Probability & Statistics

Chapter 1

Representation of data

Cambridge AS level

1.4 REPRESENTATION OF CONTINUOUS DATA: CUMULATIVE FREQUENCY GRAPHS

A cumulative frequency graph can be used to represent continuous data. **Cumulative frequency** is the total frequency of all values less than a given value.

If we are given grouped data, we can construct the cumulative frequency diagram by plotting cumulative frequencies (abbreviated to cf) against upper class boundaries for all intervals. We can join the points consecutively with straight-line segments to give a cumulative frequency polygon or with a smooth curve to give a cumulative frequency curve.

1.4 REPRESENTATION OF CONTINUOUS DATA: CUMULATIVE FREQUENCY GRAPHS

From a cumulative frequency graph we can estimate the number or proportion of values that lie above or below a given value, or between two values.

TIP: A common mistake is to plot points at class mid-values but the total frequency up to each mid-value is not precise – it is an estimate.

Example 1:

The following table shows the lengths of 80 leaves from a particular tree, given to the nearest centimetre.

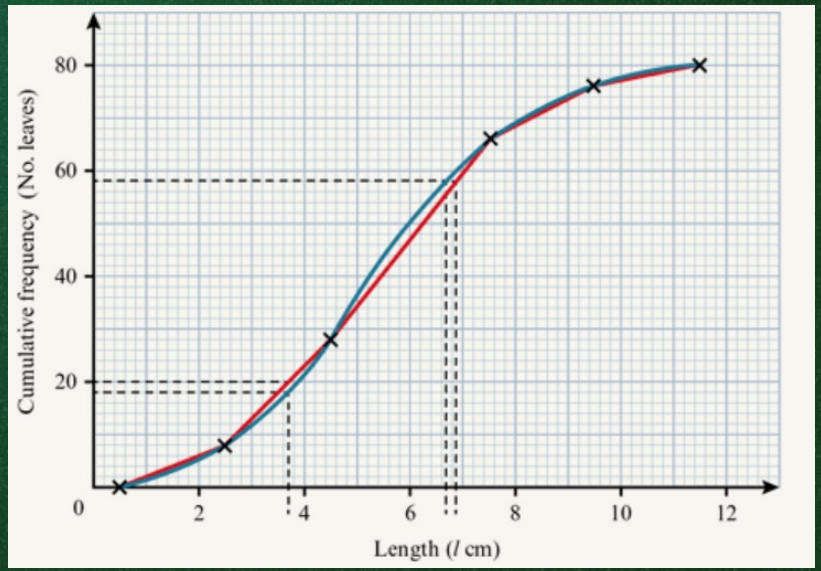
Lengths (cm)	1–2	3–4	5–7	8–9	10–11
No. leaves (f)	8	20	38	10	4

Draw a cumulative frequency curve and a cumulative frequency polygon. Use each of these to estimate:

a. the number of leaves that are less than 3.7cm long

b. the lower boundary of the lengths of the longest 22 leaves.

Lengths (l cm)	Addition of frequencies	No. leaves (cf)
$l < 0.5$	0	0
$l < 2.5$	0 + 8	8
$l < 4.5$	0 + 8 + 20	28
$l < 7.5$	0 + 8 + 20 + 38	66
$l < 9.5$	0 + 8 + 20 + 38 + 10	76
$l < 11.5$	0 + 8 + 20 + 38 + 10 + 4	80



Example 1:

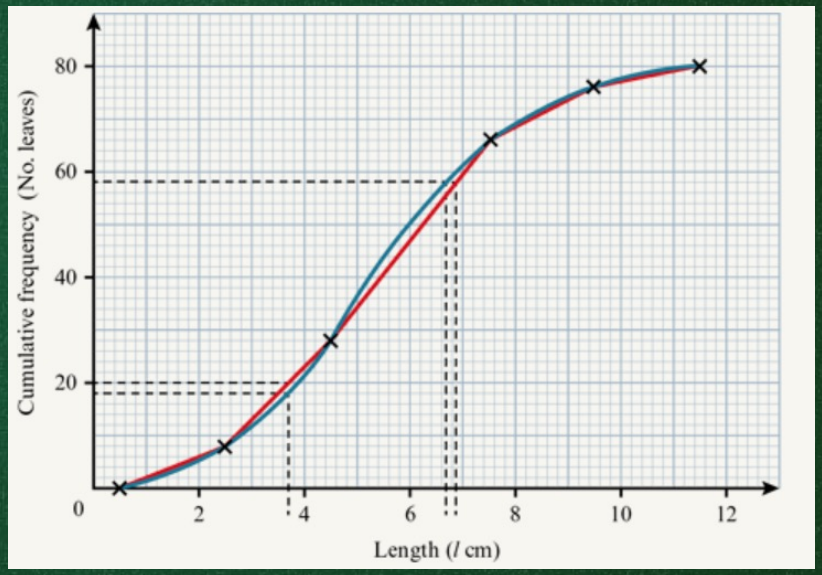
The following table shows the lengths of 80 leaves from a particular tree, given to the nearest centimetre.

Draw a cumulative frequency curve and a cumulative frequency polygon. Use each of these to estimate:

a. the number of leaves that are less than 3.7cmlong

b. the lower boundary of the lengths of the longest 22 leaves.

Lengths (l cm)	Addition of frequencies	No. leaves (cf)
$l < 0.5$	0	0
$l < 2.5$	0 + 8	8
$l < 4.5$	0 + 8 + 20	28
$l < 7.5$	0 + 8 + 20 + 38	66
$l < 9.5$	0 + 8 + 20 + 38 + 10	76
$l < 11.5$	0 + 8 + 20 + 38 + 10 + 4	80



a.The polygon gives an estimate of 20 leaves. The curve gives an estimate of 18 leaves.

b. The polygon gives an estimate of 6.9cm. The curve gives an estimate of 6.7cm.

1.4 REPRESENTATION OF CONTINUOUS DATA: CUMULATIVE FREQUENCY GRAPHS

TIP: Do include the lowest class boundary, which has a cumulative frequency of 0, and plot this point on your graph.

- When constructing a cumulative frequency graph, you are advised to use sensible scales that allow you to plot and read values accurately.

- Estimates from a polygon and a curve will not be the same, as they coincide only at the plotted points.

- Note that all of these answers are only estimates, as we do not know the exact shape of the cumulative frequency graph between the plotted points.

Homework Page 16 – Exercise 1C

1 The reaction times, t seconds, of 66 participants were measured in an experiment and presented below.

Time (t seconds)	No. participants (cf)
$t < 1.5$	0
$t < 3.0$	3
$t < 4.5$	8
$t < 6.5$	32
$t < 8.5$	54
$t < 11.0$	62
$t < 13.0$	66

a Draw a cumulative frequency polygon to represent the data.

b Use your graph to estimate:

i the number of participants with reaction times between 5.5 and 7.5 seconds

ii the lower boundary of the slowest 20 reaction times.

2 The following table shows the widths of the 70 books in one section of a library, given to the nearest centimetre.

Width (cm)	10–14	15–19	20–29	30–39	40–44
No. books (f)	3	13	25	24	5

a Given that the upper boundary of the first class is 14.5 cm, write down the upper boundary of the second class.

b Draw up a cumulative frequency table for the data and construct a cumulative frequency graph.

c Use your graph to estimate:

i the number of books that have widths of less than 27 cm

ii the widths of the widest 20 books.

3 Measurements of the distances, x mm, between two moving parts inside car engines were recorded and are summarised in the following table. There were 156 engines of type A and 156 engines of type B.

Distance (x mm)	$x < 0.10$	$x < 0.35$	$x < 0.60$	$x < 0.85$	$x < 1.20$
Engine A (cf)	0	16	84	134	156
Engine B (cf)	0	8	52	120	156

a Draw and label two cumulative frequency curves on the same axes.

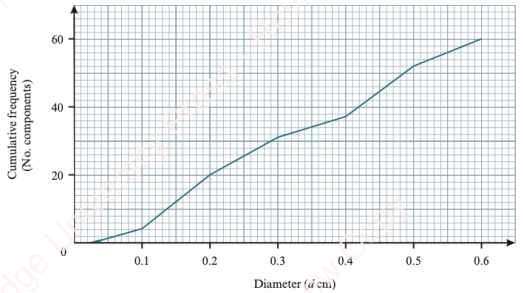
b Use your graphs to estimate:

i the number of engines of each type with measurements between 0.30 and 0.70 mm

ii the total number of engines with measurements that were less than 0.55 mm.

c Both types of engine must be repaired if the distance between these moving parts is more than a certain fixed amount. Given that 16 type A engines need repairing, estimate the number of type B engines that need repairing.

4 The diameters, d cm, of 60 cylindrical electronic components are represented in the following cumulative frequency graph.



a Find the number of components such that $0.2 \leqslant d < 0.4$, and explain why your answer is not an estimate.

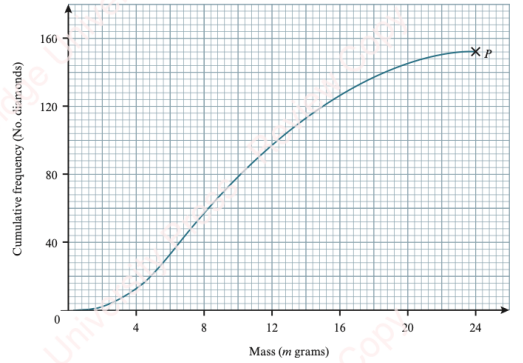
b Estimate the number of components that have:

i a diameter of less than 0.15 cm **ii** a radius of 0.16 cm or more.

c Estimate the value of k, given that 20% of the components have diameters of k mm or more.

d Give the reason why 0.1–0.2 cm is the modal class.

5 The following cumulative frequency graph shows the masses, m grams, of 152 uncut diamonds.



a Estimate the number of uncut diamonds with masses such that:

i $9 \leqslant m < 17$ **ii** $7.2 \leqslant m < 15.6$.

b The lightest 40 diamonds are classified as small. The heaviest 40 diamonds are classified as large. Estimate the difference between the mass of the heaviest small diamond and the lightest large diamond.

Homework

Page 16 – Exercise 1C

c The point marked at $P(24, 152)$ on the graph indicates that the 152 uncut diamonds all have masses of less than 24 grams.

Each diamond is now cut into two parts of equal mass. Assuming that there is no wastage of material, write down the coordinates of the point corresponding to P on a cumulative frequency graph representing the masses of these cut diamonds.

6 The densities, d g/cm^3, of 125 chemical compounds are given in the following table.

Density (d g/cm^3)	$d < 1.30$	$d < 1.38$	$d < 1.42$	$d < 1.58$	$d < 1.70$
No. compounds (cf)	32	77	92	107	125

Find the frequencies a, b, c and d given in the table below.

Density (d g/cm^3)	0–	1.30–	1.38–	1.42–1.70
No. compounds (f)	a	b	c	d

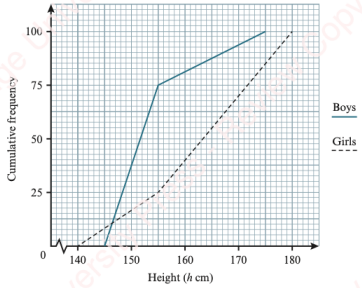
7 The daily journey times for 80 bank staff to get to work are given in the following table.

Time (t min)	$t < 10$	$t < 15$	$t < 20$	$t < 25$	$t < 30$	$t < 45$	$t < 60$
No. staff (cf)	3	11	24	56	68	76	80

a How many staff take between 15 and 45 minutes to get to work?

b Find the exact number of staff who take $\frac{x+y}{2}$ minutes or more to get to work, given that 85% of the staff take less than x minutes and that 70% of the staff take y minutes or more.

8 A fashion company selected 100 12-year-old boys and 100 12-year-old girls to audition as models. The heights, h cm, of the selected children are represented in the following graph.



a What features of the data suggest that the children were not selected at random?

b Estimate the number of girls who are taller than the shortest 50 boys.

c What is the significance of the value of h where the graphs intersect?

d The shortest 75 boys and tallest 75 girls were recalled for a second audition. On a cumulative frequency graph, show the heights of the children who were not recalled.

9 The following table shows the ages of the students currently at a university, given by percentage. Ages are rounded down to the number of whole years.

Age (years)	<18	18–19	20–21	22–24	25–28	29–35	36–44
Students (%)	0	27	51	11	5	4	2

a Represent the data in a percentage cumulative frequency polygon.

b The oldest 8% of these students qualify as 'mature'. Use your polygon to estimate the minimum age requirement for a student to be considered mature. Give your answer to the nearest month.

c Of the 324 students who are 18–19 years old, 54 are not expected to find employment within 3 months of finishing their course.

i Calculate an estimate of the number of current students who are expected to find employment within 3 months of finishing their course.

ii What assumptions must be made to justify your calculations in part **c i**? Are these assumptions reasonable? Do you expect your estimate to be an overestimate or an underestimate?

10 The distances, in km, that 80 new cars can travel on 1 litre of fuel are shown in the table.

Distance (km)	4.4–	6.6–	8.8–	12.1–	15.4–18.7
No. cars (f)	5	7	52	12	4

These distances are 10% greater than the distances the cars will be able to travel after they have covered more than 100 000 km.

Estimate how many of the cars can travel 10.5 km or more on 1 litre of fuel when new, but not after they have covered more than 100 000 km.

11 A small company produces cylindrical wooden pegs for making garden chairs. The lengths and diameters of the 242 pegs produced yesterday have been measured independently by two employees, and their results are given in the following table.

Length (l cm)	$l < 1.0$	$l < 2.0$	$l < 2.5$	$l < 3.0$	$l < 3.5$	$l < 4.0$	$l < 4.5$
No. pegs (cf)	0	0	8	40	110	216	242

Diameter (d cm)	$d < 1.0$	$d < 1.5$	$d < 2.0$	$d < 2.5$	$d < 3.0$
No. pegs (cf)	0	60	182	222	242

a On the same axes, draw two cumulative frequency graphs: one for lengths and one for diameters.

b Correct to the nearest millimetre, the lengths and diameters of n of these pegs are equal. Find the least and greatest possible value of n.

c A peg is acceptable for use when it satisfies both $l \geqslant 2.8$ and $d < 2.2$.

Explain why you cannot obtain from your graphs an accurate estimate of the number of these 242 pegs that are acceptable. Suggest what the company could do differently so that an accurate estimate of the proportion of acceptable pegs could be obtained.