Statistics 5 – Histograms and cumulative frequency

Please **complete** this homework by _____. Start it early. If you can't do a question you will then have time to ask your teacher for help or go to a drop in session.

Section 1 – Review of previous topics. Please complete all questions.

Q1 Thirteen students sat a test, marked out of 50. These are their scores:

33, 25, 36, 40, 14, 49, 31, 17, 31, 44, 35, 28, 34

a) Write down the values of Q_1, Q_2 and Q_3 for this data.

A value that is more than 1.5 times the Interquartile range (IQR) above Q_3 or more than 1.5 times the IQR below Q_1 is called an outlier.

b) Draw a box plot for this data.

Q2. In a survey, 120 shoppers in a supermarket were asked how many minutes, to the nearest minute, they had been in the store. The results are summarised in the table.

Number of minutes	Number of shoppers
1-4	5
5-9	20
10-19	28
20-29	51
30-59	16
Total	**120**

Use linear interpolation to estimate the 10% to 90% inter-percentile range of this data.

Q3. The average hourly pay at a shoe shop is £10. The standard deviation for the hourly pay is £1. The company are planning to increase the hourly rate and are choosing between two different schemes. Calculate the mean and standard deviation for each of the new schemes.

Scheme A: An increase of 50p per hour

Scheme B: A 5% increase

Q4. Show that the line $y = 2x + 11$ is a tangent to the circle $x^2 + y^2 - 6x - 4y = 32$

Section 2 – Consolidation of this week's topic. Please complete all questions. (52 marks)

Q1. For the following data, calculate frequency density. Be mindful of the class boundaries.

Waiting time (mins)	No. of patients	
1-3	30	
4	96	
5	48	
6-7	84	
8-10	27	
11-15	15	
Total	300	

(3 marks)

Q2. 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 marks)*

b) Write down the class boundary of the 1-10 hour group. *(1 mark)*

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

c) Find the width and height of the 26-30 group. *(3 marks)*

d) Estimate the mean and standard deviation of the time spent watching television by these students. *(5 marks)*

e) Use linear interpolation to estimate the median length of time spent watching television by these students. *(3 marks)*

Q3.

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

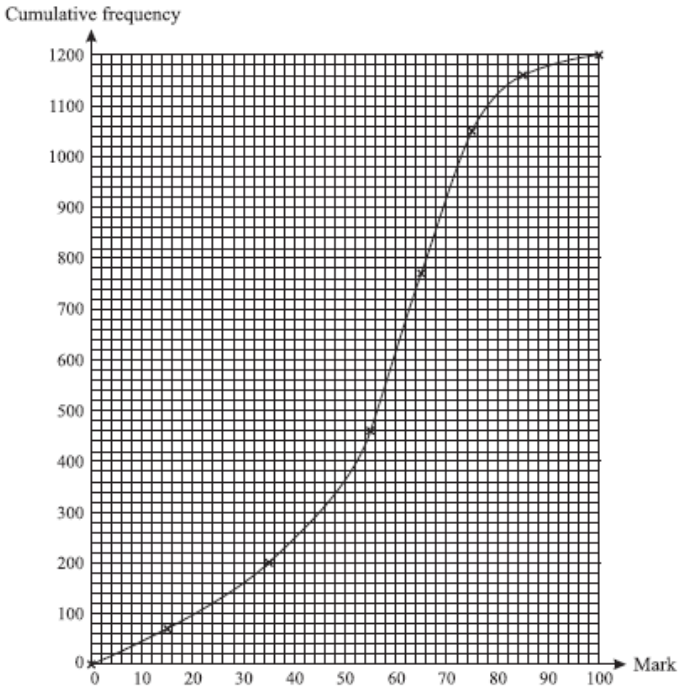
x	$10 \leqslant x \leqslant 19$	$20 \leqslant x \leqslant 24$	$25 \leqslant x \leqslant 29$	$30 \leqslant x \leqslant 49$
Number of stones	25	22	29	24

These data are to be presented on a statistical diagram.

a) For a cumulative frequency curve, state the co-ordinates of the first two points that should be plotted.

b) Why is it not possible to draw an exact box plot to illustrate the data? *(3 marks)*

Q4. The examination marks obtained by 1200 candidates are illustrated on the cumulative frequency curve where the data points are joined by a smooth curve.



Use the curve to estimate

a) The interquartile range of the marks, *(3 marks)*

b) x, if 40% of the candidates scored more than x marks *(3 marks)*

c) The number of candidates who scored more than 68 marks *(2 marks)*

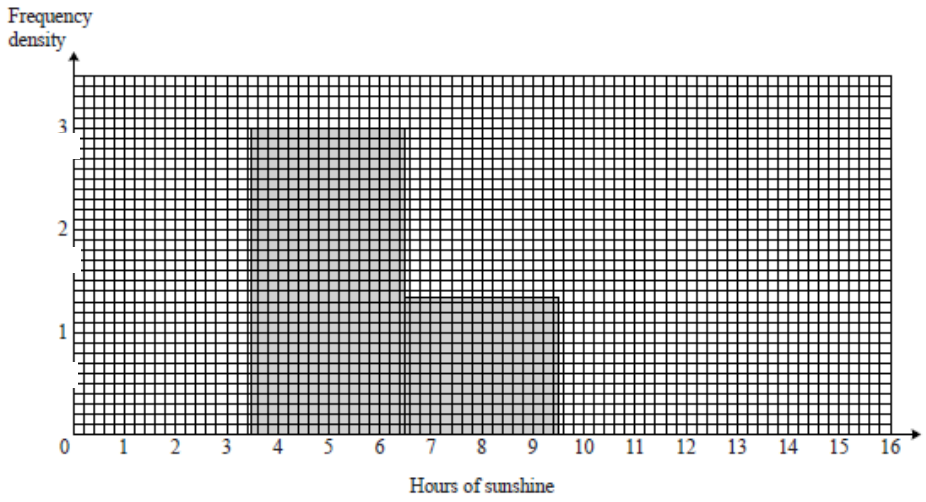
Five of the candidates are selected at random, with replacement

d) Estimate the probability that all five scored more than 68 marks *(3 marks)*

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		4	2

The diagram shows part of a histogram to illustrate the data. The scale on the frequency axis is 2cm to 1 Unit (10 squares).



a) Give a reason to justify the use of a histogram to represent these data. *(1 mark)*

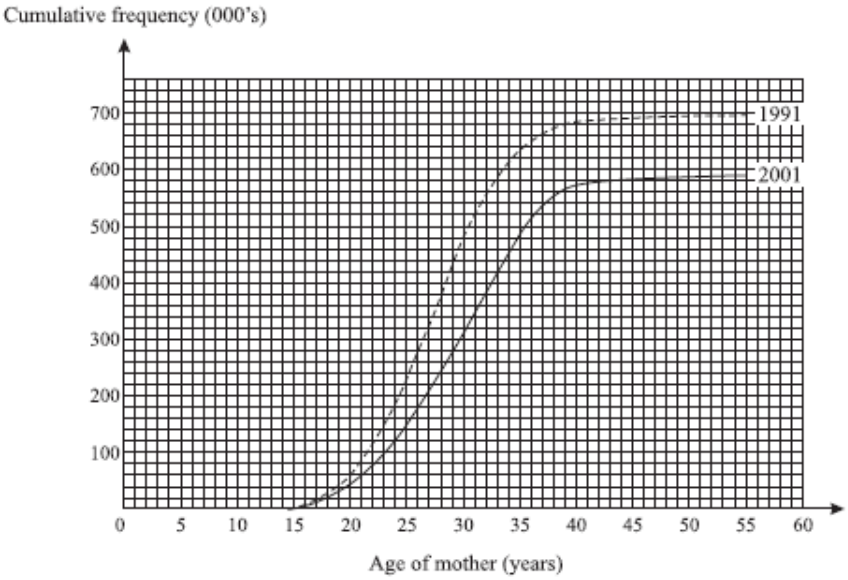
b) Complete the Histogram and the table. *(5 marks)*

c) Estimate the number of days with between 4 and 8 hours of sunshine. *(3 marks)*

d) Use your calculator to find the mean and standard deviation of hours of sunshine per day *(2 marks)*

e) On the same set of axes, draw a frequency polygon. *(2 marks)*

Q6. 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



a) In which of these two years were there more births? How many more births were there in this year? *(2 marks)*

b) 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%

i) Find the values missing from the table *(5 marks)*

ii) 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 value from part (a), give 2 reasons for your answer. *(2 marks)*

Total marks = 50 marks

Section 3 – Extension and Large Data set questions. If you are aiming for a top grade, you should attempt these questions

Q1. Here is the data for Daily mean air temperature for Bejing May - Oct 2015 from the 'Large Data set' Group the data appropriately and fill in the table, finding frequencies and Frequency density. Draw a Histogram to represent the data.

23.3	17.5
23.1	20.0
24.0	19.2
22.8	18.5
23.7	21.1
22.2	17.1
22.9	18.8
25.0	18.0
6.3	13.0
27.1	9.7
27.2	14.3
26.3	18.9
25.6	24.6
23.9	22.2
24.9	19.6
28.0	

Daily mean air temperature	Frequency	Frequency density