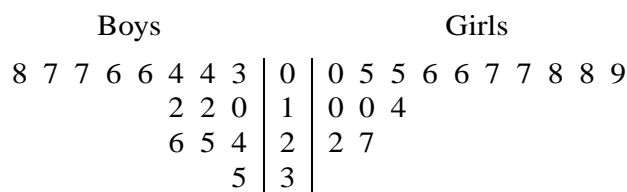


OCR Maths S1

Topic Questions from Papers

Representation of Data

- 1 The back-to-back stem-and-leaf diagram below shows the number of hours of television watched per week by each of 15 boys and 15 girls.

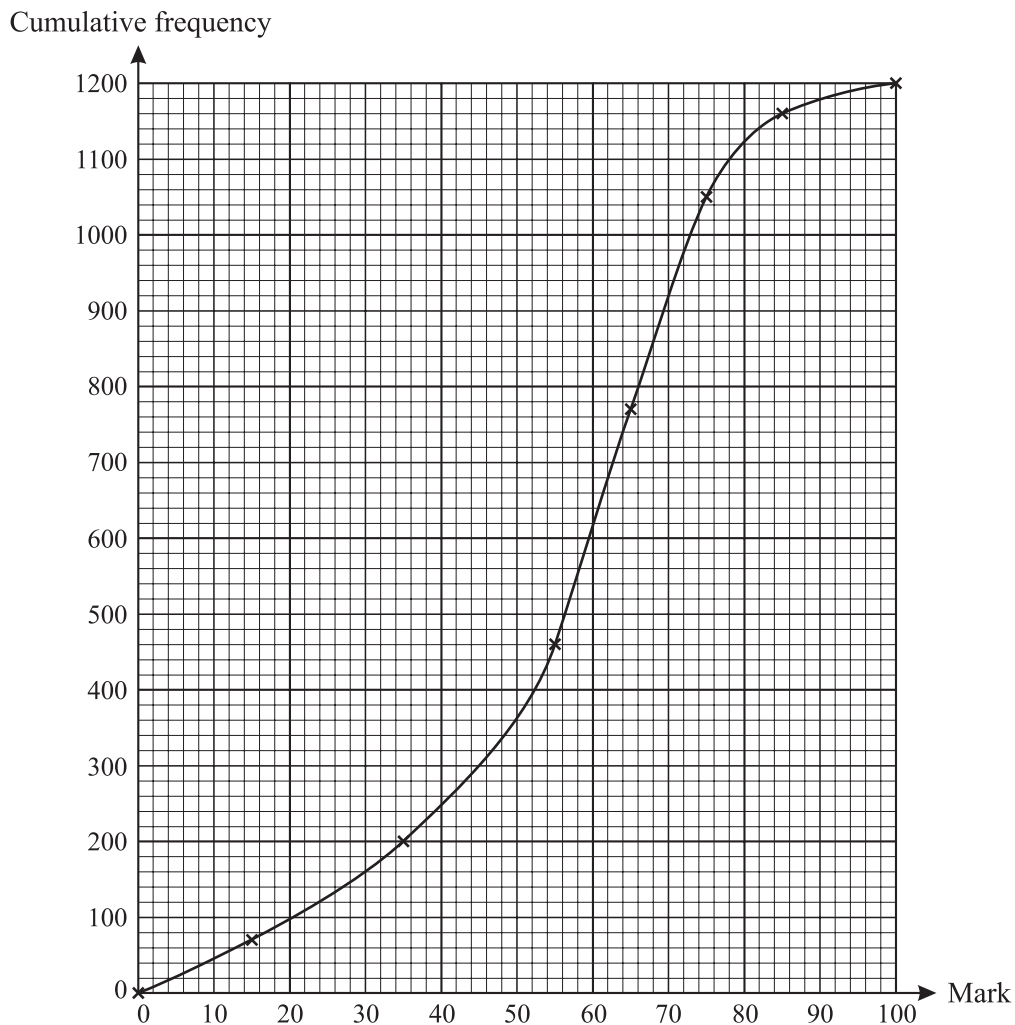


Key: 4 | 2 | 2 means a boy who watched 24 hours and a girl who watched 22 hours of television per week.

- (i) Find the median and the quartiles of the results for the boys. [3]
- (ii) Give a reason why the median might be preferred to the mean in using an average to compare the two data sets. [1]
- (iii) State one advantage, and one disadvantage, of using stem-and-leaf diagrams rather than box-and-whisker plots to represent the data. [2]

(Q2, Jan 2005)

- 2 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, June 2005)

- 3 In a UK government survey in 2000, smokers were asked to estimate the time between their waking and their having the first cigarette of the day. For heavy smokers, the results were as follows.

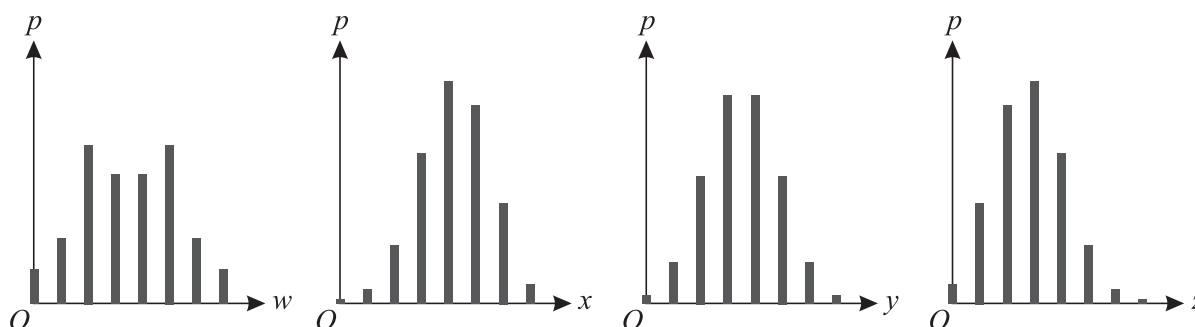
Time between waking and first cigarette	1 to 4 minutes	5 to 14 minutes	15 to 29 minutes	30 to 59 minutes	At least 60 minutes
Percentage of smokers	31	27	19	14	9

Times are given correct to the nearest minute.

- (i) Assuming that 'At least 60 minutes' means 'At least 60 minutes but less than 240 minutes', calculate estimates for the mean and standard deviation of the time between waking and first cigarette for these smokers. [6]
- (ii) Find an estimate for the interquartile range of the time between waking and first cigarette for these smokers. Give your answer correct to the nearest minute. [4]
- (iii) The meaning of 'At least 60 minutes' is now changed to 'At least 60 minutes but less than 480 minutes'. Without further calculation, state whether this would cause an increase, a decrease or no change in the estimated value of
- (a) the mean, [1]
- (b) the standard deviation, [1]
- (c) the interquartile range. [1]

(Q7, June 2006)

- 4 Each of the variables W , X , Y and Z takes eight integer values only. The probability distributions are illustrated in the following diagrams.



- (i) For which one or more of these variables is
- (a) the mean equal to the median, [1]
- (b) the mean greater than the median? [1]
- (ii) Give a reason why **none** of these diagrams could represent a geometric distribution. [1]
- (iii) Which one of these diagrams could **not** represent a binomial distribution? Explain your answer briefly. [2]

(Q4, Jan 2007)

- 5 In the 2001 census, the household size (the number of people living in each household) was recorded. The percentages of households of different sizes were then calculated. The table shows the percentages for two wards, Withington and Old Moat, in Manchester.

	Household size						
	1	2	3	4	5	6	7 or more
Withington	34.1	26.1	12.7	12.8	8.2	4.0	2.1
Old Moat	35.1	27.1	14.7	11.4	7.6	2.8	1.3

- (i) Calculate the median and interquartile range of the household size for Withington. [3]
- (ii) Making an appropriate assumption for the last class, which should be stated, calculate the mean and standard deviation of the household size for Withington. Give your answers to an appropriate degree of accuracy. [6]

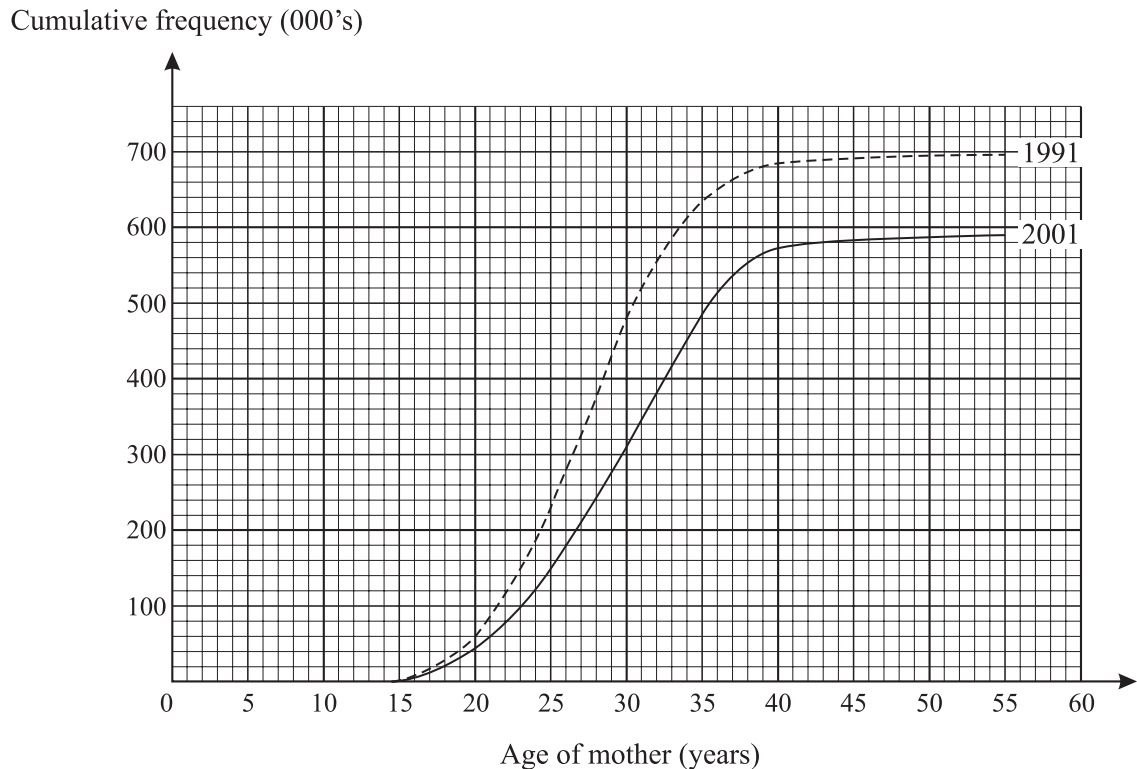
The corresponding results for Old Moat are as follows.

Median	Interquartile range	Mean	Standard deviation
2	2	2.4	1.5

- (iii) State one advantage of using the median rather than the mean as a measure of the average household size. [1]
- (iv) By comparing the values for Withington with those for Old Moat, explain briefly why the interquartile range may be less suitable than the standard deviation as a measure of the variation in household size. [1]
- (v) For one of the above wards, the value of Spearman's rank correlation coefficient between household size and percentage is -1 . Without any calculation, state which ward this is. Explain your answer. [2]

(Q8, Jan 2007)

- 6 The numbers of births, in thousands, to mothers of different ages in England and Wales, in 1991 and 2001 are illustrated by the cumulative frequency curves.



- (i) In which of these two years were there more births? How many more births were there in this year? [2]
- (ii) The following quantities were estimated from the diagram.

Year	Median age (years)	Interquartile range (years)	Proportion of mothers giving birth aged below 25	Proportion of mothers giving birth aged 35 or above
1991	27.5	7.3	33%	9%
2001				18%

- (a) Find the values missing from the table. [5]
- (b) Did the women who gave birth in 2001 tend to be younger or older or about the same age as the women who gave birth in 1991? Using the table and your values from part (a), give two reasons for your answer. [3]

(Q5, June 2007)

- 7 The stem-and-leaf diagram shows the age in completed years of the members of a sports club.

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

Key: 1 | 4 | 0 represents a male aged 41 and a female aged 40.

- (i) Find the median and interquartile range for the males. [3]
- (ii) The median and interquartile range for the females are 27 and 15 respectively. Make two comparisons between the ages of the males and the ages of the females. [2]
- (iii) The mean age of the males is 30.7 and the mean age of the females is 27.5, each correct to 1 decimal place. Give one advantage of using the median rather than the mean to compare the ages of the males with the ages of the females. [1]

A record was kept of the number of hours, X , spent by each member at the club in a year. The results were summarised by

$$n = 49, \quad \Sigma(x - 200) = 245, \quad \Sigma(x - 200)^2 = 9849.$$

- (iv) Calculate the mean and standard deviation of X . [6]

(Q8, Jan 2008)

- 8 The diameters of 100 pebbles were measured. The measurements rounded to the nearest millimetre, x , are summarised in the table.

x	$10 \leq x \leq 19$	$20 \leq x \leq 24$	$25 \leq x \leq 29$	$30 \leq x \leq 49$
Number of stones	25	22	29	24

These data are to be presented on a statistical diagram.

- (i) For a histogram, find the frequency density of the $10 \leq x \leq 19$ class. [2]
- (ii) For a cumulative frequency graph, state the coordinates of the first two points that should be plotted. [2]
- (iii) Why is it not possible to draw an exact box-and-whisker plot to illustrate the data? [1]

(Q5, June 2009)

- 9 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]
- (iv) James wished to calculate the mean and standard deviation of the given data. He first subtracted 5 from each of the digits to the left of the line in the stem-and-leaf diagram, giving the following.

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

Key : 1 | 2 means 12

The mean and standard deviation of the data in this diagram are 18.1 and 9.7 respectively, correct to 1 decimal place. Write down the mean and standard deviation of the data in the original diagram. [2]

(Q5, Jan 2009)

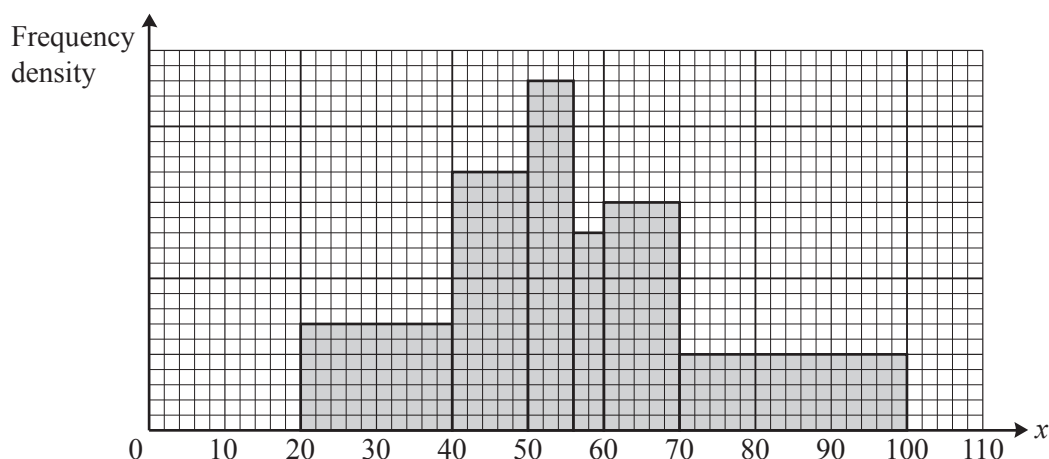
- 10** 40 people were asked to guess the length of a certain road. Each person gave their guess, l km, correct to the nearest kilometre. The results are summarised below.

l	10–12	13–15	16–20	21–30
Frequency	1	13	20	6

- (i) (a) Use appropriate formulae to calculate estimates of the mean and standard deviation of l . [6]
 (b) Explain why your answers are only estimates. [1]
- (ii) A histogram is to be drawn to illustrate the data. Calculate the frequency density of the block for the 16–20 class. [2]
- (iii) Explain which class contains the median value of l . [2]
- (iv) Later, the person whose guess was between 10 km and 12 km changed his guess to between 13 km and 15 km. Without calculation state whether the following will increase, decrease or remain the same:
- (a) the mean of l , [1]
 (b) the standard deviation of l . [1]

(Q2, Jan 2010)

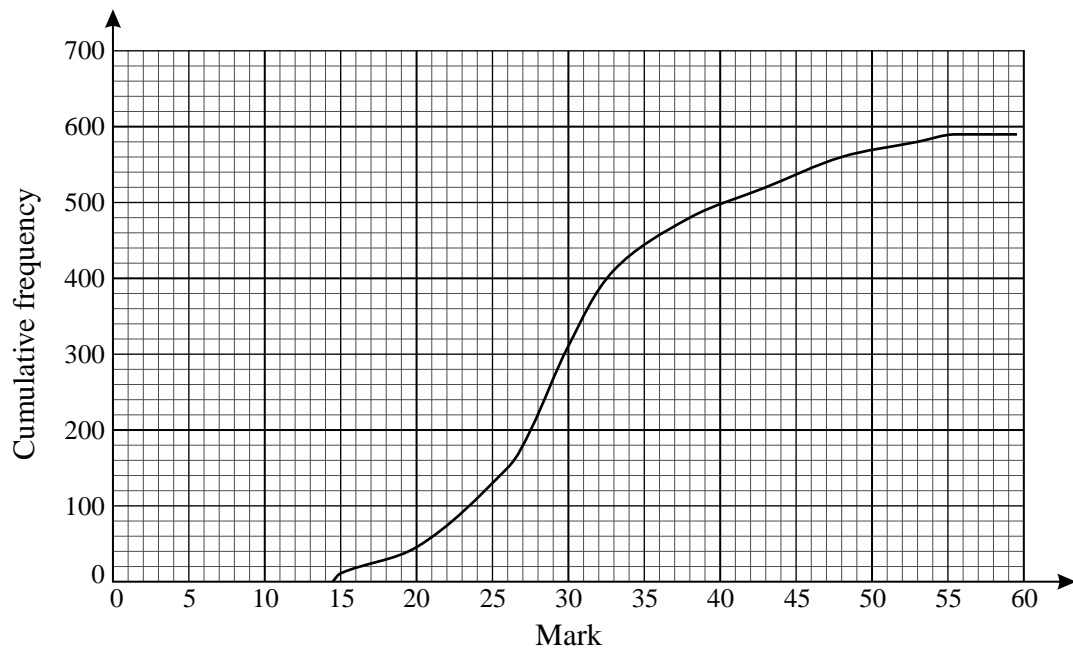
- 11** The masses, x grams, of 800 apples are summarised in the histogram.



- (i) On the frequency density axis, 1 cm represents a units. Find the value of a . [3]
 (ii) Find an estimate of the median mass of the apples. [4]

(Q6, Jan 2013)

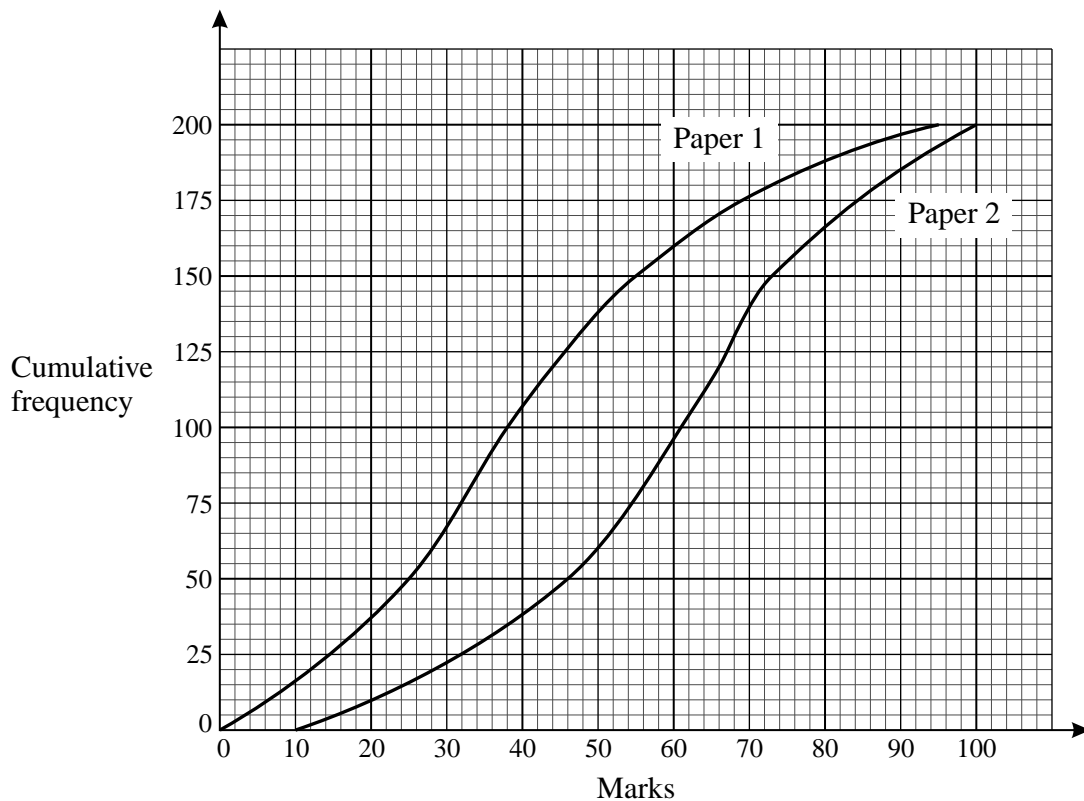
- 12** The marks of some students in a French examination were summarised in a grouped frequency distribution and a cumulative frequency diagram was drawn, as shown below.



- (i) Estimate how many students took the examination. [1]
- (ii) How can you tell that no student scored more than 55 marks? [1]
- (iii) Find the greatest possible range of the marks. [1]
- (iv) The minimum mark for Grade C was 27. The number of students who gained exactly Grade C was the same as the number of students who gained a grade lower than C. Estimate the maximum mark for Grade C. [3]
- (v) In a German examination the marks of the same students had an interquartile range of 16 marks. What does this result indicate about the performance of the students in the German examination as compared with the French examination? [3]

(Q1, June 2010)

- 13 200 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]

(Q1, Jan 2011)

- 14** The table shows information about the time, t minutes correct to the nearest minute, taken by 50 people to complete a race.

Time (minutes)	$t \leq 27$	$28 \leq t \leq 30$	$31 \leq t \leq 35$	$36 \leq t \leq 45$	$46 \leq t \leq 60$	$t \geq 61$
Number of people	0	4	28	14	4	0

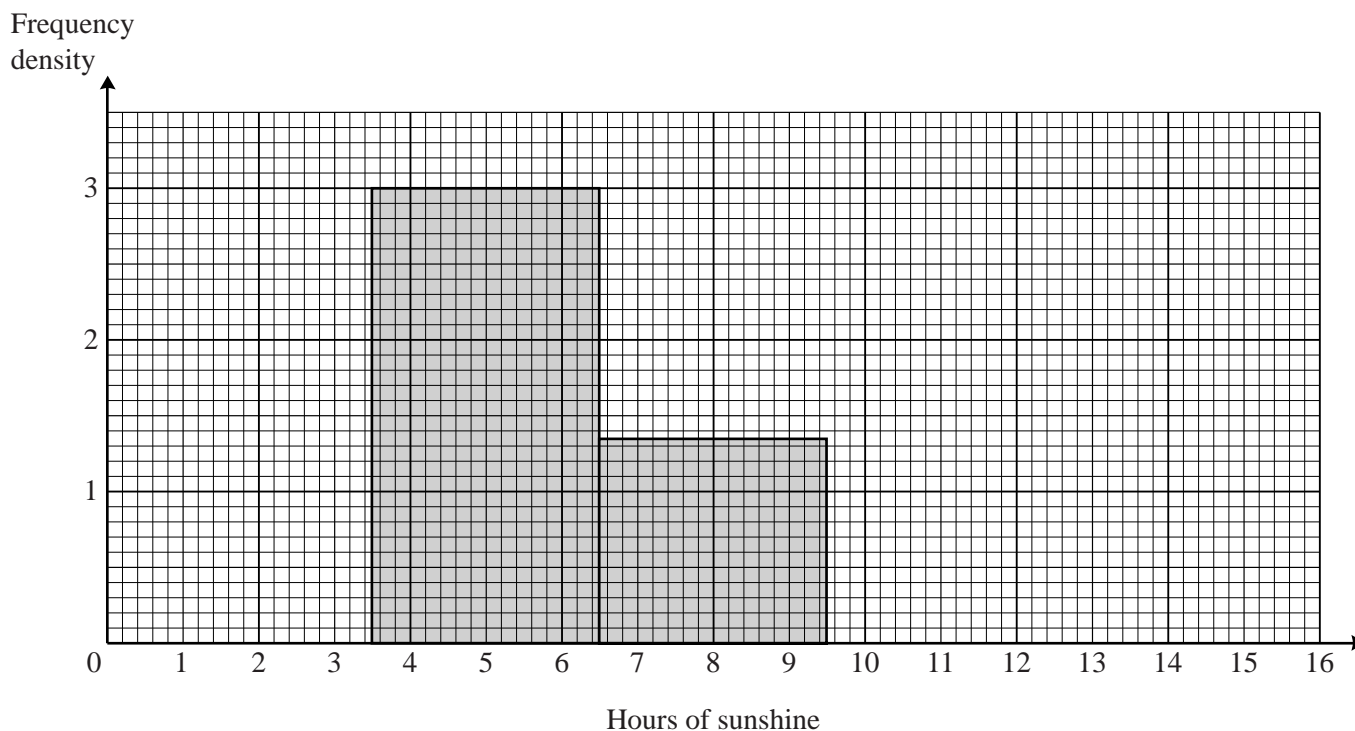
- (i) In a histogram illustrating the data, the height of the block for the $31 \leq t \leq 35$ class is 5.6 cm. Find the height of the block for the $28 \leq t \leq 30$ class. (There is no need to draw the histogram.) [3]
- (ii) The data in the table are used to estimate the median time. State, with a reason, whether the estimated median time is more than 33 minutes, less than 33 minutes or equal to 33 minutes. [3]
- (iii) Calculate estimates of the mean and standard deviation of the data. [6]
- (iv) It was found that the winner's time had been incorrectly recorded and that it was actually less than 27 minutes 30 seconds. State whether each of the following will increase, decrease or remain the same:
- (a) the mean, [1]
 - (b) the standard deviation, [1]
 - (c) the median, [1]
 - (d) the interquartile range. [1]

(Q4, June 2011)

- 15** 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]

(Q5, Jan 2012)

- 16** The test marks of 14 students are displayed in a stem-and-leaf diagram, as shown below.

0		
1		2 6
2		1 3 5
3		w x 4 8 y z
4		6 7 7

Key: 1 | 6 means 16 marks

- (i) Find the lower quartile. [1]
- (ii) Given that the median is 32, find the values of w and x . [2]
- (iii) Find the possible values of the upper quartile. [2]
- (iv) State one advantage of a stem-and-leaf diagram over a box-and-whisker plot. [1]
- (v) State one advantage of a box-and-whisker plot over a stem-and-leaf diagram. [1]

(Q3, June 2012)