www.mrc-papers.com

Probability & Statistics 1

TOPIC- Representation of data

Histogram, Cumulative frequency

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

1 The times in minutes for seven students to become proficient at a new computer game were measured. The results are shown below.

3-62-10

15 10 48 10 19 14 16

(i) Find the mean and standard deviation of these times. [2]

(ii) State which of the mean, median or mode you consider would be most appropriate to use as a measure of central tendency to represent the data in this case. [1]

(iii) For each of the two measures of average you did not choose in part **(ii)**, give a reason why you consider it inappropriate. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

2

Ashfaq and Kuljit have done a school statistics project on the prices of a particular model of headphones for MP3 players. Ashfaq collected prices from 21 shops. Kuljit used the internet to collect prices from 163 websites.

S-63-12

(i) Name a suitable statistical diagram for Ashfaq to represent his data, together with a reason for choosing this particular diagram. [2]

(ii) Name a suitable statistical diagram for Kuljit to represent her data, together with a reason for choosing this particular diagram. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Rani and Diksha go shopping for clothes.

0 3 **(i)** Rani buys 4 identical vests, 3 identical sweaters and 1 coat. Each vest costs \$5.50 and the coat costs \$90. The mean cost of Rani's 8 items is \$29. Find the cost of a sweater. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(ii) Diksha buys 1 hat and 4 identical shirts. The mean cost of Diksha's 5 items is \$26 and the standard deviation is \$0. Explain how you can tell that Diksha spends \$104 on shirts. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

© UCLES 2017

4 In a survey, the percentage of meat in a certain type of take-away meal was found. The results, to the nearest integer, for 193 take-away meals are summarised in the table. N-63-12

Percentage of meat	1 – 5	6 – 10	11 – 20	21 – 30	31 – 50
Frequency	59	67	38	18	11

(i) Calculate estimates of the mean and standard deviation of the percentage of meat in these take-away meals. [4]

(ii) Draw, on graph paper, a histogram to illustrate the information in the table. [5]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

5 A hotel has 90 rooms. The table summarises information about the number of rooms occupied each day for a period of 200 days.

Number of rooms occupied	1 – 20	21 – 40	41 – 50	51 – 60	61 – 70	71 – 90
Frequency	10	32	62	50	28	18

5-62 -11

(i) Draw a cumulative frequency graph on graph paper to illustrate this information. [4]

(ii) Estimate the number of days when over 30 rooms were occupied. [2]

(iii) On 75% of the days at most n rooms were occupied. Estimate the value of n. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

6 On a certain day in spring, the heights of 200 daffodils are measured, correct to the nearest centimetre. The frequency distribution is given below.

Height (cm)	4 – 10	11 – 15	16 – 20	21 – 25	26 – 30
Frequency	22	32	78	40	28

(i) Draw a cumulative frequency graph to illustrate the data. [4]

(ii) 28% of these daffodils are of height h cm or more. Estimate h. [2]

(iii) You are given that the estimate of the mean height of these daffodils, calculated from the table, is 18.39 cm. Calculate an estimate of the standard deviation of the heights of these daffodils. [3]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

07 Robert has a part-time job delivering newspapers. On a number of days he noted the time, correct to the nearest minute, that it took him to do his job. Robert used his results to draw up the following table; two of the values in the table are denoted by a and b.

N-61-15

Time (t minutes)	60 – 62	63 – 64	65 – 67	68 – 71
Frequency (number of days)	3	9	6	b
Frequency density	1	a	2	1.5

(i) Find the values of a and b. [3]

(ii) On graph paper, draw a histogram to represent Robert's times. [3]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

0 8 The number of people a football stadium can hold is called the 'capacity'. The capacities of 130 football stadiums in the UK, to the nearest thousand, are summarised in the table. N– 62–16

Capacity	3000–7000	8000–12 000	13 000–22 000	23 000–42 000	43 000–82 000
Number of stadiums	40	30	18	34	8

(i) On graph paper, draw a histogram to represent this information. Use a scale of 2 cm for a capacity of 10 000 on the horizontal axis. [5]

(ii) Calculate an estimate of the mean capacity of these 130 stadiums. [2]

(iii) Find which class in the table contains the median and which contains the lower quartile. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

0 9 The table summarises the lengths in centimetres of 104 dragonflies. 5- 61- 15

Length (cm)	2.0 – 3.5	3.5 – 4.5	4.5 – 5.5	5.5 – 7.0	7.0 – 9.0
Frequency	8	25	28	31	12

(i) State which class contains the upper quartile. [1]

(ii) Draw a histogram, on graph paper, to represent the data. [4]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

10 The heights to the nearest metre of 134 office buildings in a certain city are summarised in the table below.

N = 63-15

Height (m)	21 – 40	41 – 45	46 – 50	51 – 60	61 – 80
Frequency	18	15	21	52	28

(i) Draw a histogram on graph paper to illustrate the data. [4]

(ii) Calculate estimates of the mean and standard deviation of these heights. [5]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

11 The weights in grams of a number of stones, measured correct to the nearest gram, are represented in the following table.

Weight (grams)	1 – 10	11 – 20	21 – 25	26 – 30	31 – 50	51 – 70
Frequency	$2x$	$4x$	$3x$	$5x$	$4x$	x

A histogram is drawn with a scale of 1 cm to 1 unit on the vertical axis, which represents frequency density. The 1 – 10 rectangle has height 3 cm.

N – 61 – 10

(i) Calculate the value of x and the height of the 51 – 70 rectangle. [4]

(ii) Calculate an estimate of the mean weight of the stones. [3]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

12 The weights, x kilograms, of 144 people were recorded. The results are summarised in the cumulative frequency table below.

S-G3-13

Weight (x kilograms)	$x < 40$	$x < 50$	$x < 60$	$x < 65$	$x < 70$	$x < 90$
Cumulative frequency	0	12	34	64	92	144

(i) On graph paper, draw a cumulative frequency graph to represent these results. [2]

(ii) 64 people weigh more than c kg. Use your graph to find the value of c. [2]

(iii) Calculate estimates of the mean and standard deviation of the weights. [6]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

13 The weights of 220 sausages are summarised in the following table. N-62-11

Weight (grams)	<20	<30	<40	<45	<50	<60	<70
Cumulative frequency	0	20	50	100	160	210	220

(i) State which interval the median weight lies in. [1]

(ii) Find the smallest possible value and the largest possible value for the interquartile range. [2]

(iii) State how many sausages weighed between 50 g and 60 g. [1]

(iv) On graph paper, draw a histogram to represent the weights of the sausages. [4]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

14 The following cumulative frequency table shows the examination marks for 300 candidates in country A and 300 candidates in country B.

5-63-11

Mark	<10	<20	<35	<50	<70	<100
Cumulative frequency, A	25	68	159	234	260	300
Cumulative frequency, B	10	46	72	144	198	300

(i) Without drawing a graph, show that the median for country B is higher than the median for country A. [2]

(ii) Find the number of candidates in country A who scored between 20 and 34 marks inclusive. [1]

(iii) Calculate an estimate of the mean mark for candidates in country A. [4]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

15 Seventy samples of fertiliser were collected and the nitrogen content was measured for each sample. The cumulative frequency distribution is shown in the table below.

Nitrogen content	$\leqslant 3.5$	$\leqslant 3.8$	$\leqslant 4.0$	$\leqslant 4.2$	$\leqslant 4.5$	$\leqslant 4.8$
Cumulative frequency	0	6	18	41	62	70

5– 63–15

(i) On graph paper draw a cumulative frequency graph to represent the data. [3]

(ii) Estimate the percentage of samples with a nitrogen content greater than 4.4. [2]

(iii) Estimate the median. [1]

(iv) Construct the frequency table for these results and draw a histogram on graph paper. [5]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

16 There are 5000 schools in a certain country. The cumulative frequency table shows the number of pupils in a school and the corresponding number of schools. S-G1-11

Number of pupils in a school	⩽100	⩽150	⩽200	⩽250	⩽350	⩽450	⩽600
Cumulative frequency	200	800	1600	2100	4100	4700	5000

(i) Draw a cumulative frequency graph with a scale of 2 cm to 100 pupils on the horizontal axis and a scale of 2 cm to 1000 schools on the vertical axis. Use your graph to estimate the median number of pupils in a school. [3]

(ii) 80% of the schools have more than n pupils. Estimate the value of n correct to the nearest ten. [2]

(iii) Find how many schools have between 201 and 250 (inclusive) pupils. [1]

(iv) Calculate an estimate of the mean number of pupils per school. [4]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

17 The amounts spent by 160 shoppers at a supermarket are summarised in the following table.

Amount spent ($\$x$)	$0 < x \leqslant 30$	$30 < x \leqslant 50$	$50 < x \leqslant 70$	$70 < x \leqslant 90$	$90 < x \leqslant 140$
Number of shoppers	16	40	48	26	30

S—61-16

(i) Draw a cumulative frequency graph of this distribution. [4]

(ii) Estimate the median and the interquartile range of the amount spent. [3]

(iii) Estimate the number of shoppers who spent more than \$115. [2]

(iv) Calculate an estimate of the mean amount spent. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

18 The amounts spent by 160 shoppers at a supermarket are summarised in the following table.

Amount spent ($\$x$)	$0 < x \leqslant 30$	$30 < x \leqslant 50$	$50 < x \leqslant 70$	$70 < x \leqslant 90$	$90 < x \leqslant 140$
Number of shoppers	16	40	48	26	30

S-61-16

(i) Draw a cumulative frequency graph of this distribution. [4]

(ii) Estimate the median and the interquartile range of the amount spent. [3]

(iii) Estimate the number of shoppers who spent more than \$115. [2]

(iv) Calculate an estimate of the mean amount spent. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

19 The tables summarise the heights, h cm, of 60 girls and 60 boys.

N-G 3-16

Height of girls (cm)	$140 < h \leqslant 150$	$150 < h \leqslant 160$	$160 < h \leqslant 170$	$170 < h \leqslant 180$	$180 < h \leqslant 190$
Frequency	12	21	17	10	0

Height of boys (cm)	$140 < h \leqslant 150$	$150 < h \leqslant 160$	$160 < h \leqslant 170$	$170 < h \leqslant 180$	$180 < h \leqslant 190$
Frequency	0	20	23	12	5

(i) On graph paper, using the same set of axes, draw two cumulative frequency graphs to illustrate the data. [4]

(ii) On a school trip the students have to enter a cave which is 165 cm high. Use your graph to estimate the percentage of the girls who will be unable to stand upright. [3]

(iii) The students are asked to compare the heights of the girls and the boys. State one advantage of using a pair of box-and-whisker plots instead of the cumulative frequency graphs to do this. [1]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

20 The table summarises the times that 112 people took to travel to work on a particular day. N-6241

Time to travel to work (t minutes)	$0 < t \leqslant 10$	$10 < t \leqslant 15$	$15 < t \leqslant 20$	$20 < t \leqslant 25$	$25 < t \leqslant 40$	$40 < t \leqslant 60$
Frequency	19	12	28	22	18	13

(i) State which time interval in the table contains the median and which time interval contains the upper quartile. [2]

(ii) On graph paper, draw a histogram to represent the data. [4]

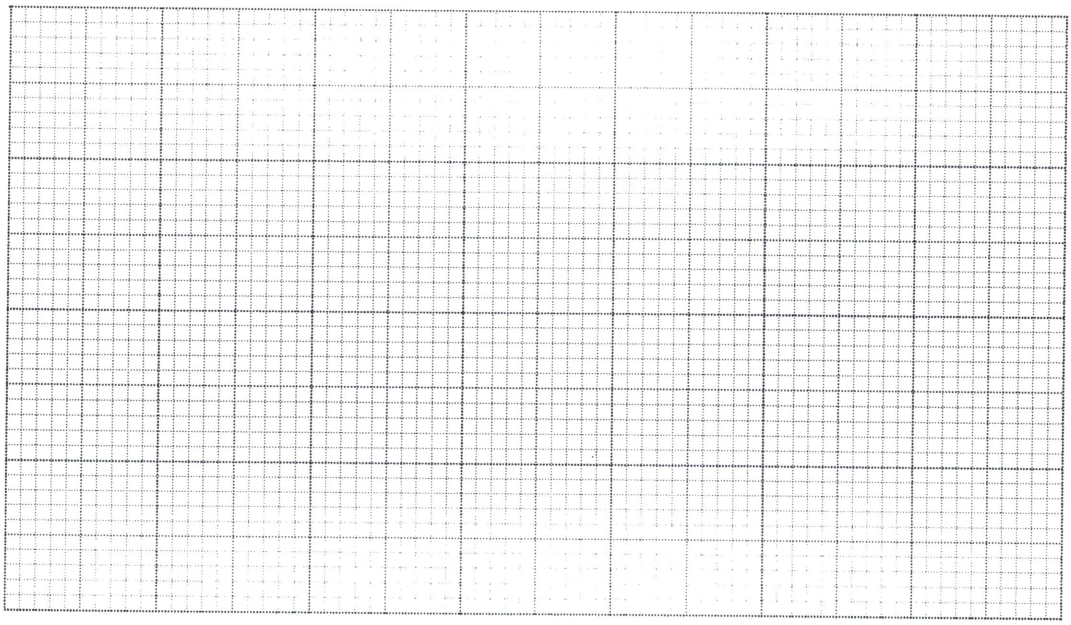
(iii) Calculate an estimate of the mean time to travel to work. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

21 The times taken, t seconds, by 1140 people to solve a puzzle are summarised in the table.

Time (t seconds)	$0 \leqslant t < 20$	$20 \leqslant t < 40$	$40 \leqslant t < 60$	$60 \leqslant t < 100$	$100 \leqslant t < 140$
Number of people	320	280	220	220	100

(i) On the grid, draw a histogram to illustrate this information. [4]



(ii) Calculate an estimate of the mean of t. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

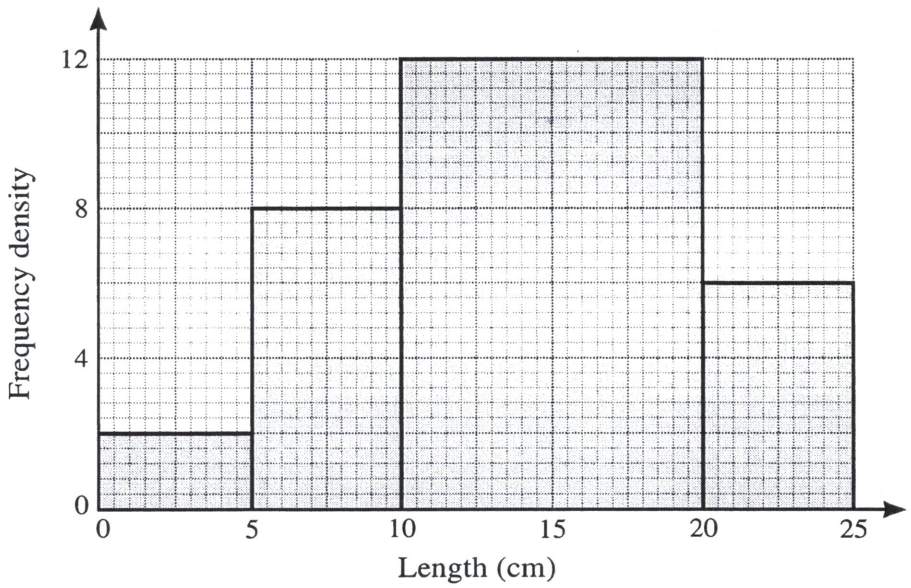
.....

.....

© UCLES 2017

[Turn over

22 The following histogram represents the lengths of worms in a garden.



(i) Calculate the frequencies represented by each of the four histogram columns. [2]

.....

.....

.....

.....

.....

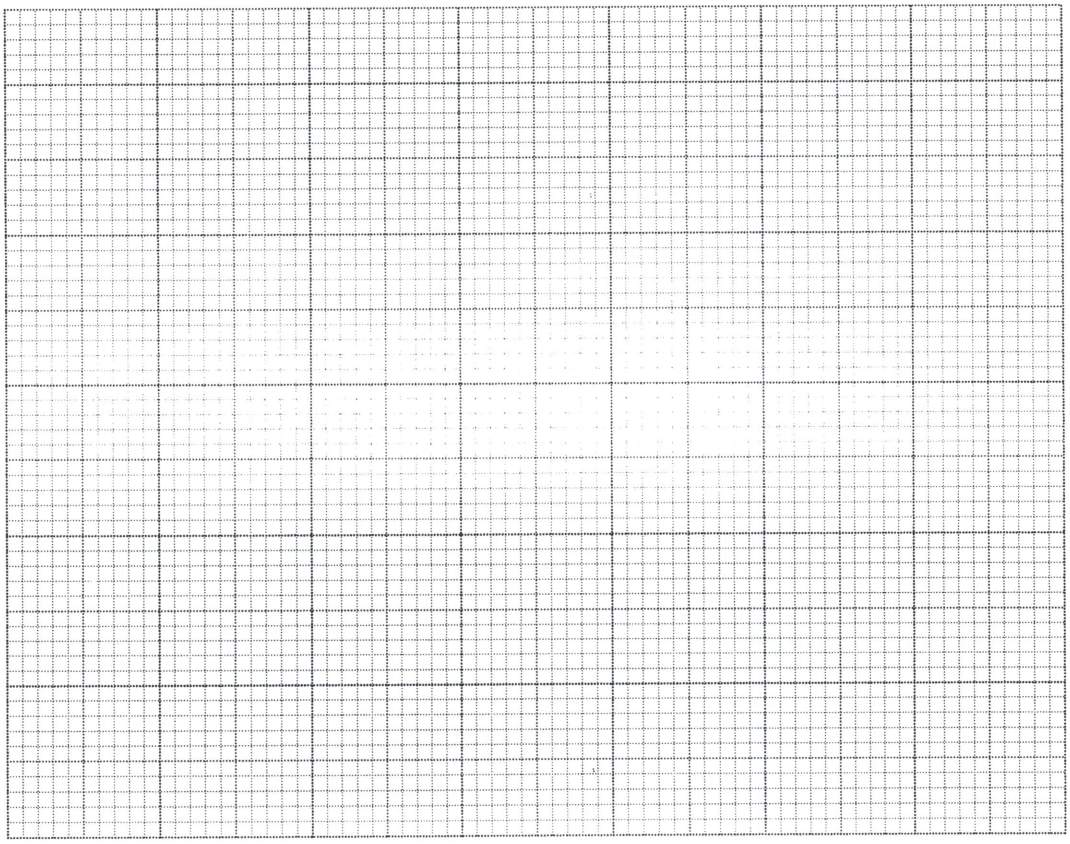
.....

.....

.....

(ii) On the grid on the next page, draw a cumulative frequency graph to represent the lengths of worms in the garden. [4]

© UCLES 2017



(iii) Use your graph to estimate the median and interquartile range of the lengths of worms in the garden. [3]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

[Question 7 (iv) is printed on the next page.]

(iv) Calculate an estimate of the mean length of worms in the garden. [2]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

To avoid the issue of disclosure of answer-related information to candidates, all copyright acknowledgements are reproduced online in the Cambridge International Examinations Copyright Acknowledgements Booklet. This is produced for each series of examinations and is freely available to download at www.cie.org.uk after the live examination series.

Cambridge International Examinations is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.

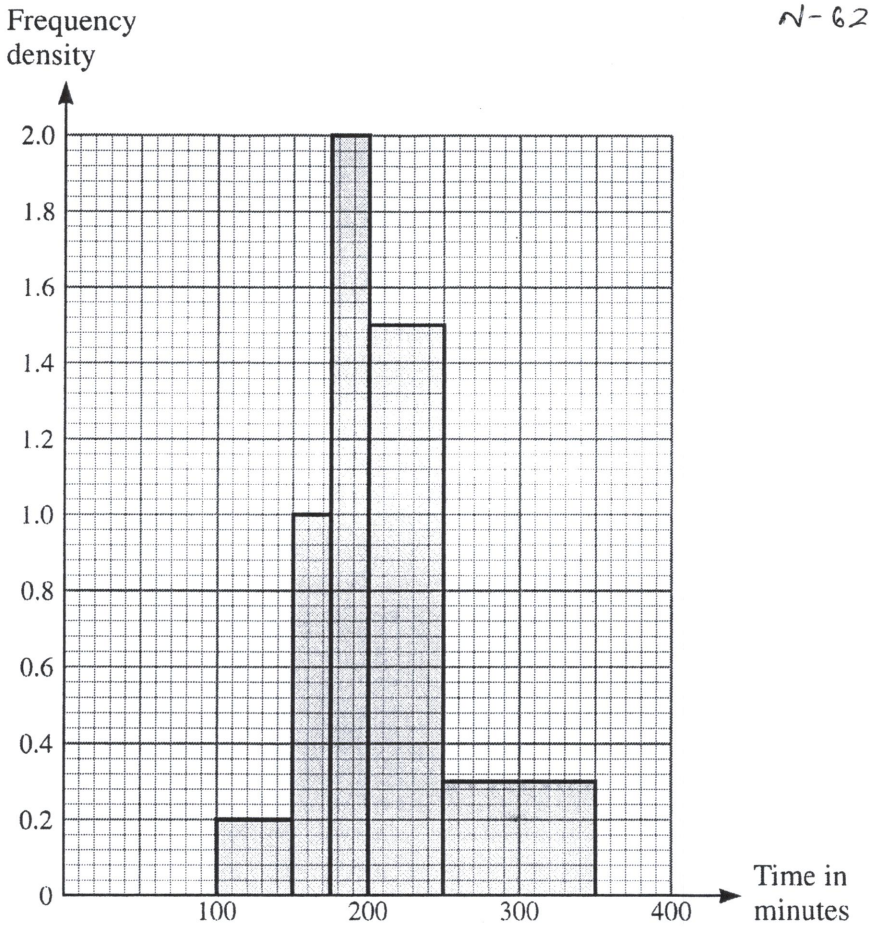
© UCLES 2017

Representation of data

The following histogram summarises the times, in minutes, taken by 190 people to complete a race.

23

N-62-13



(i) Show that 75 people took between 200 and 250 minutes to complete the race. [1]

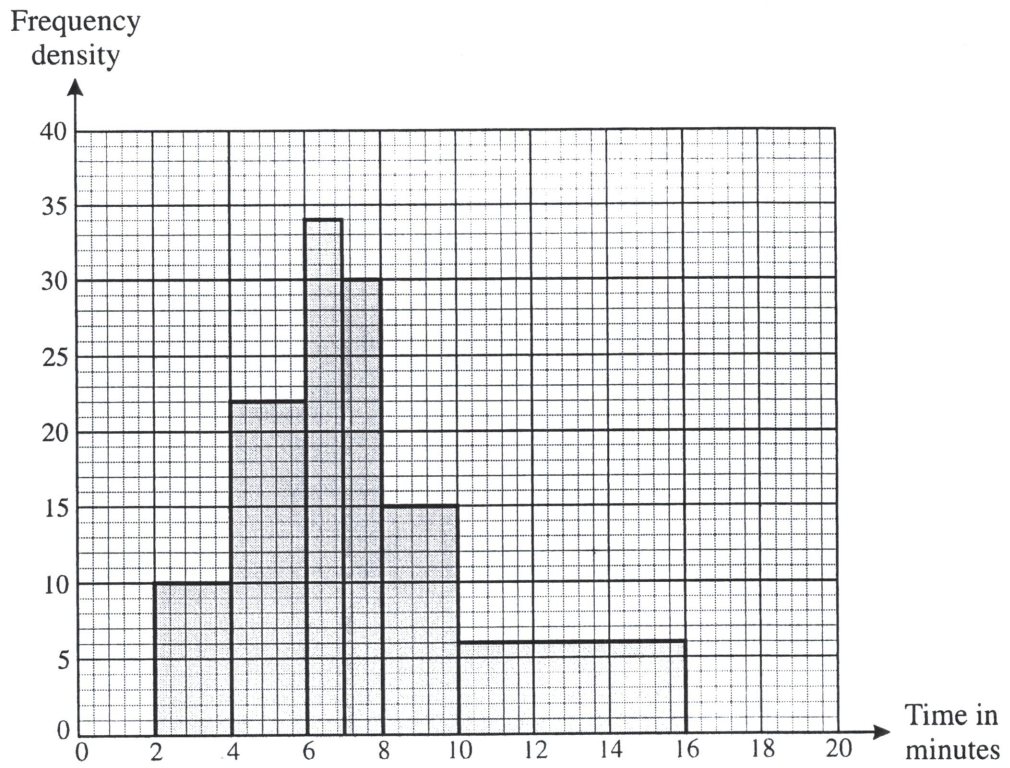
(ii) Calculate estimates of the mean and standard deviation of the times of the 190 people. [6]

(iii) Explain why your answers to part **(ii)** are estimates. [1]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

24 The following histogram illustrates the distribution of times, in minutes, that some students spent taking a shower.

N-G3-10



(i) Copy and complete the following frequency table for the data. [3]

Time (t minutes)	$2 < t \leqslant 4$	$4 < t \leqslant 6$	$6 < t \leqslant 7$	$7 < t \leqslant 8$	$8 < t \leqslant 10$	$10 < t \leqslant 16$
Frequency						

(ii) Calculate an estimate of the mean time to take a shower. [2]

(iii) Two of these students are chosen at random. Find the probability that exactly one takes between 7 and 10 minutes to take a shower. [3]

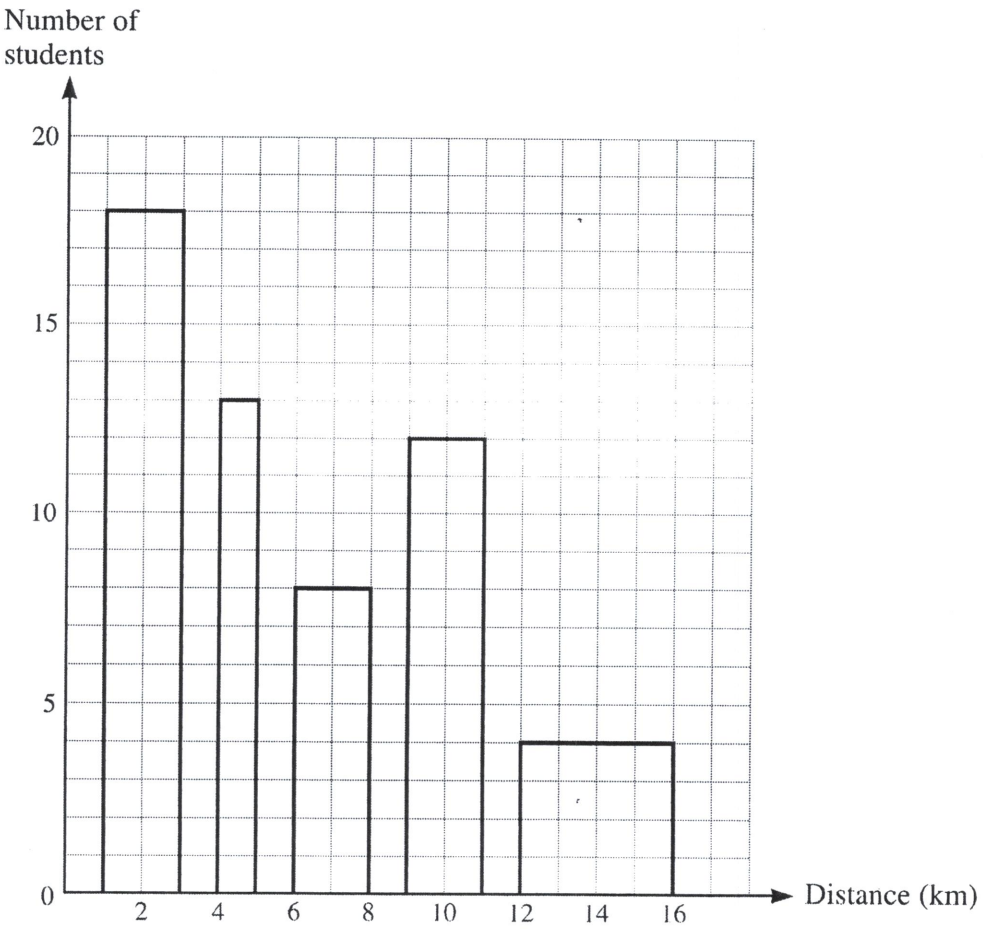
Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

Representation of data

2 The distance of a student's home from college, correct to the nearest kilometre, was recorded for each of 55 students. The distances are summarised in the following table. N-63-13

Distance from college (km)	1 – 3	4 – 5	6 – 8	9 – 11	12 – 16
Number of students	18	13	8	12	4

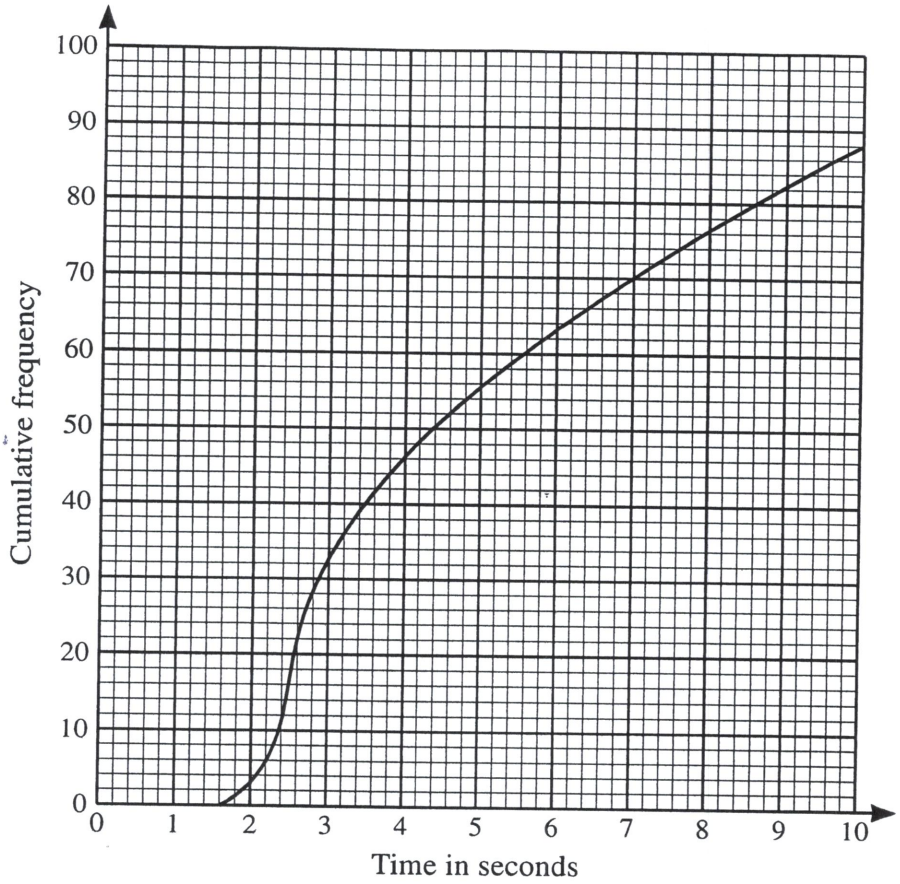
Dominic is asked to draw a histogram to illustrate the data. Dominic's diagram is shown below.



Give two reasons why this is not a correct histogram. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

26



In an open-plan office there are 88 computers. The times taken by these 88 computers to access a particular web page are represented in the cumulative frequency diagram. 5-6 2-15

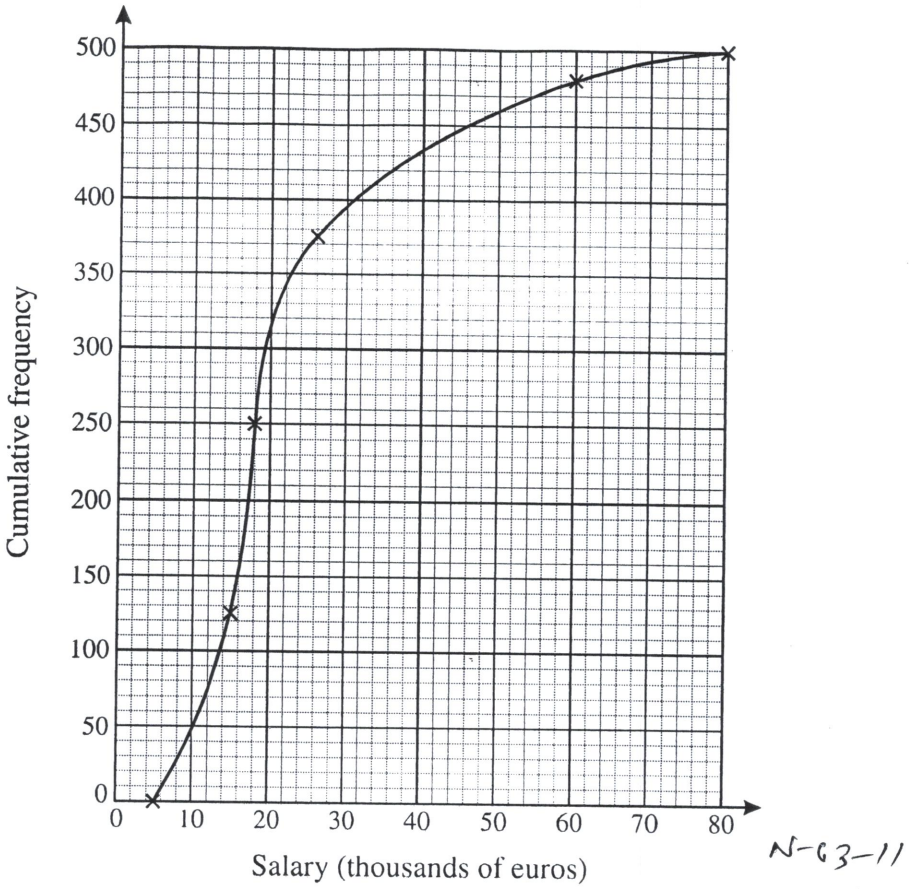
(i) On graph paper draw a box-and-whisker plot to summarise this information. [4]

An 'outlier' is defined as any data value which is more than 1.5 times the interquartile range above the upper quartile, or more than 1.5 times the interquartile range below the lower quartile.

(ii) Show that there are no outliers. [2]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

27



N-03-11

The cumulative frequency graph shows the annual salaries, in thousands of euros, of a random sample of 500 adults with jobs, in France. It has been plotted using grouped data. You may assume that the lowest salary is 5000 euros and the highest salary is 80 000 euros.

(i) On graph paper, draw a box-and-whisker plot to illustrate these salaries. [4]

(ii) Comment on the salaries of the people in this sample. [1]

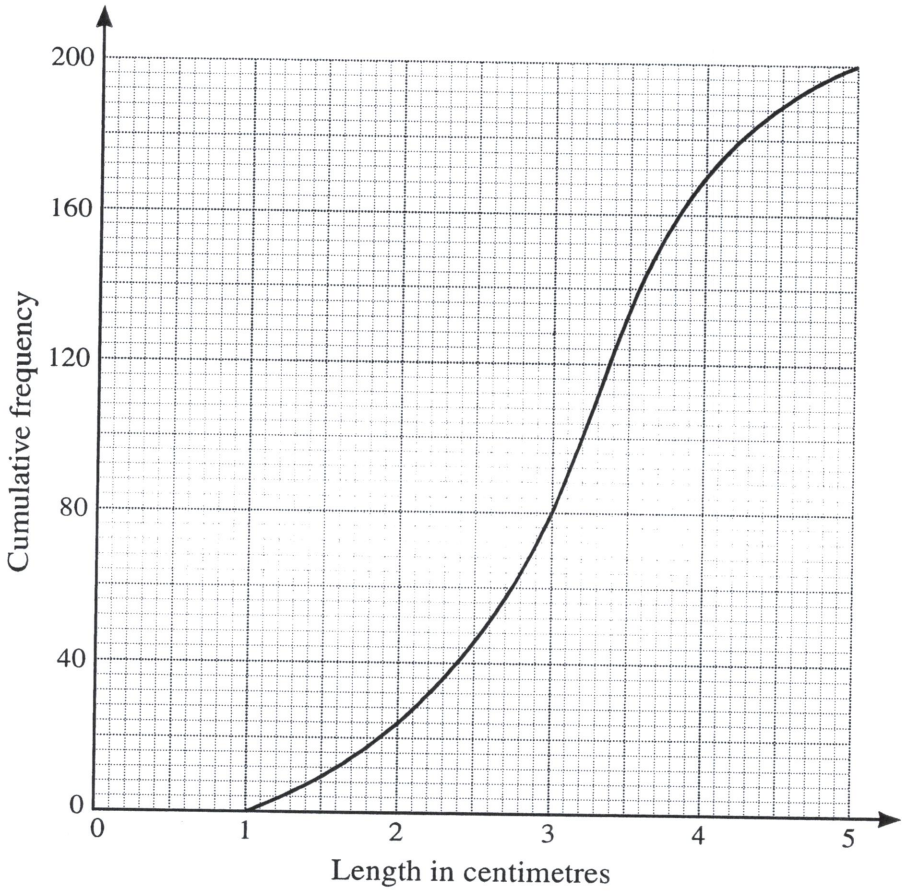
(iii) An 'outlier' is defined as any data value which is more than 1.5 times the interquartile range above the upper quartile, or more than 1.5 times the interquartile range below the lower quartile.

(a) How high must a salary be in order to be classified as an outlier? [3]

(b) Show that none of the salaries is low enough to be classified as an outlier. [1]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com

28 Anabel measured the lengths, in centimetres, of 200 caterpillars. Her results are illustrated in the cumulative frequency graph below.



(i) Estimate the median and the interquartile range of the lengths. [3]

.....

.....

.....

(ii) Estimate how many caterpillars had a length of between 2 and 3.5 cm. [1]

.....

.....

(iii) 6% of caterpillars were of length l centimetres or more. Estimate l. [2]

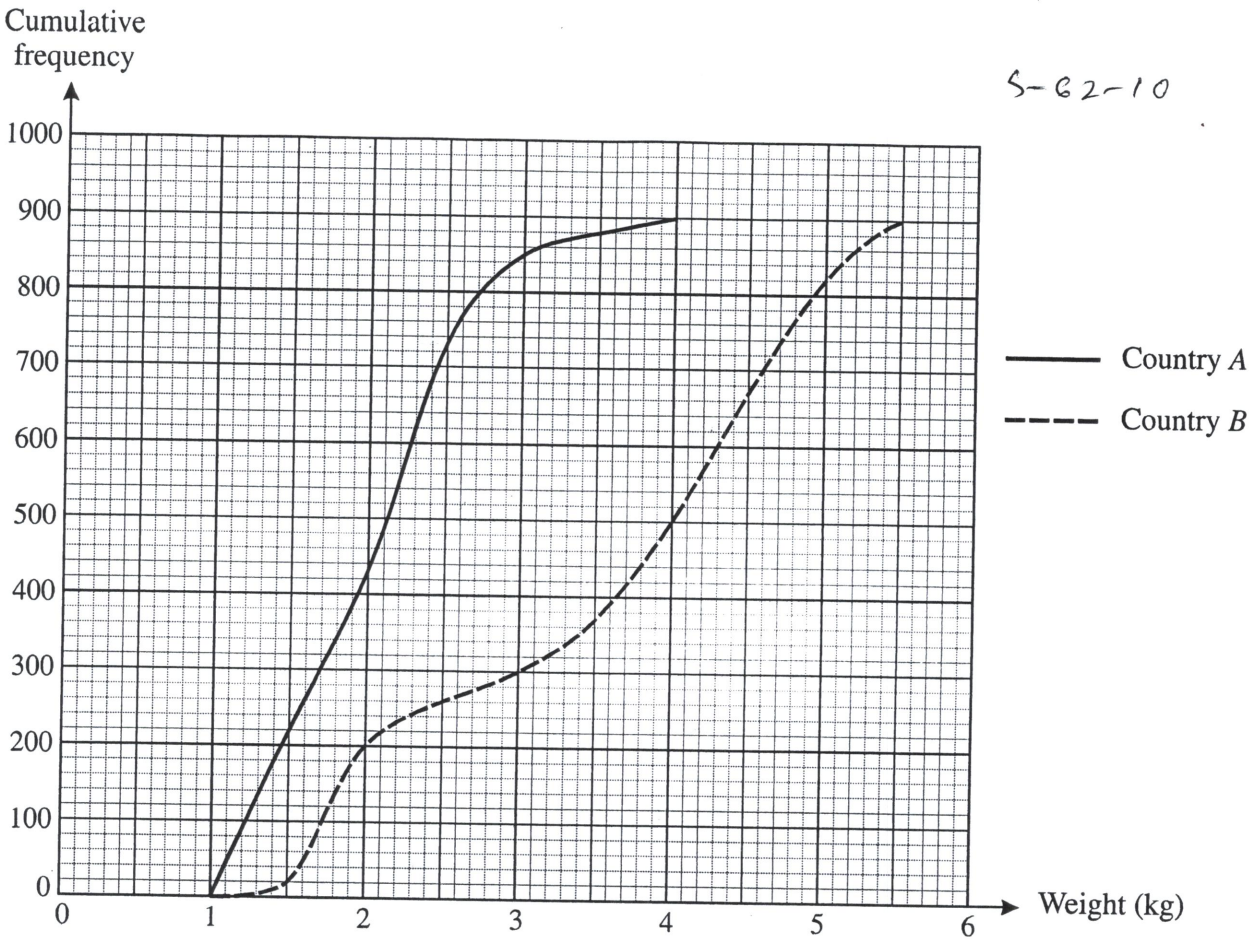
.....

.....

.....

© UCLES 2017

29



The birth weights of random samples of 900 babies born in country *A* and 900 babies born in country *B* are illustrated in the cumulative frequency graphs. Use suitable data from these graphs to compare the central tendency and spread of the birth weights of the two sets of babies. [6]

Copyright©MRC® :(**9709/06 –S1**) - 2010-2017 +97455258711 rashed.saba@gmail.com