Chapter 1
Representation of data

- Display numerical data in stem-and-leaf diagrams, histograms and cumulative frequency graphs.
- Interpret statistical data presented in various forms.
- Select an appropriate method for displaying data.

1.1 Types of data and 1.2 Representation of discrete data: stem-and-leaf diagrams

WORKED EXAMPLE 1.1

Correct to the nearest centimetre, the length of each of the 80 pencils in a box is 18 cm.

a State the lower boundary and the upper boundary of the length of a pencil.

b What is the least possible total length of all the 80 pencils together?

c The 80 pencils are shared by six children so that each receives an odd number, and no two children receive the same number of pencils. Draw an ordered stem-and-leaf diagram showing one of the possibilities for the number of pencils given to the children.

Answer

a Lower boundary = 17.5 cm
Upper boundary = 18.5 cm

Lengths rounded to 18 cm mean $17.5 \leqslant \text{length} < 18.5$ cm.

b $80 \times 17.5 = 1400$ cm

c

0	1 5 9
1	3
2	3 9

Key: **1**|3 represents 13 pencils

The diagram must show six different, ordered odd numbers with a sum of 80. One possible solution, shown here, is to use 1, 5, 9, 13, 23 and 29.

EXERCISE 1A

1 Sara has collected three sets of data from the children in her daughter's class at school. These are: A: their first names; B: their heights; C: their shoe sizes. Match each set of data with the one word from the following list that best describes it.

X: discrete Y: qualitative Z: continuous

2 **a** Correct to the nearest ten metres, the perimeter of a rectangular football pitch is 260 metres. Complete the following inequality which shows the lower and upper boundaries of the actual perimeter:

…… m $\leqslant$ perimeter $<$ …… m

b Eliana has 16 coins. The mass of each coin, correct to 1 decimal place, is 2.4 grams. Find the least possible total mass of the 16 coins.

3 A car was driven a distance of 364 km in five hours. The distance driven is correct to the nearest kilometre and the time taken is correct to the nearest hour.

Find the lower boundary and the upper boundary of the average speed of the car.

4 The numbers of items purchased by the first 11 customers at a shop this morning were 6, 2, 13, 5, 1, 7, 2, 11, 16, 20 and 15.

a Display these data in a stem-and-leaf diagram and include an appropriate key.

b Find the number of items purchased by the 12th customer, given that the first 12 customers at the shop purchased a total of 111 items.

5 **a** Correct to the nearest metre, the length and diagonal of a rectangular basketball court measure 18 m and 23 m, respectively. Calculate the lower boundary of the width of the court, correct to the nearest 10 cm.

b Correct to the nearest 0.01 cm, the radius of a circular coin is 0.94 cm. Find the least number of complete revolutions that the coin must be turned through, so that a point on its circumference travels a distance of at least 9.5 metres.

6 Bobby counts the number of grapes in 12 bunches that are for sale in a shop. To illustrate the data, he first produces the following diagram.

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

a State whether the data are:

i qualitative or quantitative

ii discrete or continuous.

b Complete Bobby's work by ordering the data in a stem-and-leaf diagram and adding a key.

7 Construct stem-and-leaf diagrams for the following data sets.

a The speeds, in kilometres per hour, of 20 cars, measured on a city street:

41, 15, 4, 27, 21, 32, 43, 37, 18, 25, 29, 34, 28, 30, 25, 52, 12, 36, 6, 25

b The times taken, in hours (to the nearest tenth), to carry out repairs to 17 pieces of machinery:

0.9, 1.0, 2.1, 4.2, 0.7, 1.1, 0.9, 1.8, 0.9, 1.2, 2.3, 1.6, 2.1, 0.3, 0.8, 2.7, 0.4

M **8** The contents of 30 medium-size packets of soap powder were weighed and the results, in kilograms correct to 4 significant figures, were as follows.

1.347	1.351	1.344	1.362	1.338	1.341	1.342	1.356	1.339	1.351
1.354	1.336	1.345	1.350	1.353	1.347	1.342	1.353	1.329	1.346
1.332	1.348	1.342	1.353	1.341	1.322	1.354	1.347	1.349	1.370

a Construct a stem-and-leaf diagram for the data.

b Why would there be no point in drawing a stem-and-leaf diagram for the data rounded to 3 significant figures?

PS **9** Two films are shown on screen A and screen B at a cinema each evening. The numbers of people viewing the films on 12 consecutive evenings are shown in the back-to-back stem-and-leaf diagram.

Screen A (12)		Screen B (12)
	3	7
8 3	4	
7 6 4 0	5	3 4
7 4 1	6	4 5 6 7 8
9 2	7	1 6 8 9

Key: 1|6|4 represents 61 viewers for A and 64 viewers for B

A second stem-and-leaf diagram (with rows of the same width as the previous diagram) is drawn showing the total number of people viewing films at the cinema on each of these 12 evenings. Find the least and greatest possible number of rows that this second diagram could have.

TIP

On the evening when 30 people viewed films on screen A, there could have been as few as 37 or as many as 79 people viewing films on screen B.

10 The masses, to the nearest 0.1 g, of 30 Yellow-rumped and 30 Red-fronted Tinkerbirds were recorded by Biology students. Their results are given in the tables below:

Yellow-rumped					
17.0	13.1	15.7	14.7	17.4	15.3
14.2	16.2	16.9	16.8	16.5	14.2
15.5	16.9	13.5	18.1	17.6	17.9
13.0	15.1	18.2	17.8	18.1	12.3
12.2	17.7	16.5	16.7	14.8	16.6

Red-fronted					
18.2	14.7	15.9	14.7	15.2	14.5
17.3	13.2	14.0	20.2	17.5	15.6
19.3	19.4	17.4	16.5	16.8	15.8
15.3	13.3	16.0	18.6	13.6	18.0
14.1	15.5	20.0	13.9	16.8	14.7

a Display the masses in a back-to-back stem-and-leaf diagram.

b How many Yellow-rumped Tinkerbirds are heavier than the lightest 80% of the Red-fronted Tinkerbirds?

c The students decide to display the birds' masses, correct to the nearest 5 grams, in two bar charts.

i Write down the frequencies for the three classes of Yellow-rumped Tinkerbirds.

ii Explain why it is possible for the 20 g class of Red-fronted Tinkerbirds to have the same frequency as the 20 g class of Yellow-rumped Tinkerbirds.

iii Given that there are equal numbers of Tinkerbirds in the two 20 g classes, construct two bar charts on the same axes showing the masses to the nearest 5 g.

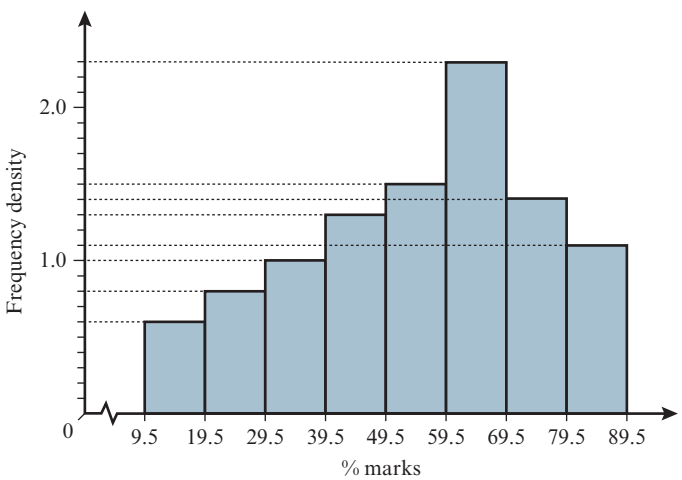
1.3 Representation of continuous data: histograms

WORKED EXAMPLE 1.2

The percentage marks scored by 100 candidates in an examination are shown in equal-width intervals in the following histogram.

The marks are to be regrouped in four unequal-width intervals, as shown in the table.

Marks (%)	f	Frequency density
$9.5 \leqslant x < 19.5$	6	$6 \div 10 = 0.6$
$19.5 \leqslant x < 39.5$	a	
$39.5 \leqslant x < 69.5$	b	
$69.5 \leqslant x < 89.5$	c	



a Find the frequencies a, b and c.

b By calculating the three missing frequency densities, illustrate these data in a histogram with four unequal-width intervals.

Answer

a

Marks (%)	Column area	No. candidates (f)
$9.5 \leqslant x < 19.5$	$10 \times 0.6 = 6$	6
$19.5 \leqslant x < 29.5$	$10 \times 0.8 = 8$	8
$29.5 \leqslant x < 39.5$	$10 \times 1.0 = 10$	10
$39.5 \leqslant x < 49.5$	$10 \times 1.3 = 13$	13
$49.5 \leqslant x < 59.5$	$10 \times 1.5 = 15$	15
$59.5 \leqslant x < 69.5$	$10 \times 2.3 = 23$	23
$69.5 \leqslant x < 79.5$	$10 \times 1.4 = 14$	14
$79.5 \leqslant x < 89.5$	$10 \times 1.1 = 11$	11
	Total area = 100	$\Sigma f = 100$

Column area $\propto$ class frequency, and we know that the sum of the frequencies is 100. This allows us to draw up a grouped frequency table, which corresponds with the original histogram.

$a = 8 + 10 = 18$

$b = 13 + 15 + 23 = 51$

$c = 14 + 11 = 25$

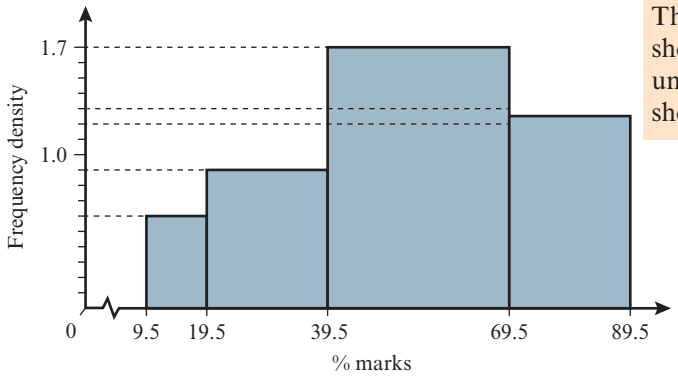
Add together the relevant frequencies.

b For $19.5 \leqslant x < 39.5$, $fd = \dfrac{18}{39.5 - 19.5} = \dfrac{18}{20} = 0.9$

For $39.5 \leqslant x < 69.5$, $fd = \dfrac{51}{69.5 - 39.5} = \dfrac{51}{30} = 1.7$

For $69.5 \leqslant x < 89.5$, $fd = \dfrac{25}{89.5 - 69.5} = \dfrac{25}{20} = 1.25$

Frequency density (fd) $= \dfrac{\text{class frequency}}{\text{class width}}$



The required histogram showing the marks in four unequal-width intervals is shown.

EXERCISE 1B

1 The speeds, in km h^{-1}, of 200 vehicles travelling on a highway were measured by a radar device. The results are summarised in the following table. Draw a histogram to illustrate the data.

Speed	45–	60–	75–	90–	105–	120 or more
Frequency	12	32	56	72	20	8

2 The mass of each of 60 pebbles collected from a beach was measured. The results, correct to the nearest gram, are summarised in the following table. Draw a histogram of the data.

Mass	5–9	10–14	15–19	20–24	25–29	30–34	35–44
Frequency	2	5	8	14	17	11	3

3 Thirty calls made by a telephone saleswoman were monitored. The lengths in minutes, to the nearest minute, are summarised in the following table.

Length of call	0–2	3–5	6–8	9–11	12–15
No. calls	17	6	4	2	1

a State the boundaries of the first two classes.

b Illustrate the data with a histogram.

4 The haemoglobin levels in the blood of 45 hospital patients were measured. The results, correct to 1 decimal place and ordered for convenience, are as follows.

9.1 10.1 10.7 10.7 10.9 11.3 11.3 11.4 11.4 11.4 11.6 11.8 12.0 12.1 12.3
12.4 12.7 12.9 13.1 13.2 13.4 13.5 13.5 13.6 13.7 13.8 13.8 14.0 14.2 14.2
14.2 14.6 14.6 14.8 14.8 15.0 15.0 15.0 15.1 15.4 15.6 15.7 16.2 16.3 16.9

a Form a grouped frequency table with eight classes.

b Draw a histogram of the data.

5 The table shows the age distribution, in whole numbers of years, of the 200 members of a chess club.

Age	16–19	20–29	30–39	40–49	50–59	over 59
No. members	12	40	44	47	32	25

a Form a table showing the class boundaries and frequency densities.

b Draw a histogram of the data.

6 The masses, w kilograms, of a selection of camera lenses are given in the following table.

Mass (w kg)	0.7–	1.2–	1.6–	1.9–	2.1–2.3
No. lenses (f)	24	56	75	36	9

a Give the masses of:

i the lightest 80 lenses

ii the heaviest 22.5% of the lenses.

b Illustrate the data in a fully labelled histogram.

c Estimate the number of lenses with masses of more than 2.0 kg.

7 The following table shows the lengths, to the nearest centimetre, of some chopsticks. Some of the class densities and some of the column heights that will be used to represent the data in a histogram are also shown.

Lengths (cm)	10–12	13–16	17–21	22–27
No. chopsticks (f)	12	24	15	c
Class density	4	6	b	2.5
Column height (cm)	8	a	6	d

Find the values of a, b, c and d.

PS **8** A histogram has been drawn with four columns that have unequal-width intervals. The areas of the four columns are, from left-to-right, 24, 44, 32 and 14 units^2.

a Find the total frequency of the data, given that the first column represents a frequency of 84.

b Explain why it would not be possible for the first column to represent a frequency of 54.

M **9** The blood glucose levels, in mmol/litre, of patients who attended a hospital during a particular week were recorded and are shown to the nearest tenth of a unit in the table.

Blood glucose (mmol/litre)	4.3–4.7	4.8–5.0	5.1–5.6	5.7–p	q–7.5
No. patients	35	27	66	45	r

a The data have a set of frequency densities that are symmetrical. Find the values of p, q and r.

b Of these patients, 10% are advised to take a new medication that may help to reduce their blood glucose levels. Without drawing a histogram, calculate an estimate of the range of these patients' current blood glucose levels, giving both boundary values correct to the nearest hundredth of a unit.

1.4 Representation of continuous data: cumulative frequency graphs

WORKED EXAMPLE 1.3

The distances, x km, covered by 250 delivery vehicles during the past week are summarised in the following cumulative frequency table.

Distance (x km)	<500	<800	<1200	<2000	<2400	<3000	<4000
No. vehicles (cf)	50	80	120	200	210	225	250

a What proportion of the vehicles covered distances from 1200 to 3000 km?

b Without drawing, show that a cumulative frequency polygon used to represent these data consists of just two straight-line segments.

Answer

a $\frac{105}{250} \times 100 = 42\%$

225 − 120 = 105 vehicles

b (0, 0) to (500, 50), $m = \frac{50 - 0}{500 - 0} = 0.1$

(500, 50) to (800, 80), $m = \frac{80 - 50}{800 - 500} = 0.1$

(800, 80) to (1200, 120), $m = \frac{120 - 80}{1200 - 800} = 0.1$

(1200, 120) to (2000, 200), $m = \frac{200 - 120}{2000 - 1200} = 0.1$

(2000, 200) to (2400, 210), $m = \frac{210 - 200}{2400 - 2000} = 0.025$

(2400, 210) to (3000, 225), $m = \frac{225 - 210}{3000 - 2400} = 0.025$

(3000, 225) to (4000, 250), $m = \frac{250 - 225}{4000 - 3000} = 0.025$

Investigate the gradient, m, between pairs of consecutively plotted points.

The polygon is made of one straight-line segment from (0, 0) to (2000, 200) and another from (2000, 200) to (4000, 250).

Write a brief conclusion to explain what the working shows about the polygon.

EXERCISE 1C

1 Estimates of the age distribution of a country for the year 2030 are given in the following table.

Age	under 16	16 – 39	40 – 64	65 – 79	80 and over
Percentage	14.3	33.1	35.3	11.9	5.4

a Draw a percentage cumulative frequency graph.

b It is expected that people who have reached the age of 60 will be drawing a state pension in 2030. If the projected population of the country is 42.5 million, estimate the number who will then be drawing this pension.

2 The records of the sales in a small grocery store for the 360 days that it opened during the year 2017 are summarised in the following table.

Sales, x (in \$100s)	$x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$
No. days	15	27	64	72
Sales, x (in \$100s)	$5 \leqslant x < 6$	$6 \leqslant x < 7$	$7 \leqslant x < 8$	$x \geqslant 8$
No. days	86	70	16	10

Days for which sales fall below \$325 are classified as 'poor' and those for which sales exceed \$775 are classified as 'good'. With the help of a cumulative frequency graph, estimate the number of poor days and the number of good days in 2017.

3 A company has 132 employees working in its city branch. The distances, x kilometres, that employees travel to work are summarised in the following grouped frequency table.

x (km)	< 5	$5 - 9$	$10 - 14$	$15 - 19$	$20 - 24$	> 24
Frequency	12	29	63	13	12	3

Draw a cumulative frequency graph. Use it to find the number of kilometres below which

a one-quarter of the employees travel to work

b three-quarters of the employees travel to work.

4 At a youth club, 80 girls and 80 boys were asked to estimate the duration of one minute by counting to 60 with their eyes closed. The results are shown in the following table.

Duration of count (seconds)	< 50	< 54	< 58	< 62	< 66
No. girls (cf)	0	10	28	71	80
No. boys (cf)	0	32	52	63	80

a Draw and label a horizontal axis from 50 to 66 seconds using 1 cm for 1 second, and a vertical axis for cumulative frequency from 0 to 80 using 1 cm for 5 units.

b Plot five points for the girls and five points for the boys, then draw and label two cumulative frequency curves.

c Use your graphs to estimate the number of:

i girls who counted to 60 too quickly **ii** boys who counted to 60 too slowly.

5 The cumulative frequency table shows the numbers of stamps, x, in 100 collections.

No. stamps (x)	$x < 40$	$x < 50$	$x < 80$	$x < 120$	$x < 150$	$x < 200$
No. collections (cf)	0	8	19	32	88	100

a How many of the collections contain 50 or more but fewer than 80 stamps?

b Find the least possible value of n, given that each of the largest 68 collections contains more than n stamps.

6 At a warehouse, cardboard is packed into bales for recycling. A full bale has a mass of 25 kg. The cumulative frequency table shows the number of full bales packed in the 52 weeks of last year. The data are to be displayed in a cumulative frequency polygon showing the mass of cardboard packed.

No. full bales	125 – 135	136 – 144	145 – 211	212 – 300
No. weeks (*cf*)	1	12	49	52

a Write down the coordinates of the two plotted points between which the polygon will be steepest.

b By drawing the polygon, or otherwise, estimate the number of weeks during which less than 4500 kg of cardboard was packed for recycling.

PS **7** The data displayed in the following stem-and-leaf diagram show the numbers of customers making purchases at two stores, A and B, over a period of 21 days.

Store A		Store B
9	**2**	5 5 5 6 7 8
3 2	**3**	0 1 3 3 4
7 7 6	**3**	5 5 8 9
4 4 3 1	**4**	0 1 1
9 9 7 7 5	**4**	6 8
4 4 3 2 0 0	**5**	0

Key: 9 | **2** | 5 represents 29 customers at A and 25 at B

a Show these data by drawing two cumulative frequency graphs on a single diagram.

On no two consecutive days during this period did the number of customers making purchases at store A decrease; on no two consecutive days during this period did the number of customers making purchases at store B increase.

b Draw up an ungrouped frequency table showing the numbers of customers who made purchases at both stores A and B on these 21 days.

P **8** **a** Under what conditions will a cumulative frequency graph representing the following set of data be a straight line?

Values of *x*	$x < x_1$	$x < x_2$	$x < x_3$	$x < x_4$	$x < x_5$
cf	0	a	b	c	d

b These data are represented by a straight-line cumulative frequency graph whose equation is $y = mx + c$. State the conditions under which the value of c is:

i negative **ii** non-negative.