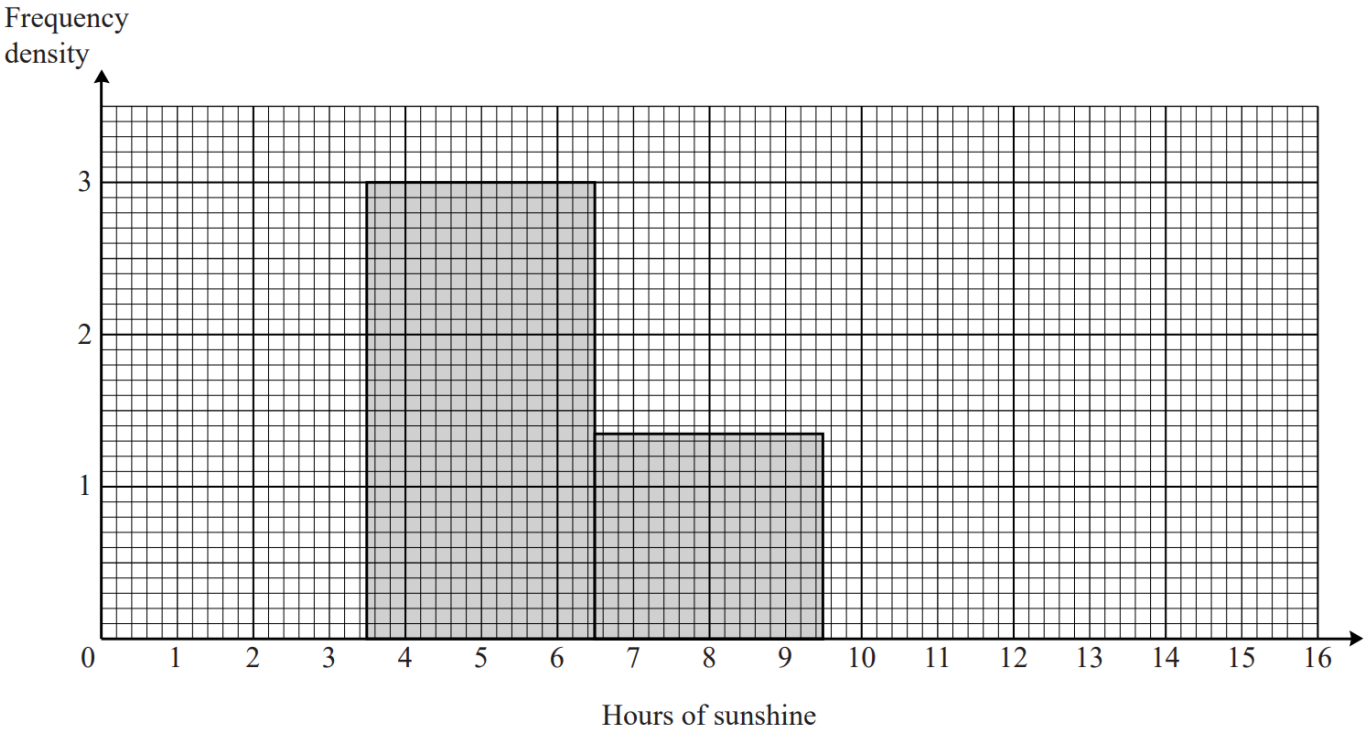
(v) The mean and standard deviation of the marks on Paper 1 were 36.5 and 28.2 respectively. Later, a marking error was discovered and it was decided to add 1 mark to each of the 200 marks on Paper 1. State the mean and standard deviation of the new marks on Paper 1. **[2]**

Q9, (OCR 4732, Jan 2012, Q5)

At a certain resort the number of hours of sunshine, measured to the nearest hour, was recorded on each of 21 days. The results are summarised in the table.

Hours of sunshine	0	1 – 3	4 – 6	7 – 9	10 – 15
Number of days	0	6	9	4	2

The diagram shows part of a histogram to illustrate the data. The scale on the frequency density axis is 2 cm to 1 unit.



(i) (a) Calculate the frequency density of the 1 – 3 class. **[1]**

(b) Fred wishes to draw the block for the 10 – 15 class on the same diagram. Calculate the height, in centimetres, of this block. **[2]**

(ii) A cumulative frequency graph is to be drawn. Write down the coordinates of the first two points that should be plotted. You are not asked to draw the graph. **[2]**

(iii) (a) Calculate estimates of the mean and standard deviation of the number of hours of sunshine. **[5]**

(b) Explain why your answers are only estimates. **[1]**