

Exercise 1A

1 List which of the following types of data are:

- a) qualitative
- b) discrete
- c) continuous.
 - i. number of emails sent
 - ii. flavour of ice cream
 - iii. mass of kittens
 - iv. test grade
 - v. collar size
 - vi. length of fingers
 - vii. time spent reading
 - viii. petal area
 - ix. membership fees
 - x. gender of cinema-goers
 - xi. age of concert-goers
 - xii. nationality of athletes

2 You are asked to write a report on students' spending on transport. Describe how you might find relevant

- a) primary data
- b) secondary data

Primary data is information you collect yourself.

Secondary data is information that has been collected by someone else.

3

- a) A market research company wants to ask a sample of people from a city about a plan to build a new leisure centre in the city.

What questions might the company want to ask?

- b) Comment on these methods for carrying out this survey.
 - i. Interview people who use the city's train station between 7 a.m. and 9 a.m. one weekday morning.
 - ii. Send questionnaires by email to all residents registered with the local council.
 - iii. Ask a local newspaper to include an article about the leisure centre and invite readers to write in with their views.

4 Some people complete online quizzes to win prize vouchers.

Do you think the results from such quizzes will be reliable?

5 Describe how you would select a representative sample and carry out a survey for

- a) a sample of 80 students from a sixth form to answer questions about library facilities
- b) a sample of 800 people from a city to answer questions about cycle lanes.

6 Describe how you could use quota sampling to ask 400 people in a town questions about building a new tram line near the town.

7

- a) Describe how you could use cluster sampling to find a large sample of people in Scotland to answer questions about public transport.
- b) Would your sample be a random sample of Scotland's population?

8 The table gives the number of students in each year group in a sixth-form college.

Year	Number of students
12	215
13	198

Priya wants to carry out a survey about canteen prices.

She says 'I will ask the first ten people I see in the canteen tomorrow at lunchtime.'

Suggest ways in which Priya can improve her method and describe any problems she may encounter.

9 An optician wants to ask a random sample of patients how satisfied they are with the service they receive from her practice.

Describe how she could select a random sample and what decisions and problems may be involved.

10 An ecologist wants to study the distribution of insects in a random sample of 15 square metres of land from a rectangular field that is 60 metres long and 35 metres wide.

How could she use random numbers to select a sample?

11 A survey is to be carried out in a city to find out what young adults think of the public sports facilities in the city.

Describe a way of using each of these methods and list the advantages and disadvantages of each method.

- a) Simple random sampling
- b) Stratified sampling
- c) Cluster sampling
- d) Quota sampling

12 The table gives the number of part-time and full-time employees of a logistics company.

Number of employees	Male		Female	
	Full-time	Part-time	Full-time	Part-time
Warehouse	312	24	96	78
Depot	64	18	31	12
Office	21	9	38	49
Distribution	22	6	15	0

The director wants a representative sample of 180 employees to complete a questionnaire about workplace wellbeing.

- a) Calculate how many employees from each sub-group are needed.
- b) What other characteristics of the workforce may be relevant?
- c) What problems are likely to arise when gathering information by this method?

13 The table gives the number of entrants to apprenticeship schemes in one year.

Age	18	19	20	Other	Total
Female	38 210	21 640	9 310	17 420	86 580
Male	51 460	28 330	12 070	23 180	115 040
Total	89 670	49 970	21 380	40 600	201 620

Source: based on www.gov.uk apprenticeship statistics

- a) A research body wants to invite 6000 of the apprentices to take part in a survey about their first experiences of apprenticeship schemes. The sample is to be stratified by age and gender.
 - i. Calculate how many apprentices from each sub-group should be invited.
 - ii. List other characteristics of the population that may be relevant.
 - iii. Describe any problems that you think the research body may have and suggest ways these might be overcome.
- b) Use the internet to find out about an annual survey carried out to collect feedback from apprentices about their training. Write a short report.

14 You are asked to carry out a survey in your school or college to find out what students think about introducing a four-day school week.

- a) Describe in detail how you would use stratified sampling to select a random sample of 20% of the students.
- b) Write a list of questions you could ask.
- c) Describe how you would display the results.

Exercise 1B

1 For each dataset find

- a) the mode
- b) the median
- c) the mean
- d) the range.

a) Price of the cheapest annual pass at each theme park in 2024–25.

£189	£142	£165	£210
£175	£198	£156	£133
£220	£119	£171	£140
£165	£205	£158	£150
£148	£260	£128	£215

b) Estimated annual sponsorship earnings, in millions of pounds, of players in a national rugby squad.

£0.6m	£0.4m	£0.7m	£2.1m
£2.4m	£1.3m	£1.5m	£1.1m
£1.1m	£0.3m	£1.6m	£6.8m
£1.2m	£1.8m	£2.2m	£1.7m
£2.7m	£2.6m	£2.3m	

2 Deshi is writing a report about the rates of pay of staff at a distribution centre.

He finds this table which gives the mean rate of pay for pickers and machine operators.

	Number of workers	Mean wage (£ per hour)
Pickers	38	£10.20
Machine operators	11	£14.80

Deshi says that the overall mean must be £12.50 per hour.

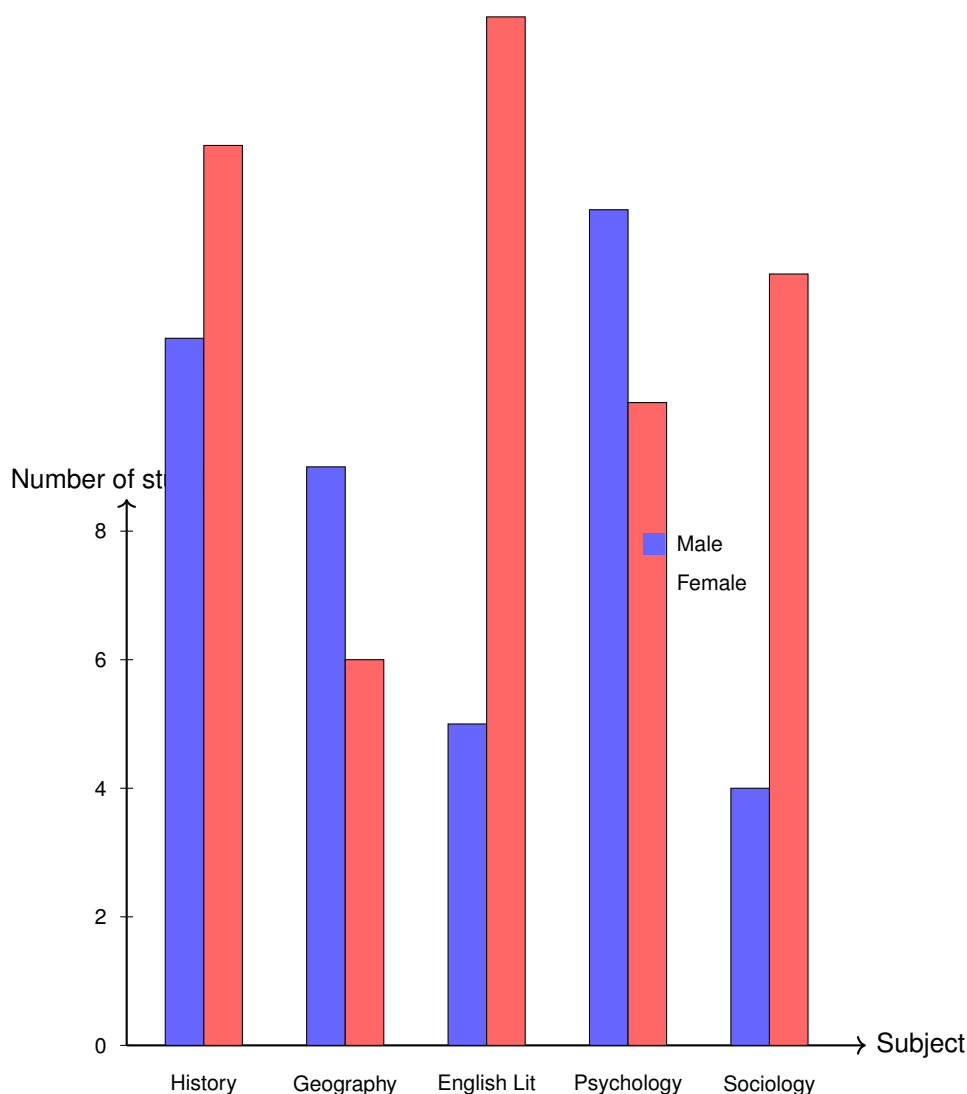
- a) Explain why Deshi is wrong.
- b) Find the overall mean wage.

Start by finding the total wages for the 49 workers.

3 Priya's mean score from 6 matches on the netball team is 18 goals. The coach says she must increase her mean score to 21 goals in the next match to keep her place on the team.

How many goals must Priya score in the next match?

4 The bar chart shows the number of students studying A level humanities subjects at a sixth-form college.



Femi says the average class size is 21, Aisha says it is 18 and Connor says it is 12.

- Explain how they can all be correct.
- Which average do you think best represents the data? Give reasons.

5

- For each dataset find the median and interquartile range.

- Number of minutes the school minibus is late each afternoon in 4 weeks:

1 0 2 0 1 0 23 2 0 1
0 1 0 0 2 1 3 0 1 0

- Number of tomatoes that ripen from 10 grow-bags:

34 31 29 33 35 8 37 30 32 36

- Amount spent on coffee by each student on a college course:

£2.10 £3.85 £2.40 £8.60
£2.30 £2.95 £1.75 £2.20
£3.05

b) Explain how you can tell that the median and interquartile range will give more useful values than the mean and range for these datasets.

6 Find the mean and standard deviation of each dataset.

a) Marks in a test out of 50:

41 28 39 44 26 38 42 30
24 33 40 23

b) Temperatures at midday:

25 °C 19 °C 21 °C 20 °C 21 °C 22 °C 24 °C

c) Journey time to college (in minutes):

33 21 29 26 31 41 24 33 37 25

d) Number of points scored in a game of netball:

29 84 39 0 24 11 71 61 43 39 0 15

7 Which of the datasets in question 6 contain

a) discrete data

b) continuous data?

8 The stem-and-leaf diagram shows the heights of the students in a class.

Heights of students

18	3
17	1 2 5 6 8 9
16	1 2 4 5 7 9
15	0 4 6 9
14	6
13	2

Key 13 | 2 means 132 cm

a) Find three different averages and three different measures of spread.

b) Which average and which measure of spread would you use to represent this dataset? Give reasons.

9 A car park attendant records the time for which each car is left in a long stay car park. The results in minutes are listed below.

38 52 35 12 47 52 15 6 37 23
51 28 30 26 51 53 20 38 53 17
51 40 12 52 7 48 53 41 29 14
39 49 18 52 29 44 36 40 51

a) Draw a stem-and-leaf diagram.

b) Find

- i. the mode
- ii. the range
- iii. the median
- iv. the interquartile range.

c) Use your answers to write a short paragraph about the results.

10 Femi can cycle to college by two different routes. He keeps a record of how long it takes him to get to college by each route for 3 weeks. The back-to-back stem-and-leaf diagram shows his results.

Time taken for journey to college

Route C	Stem	Route D
8 6 4	1	6
9 8 7 5 3	2	1 2 4 6 8 9
8 6 3 2	3	0 1 2 4 5
4 2	4	1

Key 6 | 1 | 6 means Route C takes 16 minutes, Route D takes 16 minutes

a) For each route, find

- i. the mode
- ii. the range
- iii. the median
- iv. the interquartile range.

b) i. Use the stem-and-leaf diagram and your answers to part a) to write a paragraph comparing the times taken using the two routes.

ii. Which route would you advise Femi to take when he sets off from home at 08:10 and needs to get to college by 08:45?

11 The table gives the gender and ages of the people who marry at a registry office during one week.

Mon	Tues	Wed
F24 & M30	F18 & M19	F29 & M34
F37 & M38	F25 & M31	F33 & M40
F21 & M22		M37 & M30

Thurs	Fri	Sat
F32 & M45	F28 & M27	F40 & M36
F23 & M26	F34 & F30	F20 & M25
	F30 & M31	F25 & M34
	F19 & M23	F42 & M47
		F28 & M32
		F36 & M37

Write a short report comparing the ages of the men and women. Include a back-to-back stem-and-leaf diagram, an average and a measure of spread.

Exercise 1C

- 1 The table shows the ages of female and male trainees attending a course.

Age (years)	17	18	19	20	21
Number of females	8	12	9	5	0
Number of males	10	9	7	6	11

For the female and male trainees compare

- a) the modal age
- b) the range of ages
- c) the median age
- d) the mean age.

- 2 The runs scored by Marcus and Devon in this season's cricket matches are listed below.

Marcus: 68 37 9 4 0 17 29 41 24 3 82

Devon: 22 41 23 53 24 11 49 31 21 15 39

- a) Draw two box and whisker plots to compare their scores.
- b) Compare Marcus and Devon's mean score and standard deviation.
- c) There is only one place left on the team for the next match. Who would you pick and why?

- 3 The table gives the annual wages of people who work in a garden centre.

Annual wage	£19 500	£22 000	£64 000
Number of workers	16	5	1

- a) Find the median and interquartile range.
- b) An outlier is defined as a value that is more than 1.5 times the interquartile range below the lower quartile or above the upper quartile. Show that £64 000 is an outlier.
- c) The garden centre manager says that the average wage is £21 200. Comment on the use of this value.

- 4 The table gives the number of car break-ins in two estates last year.

Month	Jan	Feb	Mar	Apr	May	Jun
Elmwood	29	18	14	11	13	9
Fairview	24	22	17	10	9	13

Month	Jul	Aug	Sep	Oct	Nov	Dec
Elmwood	12	6	16	25	32	35
Fairview	7	8	18	20	36	28

a) For each of the estates find

- i. the mean
- ii. the standard deviation
- iii. the median
- iv. the interquartile range.

b) Draw box and whisker plots.

- c)** i. Naomi says that Elmwood is safer from car break-ins than Fairview. Which of the statistics supports her comment?
- ii. Dean says that this is wrong and Fairview is safer than Elmwood. Which of the statistics supports his point of view?

5 A community centre has 12 table-tennis tables. The caretaker records the number of tables used each evening session. The table gives the results for 8 weeks.

Tables in use	12	13	14	15	16	17
No. of evenings	1	14	13	10	3	1

a) Write down

- i. the mode
- ii. the range.

b) Find

- i. the median
- ii. the interquartile range
- iii. Draw box and whisker plots to illustrate the data.

c) Find

- i. the mean
- ii. the standard deviation.

6 The labels on packets of grass seed say 'Average contents: 50 seeds.'

After receiving complaints from customers, a garden centre checks the first 100 packets in their stock. The table gives the results.

Number of seeds	47	48	49	50	51	52	53	54
Number of packets	9	15	22	31	21	0	0	2

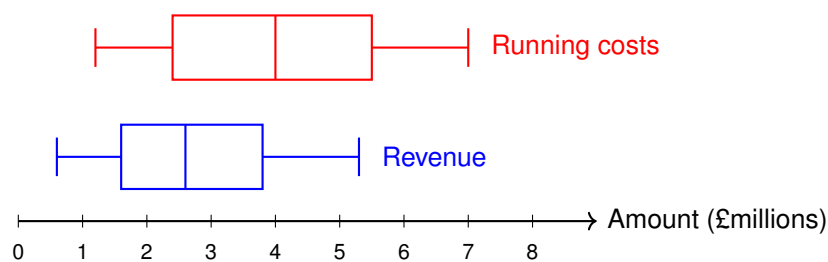
a) i. Use the data to find three averages.

- ii. Write a sentence that the garden centre could use in a letter defending the claim on the packets.

b) An outlier is a data value that is more than 1.5 times the interquartile range below the lower quartile or above the upper quartile. Show that 54 is an outlier.

c) Explain two ways in which the sampling method used by the garden centre could be improved.

7 Lewis draws this box and whisker plot to show the revenue and running costs of Championship rugby clubs in the 2023–24 season.



Compare the main features of these two distributions.

Write a paragraph to describe what the boxplots tell you about the finances of Championship rugby clubs.

8 The table gives the ages of students who go on a language exchange trip to Spain.

Age (years)	12	13	14	15	16	17	18
Number of students	8	11	16	6	12	4	0

- Theo says the modal age is 13 years and the range is 5 years.
 - What errors has Theo made?
 - Find the correct mode and range.
- Olu uses his calculator and says that the mean age is 7.4 years and the standard deviation is 4.2 years.
 - What error has Olu made?
 - Find the correct mean and standard deviation.
- Freya says the median age is 14 years and the interquartile range is 3 years.
 - What error has Freya made?
 - Find the correct median and interquartile range.
 - Draw a box and whisker plot and describe how the ages are distributed.

9 The table gives the number of errors made by call-centre staff before and after a new training scheme.

Number of errors	Number of staff	
	Before training	After training
3	0	4
4	0	9
5	2	12
6	11	9
7	14	4
8	9	0
9	2	0
10	2	0

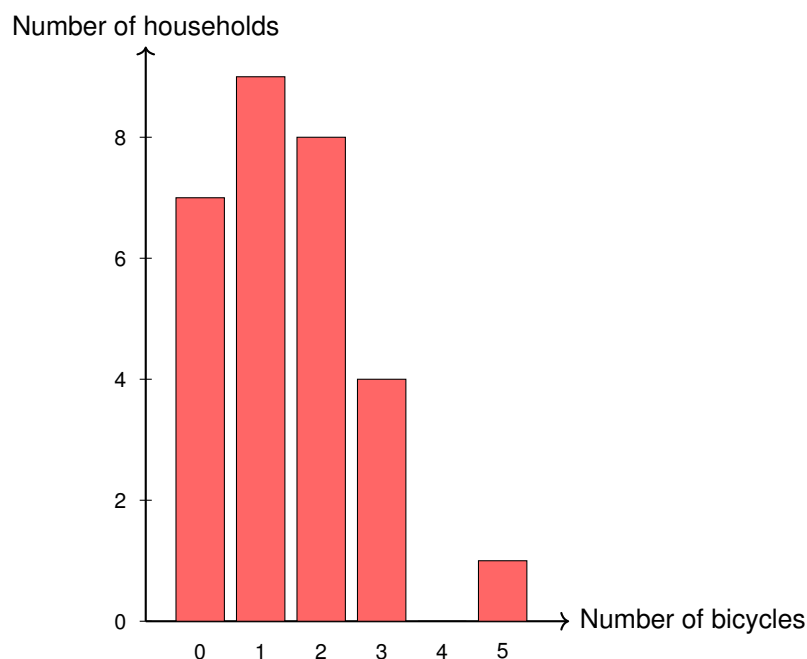
- Draw box and whisker plots to compare the results before and after training.
- Describe what the box and whisker plots show about the effect of the training on the staff's performance.

10 A psychologist carries out a memory test. She shows 12 objects to a group of volunteers. After removing the objects she asks the volunteers to list the ones they remember. A week later she asks them to write another list of the objects they remember. The table gives the results.

Number of items remembered	Number of volunteers	
	On day of test	After 1 week
1	0	2
2	0	2
3	0	4
4	0	4
5	1	5
6	2	4
7	2	2
8	5	1
9	4	0
10	4	0
11	6	0
12	6	0

- Draw box and whisker plots to illustrate these results.
- Find the mean and the standard deviation for both datasets.
- Use your answers to write a paragraph describing the results of the experiment.

11 The chart shows the results of a survey on the number of bicycles owned by households in a street.



- Find the mode, median and mean number of bicycles per household.

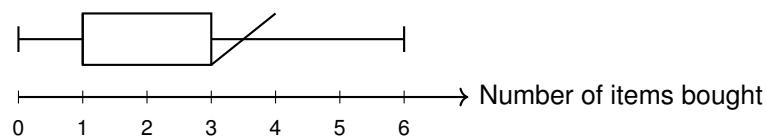
- b) Bukky says the mean is the best average to use. Give a reason why Bukky may say this.
- c) Tariq says the mode and median are both better than the mean. Give a reason why Tariq may say this.
- b) i. What is shown on the bar chart that would not be shown on a box and whisker plot?
 ii. What would be shown on a box and whisker plot that is not shown on the bar chart?

12 Folake and Marni serve customers at a market stall. Their tally chart shows the number of items customers buy on one Saturday.

Number of items	Number of customers
0	_____ (16)
1	_____ (24)
2	_____ (35)
3	_____ (12)
4	_____ (3)
5	_____ (0)
6	_____ (1)

The stall owner asks them to draw a chart to display these results.

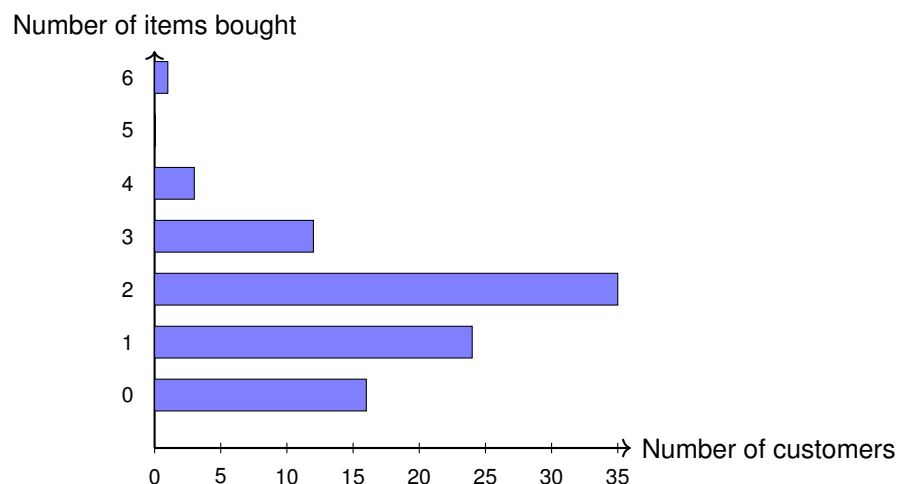
- a) i. Folake draws this box and whisker plot.



Marni says there should be a vertical line across the middle of the box. Is Marni correct? Show how you decide.

- ii. Describe a different error Folake has made.

- b) Marni says this bar chart shows the data better than the box and whisker plot.



- i. Comment on this diagram and describe any improvements that could be made.
- ii. Which of the two diagrams do you think would be most useful to the stall owner? Give a reason.

13 Hana and Jess go running each week. They always end the session with a sprint finish. The table gives their times for the last 32 weeks.

Time taken to sprint 400m (seconds)	Number of weeks	
	Hana	Jess
68	4	0
69	7	5
70	10	13
71	8	9
72	3	5
73	0	0

Hana wants to use a chart to illustrate these results.

- a) Use the data to draw
 - i. two box and whisker plots
 - ii. a bar chart.
- b) Which type of chart would you advise Hana to use? Give reasons.
- b) Jess wants to summarise the data by giving an average and a measure of spread.
Find suitable values for her to use and explain your choice.

14 Some high-street clothing brands are said to use 'vanity sizing'. This means that the measurements of their clothes are more generous than those suggested by British Standards.

The table gives the actual measurements in inches of jeans sold as size 10 and size 12 at 8 high-street shops. The British Standard (BS) sizes are also given in centimetres. (1 inch = 2.54 cm)

Shop	Jeans size 10		Jeans size 12	
	Waist	Hips	Waist	Hips
A	27.5	37.5	29.5	39.5
B	28.0	37.5	30.0	39.0
C	27.5	36.0	29.0	38.0
D	28.5	36.5	30.5	38.5
E	28.5	36.5	30.5	38.5
F	29.0	38.5	31.0	40.5
G	29.0	38.5	30.5	39.8
H	29.5	38.5	32.0	40.5
BS	26.5	35.5	28.0	37.0

Write a short report about vanity sizing. Use statistical measures to support your findings. Include diagrams if you wish.

15

- a)** Ask all members of your class to estimate one or more of the following:
- i.** the number of messages they send in a day
 - ii.** the time they spend on social media in a week
 - iii.** the distance they travel to college in a year
 - iv.** the amount they spend on streaming subscriptions in a year.
- b)** Use averages, measures of spread and/or statistical diagrams to summarise the results. Compare your estimates with those from the whole class.

Exercise 1D

1 An allotment holder records the mass of pumpkins grown by members on each weekend during a growing season. The table gives the results.

The mass of 5 kg is included in the first group, not the second.

Mass m (kg)	Number of weekends
$0 < m \leq 5$	2
$5 < m \leq 10$	5
$10 < m \leq 15$	14
$15 < m \leq 20$	24
$20 < m \leq 25$	17
$25 < m \leq 35$	4

- a) Write down the modal class.
- b) Draw a cumulative frequency graph.
- c) Use your graph to find
 - i. the median
 - ii. the lower quartile
 - iii. the upper quartile
 - iv. the interquartile range
 - v. the 30th percentile
 - vi. the 70th percentile.
- d) Find an estimate for
 - i. the mean
 - ii. the standard deviation.

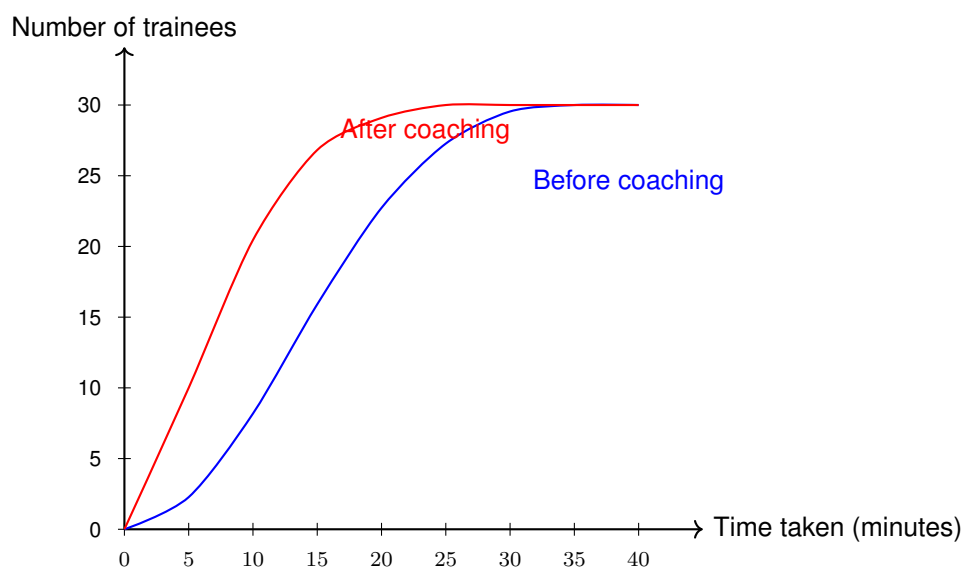
2 An athletics club records the number of spectators who attend each home meeting. The table gives the results for one season.

Attendance	No of meetings
Under 200	1
200–400	3
400–600	11
600–800	21
800–1200	6
1200–1600	2
over 1600	0

- a) Write down the modal class.

- b) i. Draw a cumulative frequency graph.
 ii. Use your graph to find the median and interquartile range.
 iii. The organiser says that over 8% of the meetings have had an attendance of over 1000. Does the graph confirm this? Show how you decide.
- c) Find an estimate for
 i. the mean
 ii. the standard deviation.
- d) The organiser says that last season the mean was 720 and the standard deviation was 240. How have attendances changed since last season?

3 After trainee chefs were timed on preparing a dish, they were given extra coaching, then timed again. The cumulative frequency graph shows the results.



- a) Find the median and interquartile range of the times before and after the coaching.
 b) Write a paragraph to describe the effect the coaching has had on the trainees' performance.

4 Students on a childcare course weighed the 4-year-old children at a nursery. The table shows their results with masses given to the nearest 0.1 kg.

Mass (kg)	No of children
13.0–13.9	3
14.0–14.9	6
15.0–15.9	12
16.0–16.9	14
17.0–17.9	7
18.0–18.9	5
19.0–19.9	2

- a) Find estimates for the mean mass and standard deviation for this group.
 b) i. Draw a cumulative frequency diagram.

- ii. A childcare book gives the following percentile masses for 4-year-old children.

Percentile	5	10	25	50	75	90	95
Mass (kg)	13.9	14.4	15.2	16.2	17.4	18.5	19.2

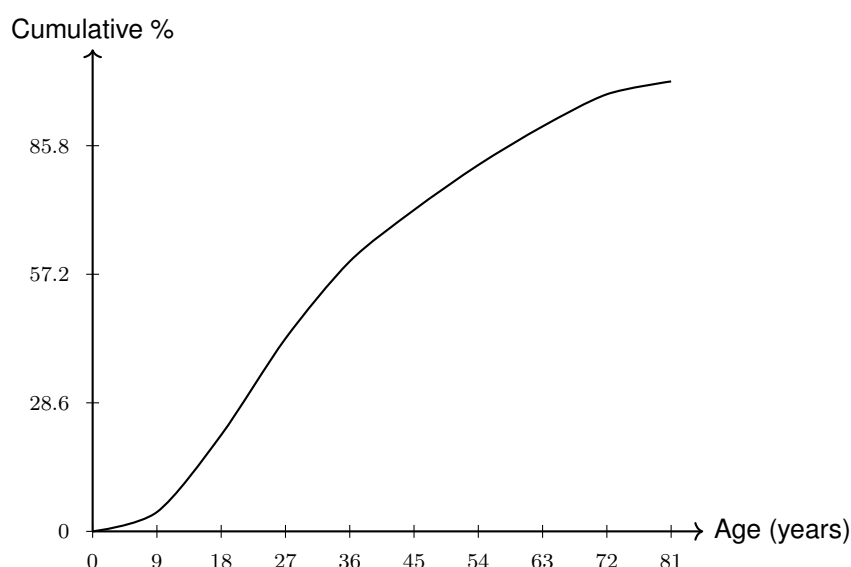
Compare values from your graph with those given in this table.

- 5 A survey asks students how long they spent gaming and how long they spent reading on one day.

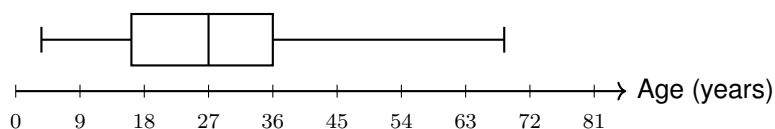
Time (min)	Number of students	
	Gaming	Reading
0–29	7	0
30–59	22	19
60–89	21	25
90–119	9	6
120–149	1	4
150–179	0	1

- a) Write down the modal class for each dataset.
- b) i. Compare estimates for the mean and standard deviation of the datasets.
 ii. Describe what these measures tell you about the way the students spend their time.
- c) i. Use cumulative frequency graphs to find the median and interquartile range for each dataset.
 ii. Do these measures support your comments in b) ii?

- 6 These diagrams both show the results of a survey about the ages of people who use a video-sharing app.



Ages of users of a video-sharing app

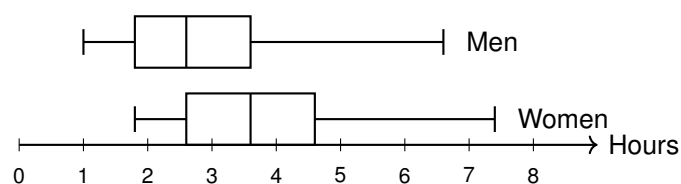


- a) Describe the main features of this distribution.
- b) i. Priya says that a cumulative frequency graph shows the results better than a box and whisker plot. Give a reason why Priya might say this.
- ii. Connor disagrees. He says that a box and whisker plot illustrates the data in a better way. Give a reason why Connor might say this.
- c) What information about the survey is not given on either of the diagrams?

7 Each year thousands of people raise money for charity by taking part in a national cycling event. The table gives the times taken by women and men in one year.

Time taken t (hours)	Frequency	
	Women	Men
$1 < t \leq 1\frac{1}{2}$	9	48
$1\frac{1}{2} < t \leq 2$	58	1102
$2 < t \leq 2\frac{1}{2}$	612	2890
$2\frac{1}{2} < t \leq 3$	1740	4215
$3 < t \leq 3\frac{1}{2}$	2304	3580
$3\frac{1}{2} < t \leq 4$	2510	2790
$4 < t \leq 4\frac{1}{2}$	1690	1480
$4\frac{1}{2} < t \leq 5$	1080	740
$5 < t \leq 6$	870	510
$6 < t \leq 7$	165	95
$7 < t \leq 8$	1	1

- a) i. Draw a cumulative frequency graph to illustrate these results.
- ii. Use your graph to compare the performance of the women and the men.
- b) The box and whisker plots show the results from the previous year's event.



- i. Use these box and whisker plots to compare the performance of the women and the men in the previous year.
- ii. Compare the performance of the women and men in the previous year and this year.
- 8 In an investigation about part-time work Sami and Lola find this information on the internet.

Percentile	Number of hours worked		Rate of pay (£ per hour)	
	Male	Female	Male	Female
10	4.8	6.1	7.20	7.20
20	8.0	9.9	7.35	7.55
25	9.2	12.0	7.50	7.85
30	10.8	14.0	7.80	8.10
40	14.3	15.9	8.35	8.78
50	17.1	18.6	9.12	9.65
60	19.4	20.5	10.40	10.83
70	22.3	22.7	12.55	12.55
75	23.3	23.8	14.25	14.00
80	24.5	24.7	17.10	16.05
90	28.1	27.8	25.85	22.10
Mean	16.8	17.7	13.25	12.70
No of jobs (thousands)	1750	5120	1750	5120

- a) Sami says ‘On average, the men get over 50 pence more per hour than the women.’
Lola says ‘No – on average, the women get 45 pence more per hour than the men.’
- Explain each person’s statement.
 - Give a reason why this difference may occur.
- b) Compare the interquartile ranges of rates of pay.
- c) Write a paragraph to compare the hours worked by men and women who have part-time jobs.
- d) Why is it difficult to draw box and whisker plots for this data?

9 The table gives the ages on first marriage for people in a small country, comparing two decades.

Age (years)	Number of males		Number of females	
	1990	2020	1990	2020
16–19	15 200	720	52 100	2 450
20–24	108 400	16 300	116 700	27 900
25–29	58 300	50 800	32 700	59 600
30–39	24 200	68 100	11 900	55 200
40–49	3 400	15 700	1 900	10 100
50–59	1 300	2 800	900	1 900
60–69	480	720	410	330
70 and over	100	160	130	80

- a) Calculate an estimate for the mean and the standard deviation for each dataset. State any assumptions you make.
- b) Find the median and interquartile range for each dataset.
- c) Describe what your results tell you about changes between the ages at which people married in 1990 and 2020.
- d) A sociologist wants to carry out a survey to find out why people decide to marry at particular ages. Suggest a sampling technique that could be used to find a representative sample of newlyweds and describe any problems that may arise in carrying out this survey.

10 Use the internet to find population data for a country of your choice. Write a short report comparing the age distribution of males and females, or the age distribution for different years. Include statistical measures and diagrams to support your comments.

11 A travel company has downloaded 40 years of weather data for two seaside resorts in July from the internet.

Hours of sunshine per month

Hours of sunshine (h)	Number of months	
	Brightside	Mistport
$100 < h \leq 140$	0	2
$140 < h \leq 160$	1	9
$160 < h \leq 180$	3	9
$180 < h \leq 200$	4	8
$200 < h \leq 220$	8	7
$220 < h \leq 240$	11	4
$240 < h \leq 280$	9	1
$280 < h \leq 320$	3	0
$320 < h \leq 360$	1	0

Rainfall per month

Rainfall (mm)	Number of months	
	Brightside	Mistport
$0 < x \leq 10$	2	0
$10 < x \leq 20$	4	1
$20 < x \leq 30$	5	2
$30 < x \leq 40$	7	7
$40 < x \leq 50$	6	7
$50 < x \leq 60$	1	5
$60 < x \leq 70$	2	5
$70 < x \leq 80$	3	2
$80 < x \leq 100$	6	6
$100 < x \leq 120$	3	3
$120 < x \leq 140$	1	2

- Write a paragraph comparing the weather at these two resorts in July. Use statistical diagrams and measures to support your comments.
- Find weather data for a resort of your choice and compare the results with those for Brightside and Mistport.

Exercise 1E

1 The table gives the amounts spent by customers at a department store on one day.

The last class has been assumed to have the same width as the previous class.

Amount (£ P)	No. of customers	Lower boundary	Upper boundary	Width	Freq. density
$P < 20$	64	0	20	20	$64 \div 20 =$
$20 \leq P < 40$	88	20	40	20	$88 \div 20 =$
$40 \leq P < 80$	142	40	80	40	$142 \div 40 =$
$80 \leq P < 120$	71	80	120		
$120 \leq P < 200$	19	120	200		
$P \geq 200$	6	200	280		

a) Copy and complete the table. State any assumptions you make.

b) Draw a histogram to illustrate the data.

2 The table gives estimates for the age distribution of a region's population in one year.

Age (years)	No. of people (1000s)	Lower boundary	Upper boundary	Width	Freq. density
0–4	3650	0	5	5	$3650 \div 5 = 730$
5–15	10 230	5	16	11	$10\,230 \div 11 = 930$
16–44	19 870	16	45	29	
45–64	15 110				
65–74	5 640				
75–89	4 120				
90+	480				

a) Copy and complete the table. State any assumptions you make.

b) Draw a histogram to illustrate the age distribution.

3 Reading age is a measure of literacy. All students are tested when they join Hartfield College. The table gives this year's results.

Reading age	Number of students
$10 < x \leq 12$	18
$12 < x \leq 14$	64
$14 < x \leq 16$	121
$16 < x \leq 18$	115
$18 < x \leq 20$	97
$20 < x \leq 24$	71
$24 < x \leq 30$	14

a) i. Calculate an estimate for the mean reading age and the standard deviation for this year's students.

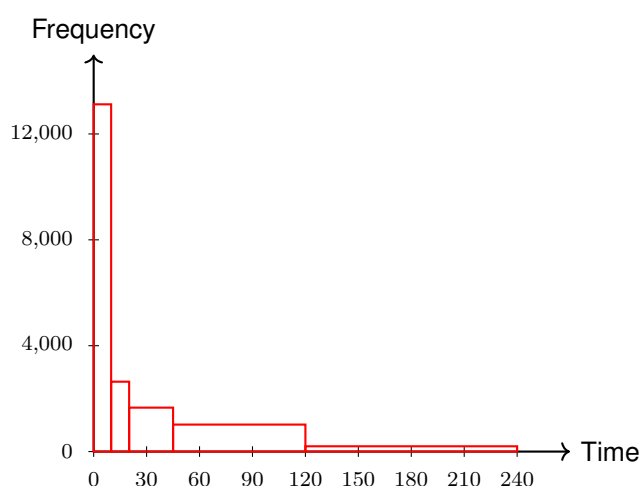
- ii. Last year's students had a mean reading age of 17.2 and the standard deviation was 3.6.
Write two statements comparing the results for this year's students with last year's.

- b) i. Draw a histogram for the data for this year's students.
ii. Use your histogram to estimate the percentage of this year's students that have a reading age of more than 17.

4 Mia finds this data for train delays at a city station in one month.

Number of minutes late	Number of trains
Early–10	9840
11–20	1980
21–45	1240
46–120	760
121–240	145
> 240	18

Mia draws this histogram to illustrate the data.



- a) i. Give reasons why this histogram is incorrect.
ii. Use the data to draw a correct histogram. Describe any problems and assumptions you make.
- b) i. Use the data to estimate the mean and standard deviation.
ii. Explain why the values you have found are estimates.

5 A council records the widths of the fire exit doors in all of its public buildings. The table gives the data with widths rounded to the nearest 10 mm.

Door width (mm)	Number of doors
Less than 700	0
700–740	19
750–840	168
850–940	132
950–1040	54
1050 and over	16

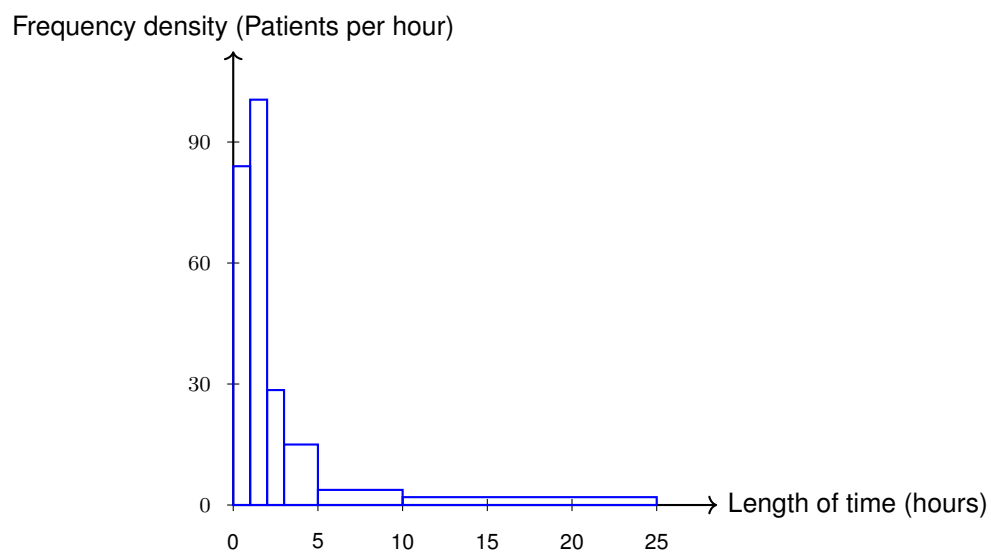
Official guidelines say that fire exit door widths must be at least 725 mm.
The council wants to increase all fire exit door widths to at least 800 mm.

Take care with class boundaries here. The 700–740 group starts at 695, but ends at 745. Can you see why?

Use a histogram to estimate:

- a)** the number of doors that do not comply with the official guidelines
- b)** the percentage of the doors that the council wants to widen.

6 The histogram shows the times that patients spent at a clinic's minor injuries unit during one day.



The target for clinics is for 95% of patients to be seen in under 4 hours.

The average time for clinics nationally is 110 minutes.

Use data from the histogram to compare this clinic's performance with these statistics.

You could draw a grouped frequency table for the data.

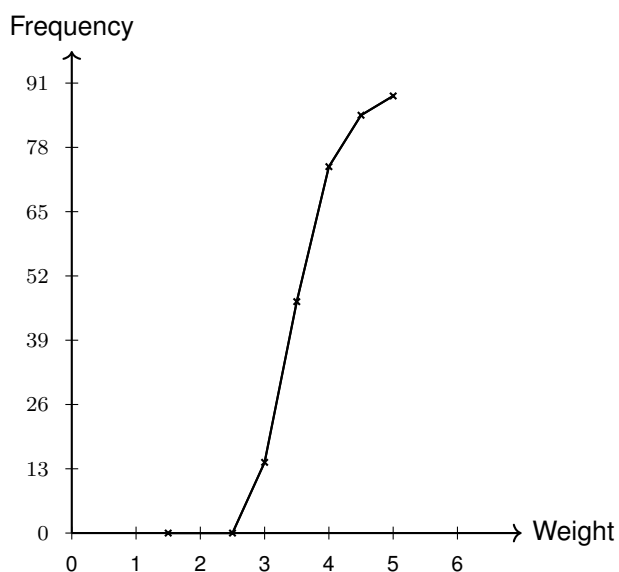
Exercise 1F

1 A tutor on a childcare course gives her students this table which shows the weights of newborn babies born in one month at a local hospital.

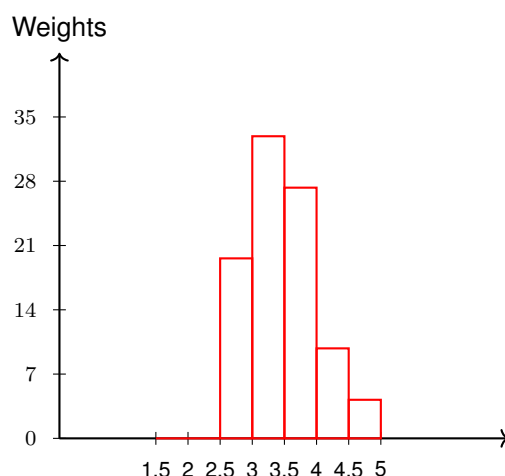
The tutor asks for a statistical diagram to show this information.

Weight (kg)	No of babies
$1.5 < x \leq 2.5$	0
$2.5 < x \leq 3$	14
$3 < x \leq 3.5$	33
$3.5 < x \leq 4$	27
$4 < x \leq 4.5$	10
$4.5 < x \leq 5$	4

Bea draws this cumulative frequency graph.



Theo draws this histogram.



- a) List the errors each student has made.
- b) Draw a better cumulative frequency graph and histogram.

2 The tables give the size of households and age distribution of the population for three adjacent postcodes in a coastal region.

Number in household	TQ4	TQ5	TQ6
1	4920	5610	1190
2	4310	6190	1980
3	1520	2540	580
4	970	1990	450
5	330	610	135
6	125	250	50
7	28	38	9
8+	14	27	6
Total	12217	17256	4400

Age group	Number of people		
	TQ4	TQ5	TQ6
0–4	1380	2290	410
5–9	1120	2110	435
10–15	1530	2860	610
16–19	1460	1970	400
20–29	3700	4360	755
30–44	4350	7040	1450
45–59	5200	8110	2280
60–74	4730	6960	2460
75–89	2410	3400	1145
90+	335	410	96
Total	26215	39510	10041

Use statistical diagrams and measures to display and analyse this information.

Write a report on your findings. Include a comparison of the distribution in ages and household sizes in the three areas.

Consolidation Exercise 1

1

- a)** Describe what is meant by
- i. a random sample
 - ii. a representative sample.
- b)** Give one example of the use of
- i. cluster sampling
 - ii. quota sampling.
- c)** The table gives the number of students in each year of a computing course.

	Year 1	Year 2	Year 3
Male	21	24	18
Female	9	8	6

The course tutor wants to select a representative sample of 24 students.

- i. How many students from each group should he select?
- ii. Suggest two ways he could select the students he needs from each group.

2 A geography student measures the heights of plants grown in shade and in sunlight.

The results in centimetres are in the table.

Shade	13	19	9	22	27
	24	14	16	21	8
	28	9	11	19	17
	10	15	18	17	14
Sunlight	16	25	25	13	19
	14	20	30	23	28
	19	23	16	12	8
	23	24	26	29	10

- a)**
- i. Draw a back-to-back stem-and-leaf diagram.
 - ii. Write down the modes for each group of plants.
 - iii. Work out the range of heights for each group.
- b)**
- i. Draw a box and whisker plot for each distribution.
 - ii. Describe and compare the distributions.

3 The table gives college students' marks in an A level German test.

Marks	Number of students
1–10	0
11–20	6
21–30	14
31–40	26
41–50	19
51–60	7

- a) i. Draw a cumulative frequency graph.
 ii. Find the 35th percentile. What does this tell you?
 iii. The pass mark for the test is 35%. Estimate the number of students who passed.
- b) Draw a box and whisker plot to illustrate the distribution of marks.

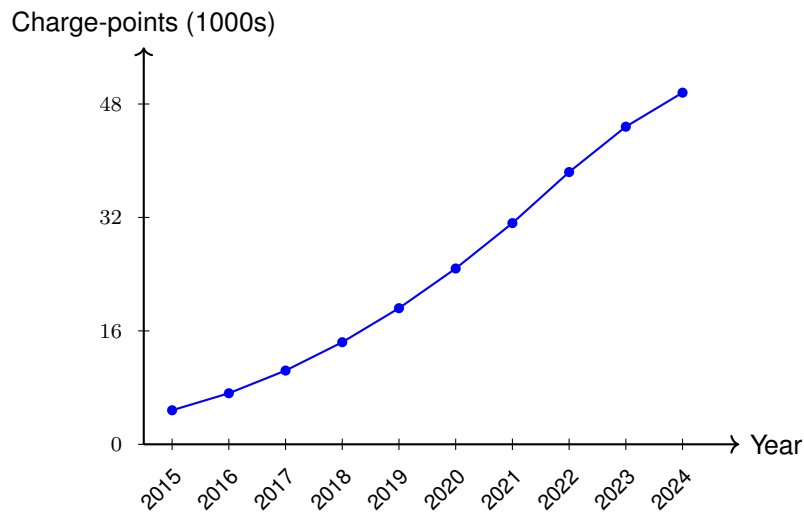
4 In a survey Marcus asks students how much they spent on takeaways last month. He groups his results in this table.

Amount (£ P)	Number of students
$P < 15$	11
$15 \leq P < 25$	23
$25 \leq P < 35$	31
$35 \leq P < 50$	16
$50 \leq P < 80$	9
$P \geq 80$	3

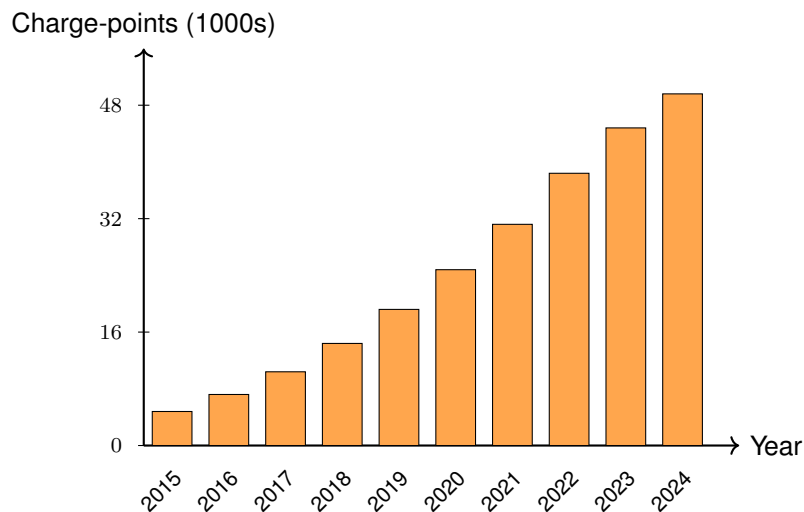
- a) Marcus says ‘My data is continuous, secondary data.’
 Is Marcus correct? Explain your answer.
- b) Draw a histogram for Marcus’s data.
- c) Find an estimate for the mean amount students spent on takeaways and the standard deviation.

5 Hannah and Reuben have drawn these diagrams to show the number of electric vehicle charge-points installed in a country between 2015 and 2024.

Hannah’s graph



Reuben's chart



- a) List the ways in which each diagram can be improved.
- b) Describe how the number of charge-points has changed over this time period.

6 A Local Authority has 28 primary schools, 12 secondary schools and 5 colleges for 16–18 year olds. The Local Authority wants to find out the opinions of teachers in the area. One suggestion is for the Local Authority to choose one school or college from each group at random, then interview three teachers from each.

- a) Give two reasons why this is not a good sample.
- b) Describe a better sampling method the Local Authority could use.

7 The table shows box office takings for documentary films.

Box office takings (£ T million)	Number of films
$0 < T < 0.05$	286
$0.05 \leq T < 0.5$	94
$0.5 \leq T < 2$	48
$2 \leq T < 5$	17
$5 \leq T < 10$	9
$10 \leq T < 15$	3
$15 \leq T < 25$	4

- a) Write down the modal class.
- b) Give an estimate for
 - i. the mean
 - ii. the standard deviation.
- c) Estimate the median and interquartile range. Describe any problems that arise.
- d) Write three sentences explaining what the data shows.

8 For routine treatment, the waiting time to see a specialist should be no more than 16 weeks from referral. The table gives the waiting time results for one clinic.

Waiting time	% of patients
Less than 4 weeks	3
4–8 weeks	19
9–13 weeks	29
14–15 weeks	33
16 weeks	11
17 weeks	4
18 weeks	1
More than 18 weeks	0

Comment on the clinic's performance.

Use statistical measures and/or diagrams to support your comments.

9 The table below gives information about average weekly earnings by region in one country.

Region	Mean weekly earnings (£)	Number of employees (1000s)
North	612	2840
Midlands	588	3120
South West	597	2510
South East	724	4680
Capital	812	5230

Here is a question you could ask based on this information:

How does the overall mean weekly earnings figure compare with a simple average of the five regional means?

Write down two more questions that you could ask.

10 Use the internet to find out more about regional pay inequality in a country of your choice.

Write a report or prepare a presentation including statistical measures and diagrams that support your findings.