3 Data Analysis and Interpretation

3.1 Mean, Median, Mode and Range

You have already looked at ways of collecting and representing data. In this section, you will go one step further and find out how to calculate statistical quantities which summarise the important characteristics of the data.

The *mean*, *median* and *mode* are three different ways of describing the average.

- To find the *mean*, add up all the numbers and divide by the number of numbers.
- To find the *median*, place all the numbers in order and select the middle number. If there are two values in the middle then the median is the mean of those two numbers.
- The *mode* is the number which appears most often.
- The *range* gives an idea of how the data are spread out and is the difference between the smallest and largest values.

Worked Example 1

Find

(a) the mean (b) the median (c) the mode (d) the range

of this set of data.

$$5,\ 6,\ 2,\ 4,\ 7,\ 8,\ 3,\ 5,\ 6,\ 6$$

Solution

(a) The mean is

$$\frac{5+6+2+4+7+8+3+5+6+6}{10}$$
$$= \frac{52}{10}$$
$$= 5.2$$

(b) To find the median, place all the numbers in order.

$$2,\ 3,\ 4,\ 5,\ 5,\ 6,\ 6,\ 6,\ 7,\ 8$$

As there are *two* middle numbers in this example, 5 and 6,

$$\text{median} = \frac{5+6}{2}$$
$$= \frac{11}{2}$$
$$= 5.5$$

(c) From the list in (b) it is easy to see that 6 appears more than any other number, so

$$\text{mode} = 6$$

(d) The range is the difference between the smallest and largest numbers, in this case 2 and 8. So the range is $8 - 2 = 6$.

Worked Example 2

Five people play golf and at one hole their scores are

3, 4, 4, 5, 7

For these scores, find

(a) the mean (b) the median

(c) the mode (d) the range.

Solution

(a) The mean is

$$\frac{3+4+4+5+7}{5}$$

$$= \frac{23}{5}$$

$$= 4.6$$

(b) The numbers are already in order and the middle number is 4.

So

$$\text{median} = 4$$

(c) The score 4 occurs most often, so,

$$\text{mode} = 4$$

(d) The range is the difference between the smallest and largest numbers, in this case 3 and 7, so

$$\text{range} = 7 - 3$$

$$= 4$$

3.1

Exercises

1. Find the mean median, mode and range of each set of numbers below.

   (a) 3, 4, 7, 3, 5, 2, 6, 10

   (b) 8, 10, 12, 14, 7, 16, 5, 7, 9, 11

   (c) 17, 18, 16, 17, 17, 14, 22, 15, 16, 17, 14, 12

   (d) 108, 99, 112, 111, 108

   (e) 64, 66, 65, 61, 67, 61, 57

   (f) 21, 30, 22, 16, 24, 28, 16, 17

2. Twenty children were asked their shoe sizes. The results are given below.

8,	6,	7,	6,	5,	$4\frac{1}{2}$,	$7\frac{1}{2}$,	$6\frac{1}{2}$,	$8\frac{1}{2}$,	10
7,	5,	$5\frac{1}{2}$	8,	9,	7,	5,	6,	$8\frac{1}{2}$	6

   For this data, find

   (a) the mean (b) the median

   (c) the mode (d) the range.

3. Eight people work in a shop. They are paid hourly rates of

   £2, £15, £5, £4, £3, £4, £3, £3.

   (a) Find

   (i) the mean (ii) the median (iii) the mode.

   (b) Which average would you use if you wanted to claim that the staff were:

   (i) well paid (ii) badly paid?

   (c) What is the range?

4. Two people work in a factory making parts for cars. The table shows how many complete parts they make in one week.

Worker	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*
Fred	20	21	22	20	21
Harry	30	15	12	36	28

   (a) Find the mean and range for Fred and Harry.

   (b) Who is most consistent?

   (c) Who makes the most parts in a week?

5. A gardener buys 10 packets of seeds from two different companies. Each pack contains 20 seeds and he records the number of plants which grow from each pack.

Company A	20	5	20	20	20	6	20	20	20	8
Company B	17	18	15	16	18	18	17	15	17	18

(a) Find the mean, median and mode for each company's seeds.

(b) Which company does the mode suggest is best?

(c) Which company does the mean suggest is best?

(d) Find the range for each company's seeds.

6. Adrian takes four tests and scores the following marks.

65, 72, 58, 77

(a) What are his median and mean scores?

(b) If he scores 70 in his next test, does his mean score increase or decrease? Find his new mean score.

(c) Which has increased most, his mean score or his median score?

7. Richard keeps a record of the number of fish he catches over a number of fishing trips. His records are:

1, 0, 2, 0, 0, 0, 12, 0, 2, 0, 0, 1, 18, 0, 2, 0, 1.

(a) Why does he object to talking about the mode and median of the number of fish caught?

(b) What are the mean and range of the data?

(c) Richard's friend, Najir, also goes fishing. The mode of the number of fish he has caught is also 0 and his range is 15.

What is the largest number of fish that Najir has caught?

8. A garage owner records the number of cars which visit his garage on 10 days. The numbers are:

204, 310, 279, 314, 257, 302, 232, 261, 308, 217

(a) Find the mean number of cars per day.

(b) The owner hopes that the mean will increase if he includes the number of cars on the next day. If 252 cars use the garage on the next day, will the mean increase or decrease?

9. The children in a class state how many children there are in their family. The numbers they state are given below.

1, 2, 1, 3, 2, 1, 2, 4, 2, 2, 1, 3, 1, 2,
2, 2, 1, 1, 7, 3, 1, 2, 1, 2, 2, 1, 2, 3

(a) Find the mean, median and mode for this data.

(b) Which is the most sensible average to use in this case?

10. The mean number of people visiting Jane each day over a five-day period is 8. If 10 people visit Jane the next day, what happens to the mean?

11. The table shows the maximum and minimum temperatures recorded in six cities one day last year.

City	*Maximum*	*Minimum*
Los Angeles	22° C	12° C
Boston	22° C	– 3° C
Moscow	18° C	– 9° C
Atlanta	27° C	8° C
Archangel	13° C	– 15° C
Cairo	28° C	13° C

(a) Work out the range of temperature for Atlanta.

(b) Which city in the table had the lowest temperature?

(c) Work out the difference between the maximum temperature and the minimum temperature for Moscow.

(LON)

12. The weights, in grams, of seven potatoes are

260, 225, 205, 240, 232, 205, 214

What is the median weight?

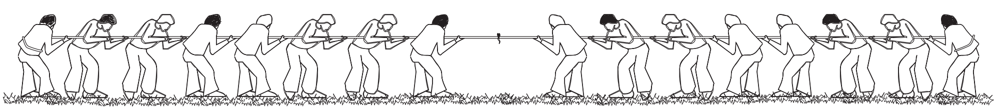
13. Here are the number of goals scored by a school football team in their matches this term.

3, 2, 0, 1, 2, 0, 3, 4, 3, 2

(a) Work out the mean number of goals.

(b) Work out the range of the number of goals scored.

(LON)

14. 

(a) The weights, in kilograms, of the 8 members of *Hereward House* tug of war team at a school sports are:

75, 73, 77, 76, 84, 76, 77, 78

Calculate the mean weight of the team.

(b) The 8 members of *Nelson House* tug of war team have a mean weight of 64 kilograms.

Which team do you think will win a tug of war between *Hereward House* and *Nelson House*? Give a reason for your answer.

(MEG)

15. Pupils in Year 8 are arranged in eleven classes. The class sizes are

23, 24, 24, 26, 27, 28, 30, 24, 29, 24, 27

(a) What is the modal class size?

(b) Calculate the mean class size.

The range of the class sizes for Year 9 is 3.

(c) What does this tell you about the class sizes in Year 9 compared with those in Year 8?

(SEG)

16. A school has to select one pupil to take part in a General Knowledge Quiz.

Kim and Pat took part in six trial quizzes. The following lists show their scores.

Kim	28	24	21	27	24	26
Pat	33	19	16	32	34	18

Kim had a mean score of 25 with a range of 7.

(a) Calculate Pat's mean score and range.

(b) Which pupil would you choose to represent the school? Explain the reason for your choice, referring to the mean scores and ranges.

(MEG)

17. Eight judges each give a mark out of 6 in an ice-skating competition.

Oksana is given the following marks.

5.3, 5.7, 5.9, 5.4, 4.5, 5.7, 5.8, 5.7

The mean of these marks is 5.5, and the range is 1.4.

The rules say that the highest mark and the lowest mark are to be deleted.

5.3, 5.7, ~~5.9~~, 5.4, ~~4.5~~, 5.7, 5.8, 5.7

(a) (i) Find the mean of the six remaining marks.

(ii) Find the range of the six remaining marks.

(b) Do you think it is better to count all eight marks, or to count only the six remaining marks? Use the means and the ranges to explain your answer.

(c) The eight marks obtained by Tonya in the same competition have a mean of 5.2 and a range of 0.6. Explain why none of her marks could be as high as 5.9.

(MEG)

18. Zena and Charles played nine rounds of crazy golf on their summer holidays. Their scores shown on the back to back stem and leaf diagram.

Zena		Charles
	3	0 0 2
1	4	1 1 1 2
9 3 1 0 0	5	2
6 5 4	6	8

Charles' lowest score was 30.

(a) What was Zena's lowest score? (b) What was Charles' modal score?

(c) What was Zena's median score?

In crazy golf the player with the lowest score wins.

Charles actually made the highest score that summer but was still chosen as the better player.

(d) Give a reason for this choice. *(SEG)*

19. Sandra and Aziz record the heights, in millimetres, of 25 seedlings.

These are the heights obtained.

42	37	53	57	62
37	46	68	54	53
49	64	51	58	37
70	42	57	51	60
36	48	55	63	56

(a) Construct a stem and leaf diagram for these results.

(b) Using your stem and leaf diagram, or otherwise, find,

(i) the median, (ii) the mode, (iii) the range.

(c) Which of the two averages, the mode or the median, do you think is more representative of the data? Give a reason for your answer.

20. The stem and leaf diagram opposite shows the number of passengers using the 8 o'clock bus to Upchester over a period of 15 weekdays.

Stem (tens)	*Leaf (units)*
0	8 9
1	1 4 4 5 8
2	1 2 3 3 3 5 7
3	1

(a) Copy and complete the frequency table below.

Number of passengers	*Frequency*
5 – 9	2
10 – 14	
15 – 19	
20 – 24	
25 – 29	
30 – 34	

(b) An inspector was sent to see how well the bus service was used.

(i) What is the probability that, on the day she chose, there were fewer than ten passengers on the bus?

(ii) What is the probability that, on the day she chose, there were twenty or more passengers on the bus?

(NEAB)

3.2 Finding the Mean from Tables and Tally Charts

Often data are collected into tables or tally charts. This section considers how to find the mean in such cases.

Worked Example 1

A football team keep records of the number of goals it scores per match during a season.

No. of Goals	*Frequency*
0	8
1	10
2	12
3	3
4	5
5	2

Find the mean number of goals per match.

Solution

The table above can be used, with a third column added.

No. of Goals	*Frequency*	*No. of Goals* $\times$ *Frequency*
0	8	$0 \times 8 = 0$
1	10	$1 \times 10 = 10$
2	12	$2 \times 12 = 24$
3	3	$3 \times 3 = 9$
4	5	$4 \times 5 = 20$
5	2	$5 \times 2 = 10$
TOTALS	**40**	**73**
	(Total matches)	(Total goals)

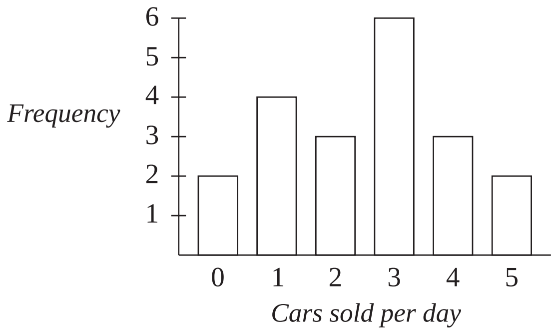
The mean can now be calculated.

$$\text{Mean} = \frac{73}{40}$$

$$= 1.825 \text{ goals per match}$$

Worked Example 2

The bar chart shows how many cars were sold by a salesman over a period of time.



Find the mean number of cars sold per day.

3.2

Solution

The data can be transferred to a table and a third column included as shown.

Cars sold daily	*Frequency*	*Cars sold* $\times$ *Frequency*
0	2	$0 \times 2 = 0$
1	4	$1 \times 4 = 4$
2	3	$2 \times 3 = 6$
3	6	$3 \times 6 = 18$
4	3	$4 \times 3 = 12$
5	2	$5 \times 2 = 10$
TOTALS	**20**	**50**
	(Total days)	(Total number of cars sold)

$$\text{Mean} = \frac{50}{20}$$
$$= 2.5 \text{ cars sold per day}$$

Worked Example 3

A police station kept records of the number of road traffic accidents in their area each day for 100 days. The figures below give the number of accidents per day.

1	4	3	5	5	2	5	4	3	2	0	3	1	2	2	3	0	5	2	1
3	3	2	6	2	1	6	1	2	2	3	2	2	2	2	5	4	4	2	3
3	1	4	1	7	3	3	0	2	5	4	3	3	4	3	4	5	3	5	2
4	4	6	5	2	4	5	5	3	2	0	3	3	4	5	2	3	3	4	4
1	3	5	1	1	2	2	5	6	6	4	6	5	8	2	5	3	3	5	4

Find the mean number of accidents per day.

Solution

The first step is to draw out and complete a tally chart. The final column shown below can then be added and completed.

Number of Accidents	*Tally*	*Frequency*	*No. of Accidents* $\times$ *Frequency*
0	||||	4	$0 \times 4 = 0$
1	𝍸 𝍸	10	$1 \times 10 = 10$
2	𝍸 𝍸 𝍸 𝍸 ||	22	$2 \times 22 = 44$
3	𝍸 𝍸 𝍸 𝍸 |||	23	$3 \times 23 = 69$
4	𝍸 𝍸 𝍸 |	16	$4 \times 16 = 64$
5	𝍸 𝍸 𝍸 ||	17	$5 \times 17 = 85$
6	𝍸 |	6	$6 \times 6 = 36$
7	|	1	$7 \times 1 = 7$
8	|	1	$8 \times 1 = 8$
	TOTALS	**100**	**323**

$$\text{Mean number of accidents per day} = \frac{323}{100} = 3.23.$$

3.2

Exercises

1. A survey of 100 households asked how many cars there were in each household The results are given below.

No. of Cars	*Frequency*
0	5
1	70
2	21
3	3
4	1

Calculate the mean number of cars per household.

2. The survey of question 1 also asked how many TV sets there were in each household. The results are given below.

No. of TV Sets	*Frequency*
0	2
1	30
2	52
3	8
4	5
5	3

Calculate the mean number of TV sets per household.

3. A manager keeps a record of the number of calls she makes each day on her mobile phone.

Number of calls per day	0	1	2	3	4	5	6	7	8
Frequency	3	4	7	8	12	10	14	3	1

Calculate the mean number of calls per day.

4. A cricket team keeps a record of the number of runs scored in each over.

No. of Runs	*Frequency*
0	3
1	2
2	1
3	6
4	5
5	4
6	2
7	1
8	1

(a) Calculate the mean number of runs per over.

(b) How many runs would the team expect to score in a 40-over match?

5. A class conducted an experiment in biology. They placed a number of 1 m by 1 m square grids on the playing field and counted the number of worms which appeared when they poured water on the ground. The results obtained are given below.

6	3	2	1	3	2	1	3	0	1
0	3	2	1	1	4	0	1	2	0
1	1	2	2	2	4	3	1	1	1
2	3	3	1	2	2	2	1	7	1

(a) Calculate the mean number of worms.

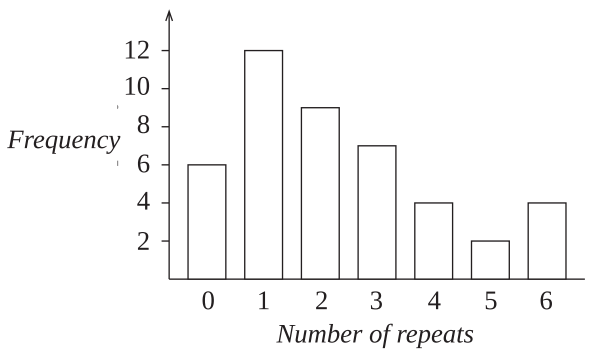
(b) How many times was the number of worms seen greater than the mean?

6. As part of a survey, a station recorded the number of trains which were late each day. The results are listed below.

0	1	2	4	1	0	2	1	1	0
1	2	1	3	1	0	0	0	0	5
2	1	3	2	0	1	0	1	2	1
1	0	0	3	0	1	2	1	0	0

Construct a table and calculate the mean number of trains which were late each day.

7. Hannah has been collecting football cards. Sometimes when she bought a new packet she found cards that she had already collected. She drew up this table to show the number of repeated cards in the packs she opened.



Calculate the mean number of repeats per packet.

8. In a season a football team scored a total of 55 goals. The table opposite gives a summary of the number of goals per match.

   (a) In how many matches did they score 2 goals?

   (b) Calculate the mean number of goals per match.

Goals per Match	*Frequency*
0	4
1	6
2	
3	8
4	2
5	1

9. A traffic warden is trying to work out the mean number of parking tickets he has issued per day. He produced the table below, but has accidentally rubbed out some of the numbers.

Tickets per day	*Frequency*	*No. of Tickets $\times$ Frequency*
0	1	
1		1
2	10	
3	7	
4		20
5	2	
6		
TOTALS	**26**	**72**

   Fill in the missing numbers and calculate the mean.

10. The number of children per family in a recent survey of 21 families is shown.

1	2	3	2	2	4	2	2
3	2	2	2	3	2	2	2
4	1	2	3	2			

   (a) What is the range in the number of children per family?

   (b) Calculate the mean number of children per family. *Show your working.*

   A similar survey was taken in 1960.

   In 1960 the range in the number of children per family was 7 and the mean was 2.7.

   (c) Describe **two** changes that have occurred in the number of children per family since 1960.

(SEG)

3.3 Mean, Median and Mode for Grouped Data

The mean and median can be estimated from tables of *grouped* data.

The class interval which contains the most values is known as the *modal class*.

Note

Worked Examples 1 and 2 use *continuous data*, since height can be of any value within a given range. Other examples of continuous data are weight, temperature, area and volume. Worked Example 3 uses *discrete data*, that is, data which can take only a particular value, such as the integers 1, 2, 3, 4, . . . in this case.

Worked Example 1

The table below gives data on the heights, in cm, of 51 children.

Class Interval	$140 \le h < 150$	$150 \le h < 160$	$160 \le h < 170$	$170 \le h < 180$
Frequency	6	16	21	8

(a) Estimate the mean height.

(b) Find the class interval that contains the median value.

(c) Find the modal class.

Solution

(a) To estimate the mean, the mid-point of each interval should be used.

Class Interval	*Mid-point*	*Frequency*	*Mid-point* $\times$ *Frequency*
$140 \le h < 150$	145	6	$145 \times 6 = 870$
$150 \le h < 160$	155	16	$155 \times 16 = 2480$
$160 \le h < 170$	165	21	$165 \times 21 = 3465$
$170 \le h < 180$	175	8	$175 \times 8 = 1400$
	Totals	**51**	**8215**

$$\text{Mean} = \frac{8215}{51}$$

$$= 161 \text{ (to the nearest cm)}$$

(b) When there are too many items of data to sensibly write them in order or the data is in grouped form as in this question, to find the position of the median value we need to use the formula $\frac{n+1}{2}$, where n is the total frequency, which here gives $\frac{51+1}{2} = 26$. Therefore the median is the 26th value. In this case the median lies in the interval $160 \text{ cm} \le h < 170 \text{ cm}$. This is because 22 values $(6+16)$ lie below

160 and a further 21 values are in the $160 \text{ cm} \le h < 170 \text{ cm}$ interval which clearly must include the 26th value.

(c) The modal class is $160 \text{ cm} \le h < 170 \text{ cm}$ as it contains the most values.

Note

When we speak of someone by age, say 8, then the person could be any age from 8 years 0 days up to 8 years 364 days (365 if the year is a leap year). You will see how this is tackled in the following example.

Worked Example 2

The age of children in a primary school were recorded in the table below.

Age	5 – 6	7 – 8	9 – 10
Frequency	29	40	38

(a) Estimate the mean.

(b) Find the class interval that contains the median value.

(c) Find the modal age.

Solution

(a) To estimate the mean, we must use the mid-point of each interval; so, for example for '5 – 6', which really means

$$5 \le \text{age} < 7$$

the mid-point is taken as 6.

Age group	*Mid-point*	*Frequency*	*Mid-point* × *Frequency*
5 – 6	6	29	$6 \times 29 = 174$
7 – 8	8	40	$8 \times 40 = 320$
9 – 10	10	38	$10 \times 38 = 380$
	Totals	**107**	**874**

$$\text{Mean} = \frac{874}{107}$$

$$= 8.2 \text{ (to 1 decimal place)}$$

(b) The median is given by the 54th value which lies within the age group 7 – 8, since $29 + 40$ is greater than 54 so this interval must contain the median.

(c) The modal age is the 7 – 8 age group because there are more children in this age group than in any other.

3.3

Worked Example 3

The number of days that children were missing from school due to sickness in one year was recorded.

Number of days off sick	1 – 5	6 – 10	11 – 15	16 – 20	21 – 25
Frequency	12	11	10	4	3

(a) Estimate the mean.

(b) Find the class interval that contains the median value.

(c) Find the modal class.

Solution

(a) The estimate is made by assuming that all the values in a class interval are equal to the midpoint of the class interval.

Class interval	*Mid-point*	*Frequency*	*Mid-point* $\times$ *Frequency*
1–5	3	12	$3 \times 12 = 36$
6–10	8	11	$8 \times 11 = 88$
11–15	13	10	$13 \times 10 = 130$
16–20	18	4	$18 \times 4 = 72$
21–25	23	3	$23 \times 3 = 69$
	Totals	**40**	**395**

$$\text{Mean} = \frac{395}{40} = 9.875 \text{ days}$$

This means that on average each pupil missed approximately 10 days due to sickness.

(b) The median lies between the 20th and the 21st values which both lie within the class interval 6 – 10, so the median is in the class interval 6 – 10.

(c) The modal class is 1–5, as this class contains the most entries.

3.3

Exercises

1. A door to door salesman keeps a record of the number of homes he visits each day.

Homes visited	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49
Frequency	3	8	24	60	21

(a) Estimate the mean number of homes visited.

(b) Find the class interval that contains the median value.

(c) What is the modal class?

2. The weights of a number of students were recorded in kg.

Mean (kg)	$30 \leq w < 35$	$35 \leq w < 40$	$40 \leq w < 45$	$45 \leq w < 50$	$50 \leq w < 55$
Frequency	10	11	15	7	4

(a) Estimate the mean weight.

(b) Find the class interval that contains the median value.

(c) What is the modal class?

3. A stopwatch was used to find the time that it took a group of children to run 100 m.

Time (seconds)	$10 \leq t < 15$	$15 \leq t < 20$	$20 \leq t < 25$	$25 \leq t < 30$
Frequency	6	16	21	8

(a) Find the class interval that contains the median value.

(b) Is the median in the modal class? Explain your answer.

(c) Estimate the mean.

4. The distances that children in a year group travelled to school is recorded.

Distance (km)	$0 \leq d < 0.5$	$0.5 \leq d < 1.0$	$1.0 \leq d < 1.5$	$1.5 \leq d < 2.0$
Frequency	30	22	19	8

(a) Does the modal class contain the median? Explain your answer.

(b) Estimate the mean.

5. The ages of the children at a youth camp are summarised in the table below.

Age (years)	6 – 8	9 – 11	12 – 14	15 – 17
Frequency	8	22	29	5

Estimate the mean age of the children.

6. The lengths of a number of leaves collected for a project are recorded.

Length (cm)	2 – 5	6 – 10	11 – 15	16 – 25
Frequency	8	20	42	12

(a) Estimate the mean

(b) Find the class interval that contains the median length of a leaf.

7. The table shows how many nights people spend at a campsite.

Number of nights	1 – 5	6 – 10	11 – 15	16 – 20	21 – 25
Frequency	20	26	32	5	2

(a) Estimate the mean.

(b) Find the class interval that contains the median value.

(c) What is the modal class?

8. (a) A teacher notes the number of correct answers given by a class on a multiple-choice test.

Correct answers	1 – 10	11 – 20	21 – 30	31 – 40	41 – 50
Frequency	2	8	15	11	3

(i) Estimate the mean.

(ii) Find the class interval that contains the median value.

(iii) What is the modal class?

(b) Another class took the same test. Their results are given below.

Correct answers	1 – 10	11 – 20	21 – 30	31 – 40	41 – 50
Frequency	3	14	20	2	1

(i) Estimate the mean.

(ii) Find the class interval that contains the median value.

(iii) What is the modal class?

(c) How do the results for the two classes compare?

9. 29 children are asked how much pocket money they were given last week. Their replies are shown in this frequency table.

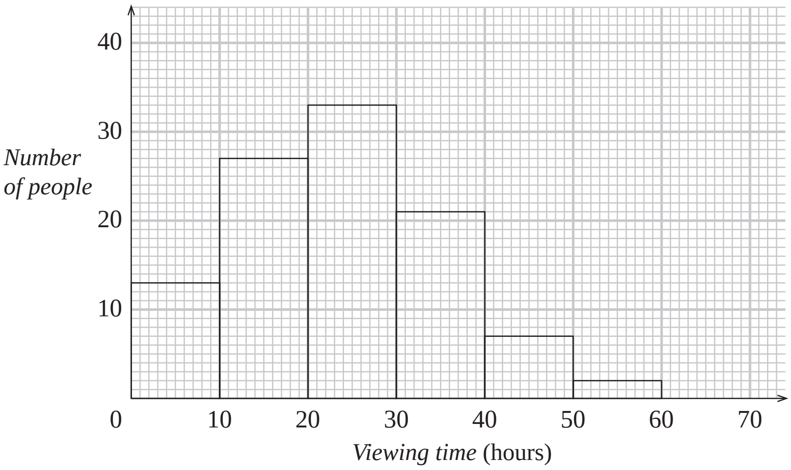
Pocket money £	*Frequency* *f*
0 – £1.00	12
£1.01 – £2.00	9
£2.01 – £3.00	6
£3.01 – £4.00	2

(a) Which is the modal class?

(b) Calculate an estimate of the mean amount of pocket money received per child.

(NEAB)

10. The graph shows the number of hours a sample of people spent viewing television one week during the summer.



(a) Copy and complete the frequency table for this sample.

Viewing time (*h* hours)	*Number of people*
$0 \le h < 10$	13
$10 \le h < 20$	27
$20 \le h < 30$	33
$30 \le h < 40$	
$40 \le h < 50$	
$50 \le h < 60$	

(b) Another survey is carried out during the winter. State **one** difference you would expect to see in the data.

(c) Use the mid-points of the class intervals to calculate the mean viewing time for these people. You may find it helpful to use the table below.

Viewing time (h hours)	*Mid-point*	*Frequency*	*Mid-point* × *Frequency*
$0 \le h < 10$	5	13	65
$10 \le h < 20$	15	27	405
$20 \le h < 30$	25	33	825
$30 \le h < 40$	35		
$40 \le h < 50$	45		
$50 \le h < 60$	55		

(SEG)

11. In an experiment, 50 people were asked to estimate the length of a rod to the nearest centimetre. The results were recorded.

Length (cm)	20	21	22	23	24	25	26	27	28	29
Frequency	0	4	6	7	9	10	7	5	2	0

(a) Find the value of the median. (b) Calculate the mean length.

(c) In a second experiment another 50 people were asked to estimate the length of the same rod. The most common estimate was 23 cm. The range of the estimates was 13 cm.

Make two comparisons between the results of the two experiments. *(SEG)*

12. The following list shows the maximum daily temperature, in °F, throughout the month of April.

56.1	49.4	63.7	56.7	55.3	53.5	52.4	57.6	59.8	52.1
45.8	55.1	42.6	61.0	61.9	60.2	57.1	48.9	63.2	68.4
55.5	65.2	47.3	59.1	53.6	52.3	46.9	51.3	56.7	64.3

(a) Copy and complete the grouped frequency table below.

Temperature, T		*Frequency*
$40 < T \le 50$		
$50 < T \le 54$		
$54 < T \le 58$		
$58 < T \le 62$		
$62 < T \le 70$		

(b) Use the table of values in part (a) to calculate an estimate of the mean of this distribution. *You **must** show your working clearly*.

(c) Draw a histogram to represent your distribution in part (a). *(MEG)*

3.4 Calculations with the Mean

This section considers calculations concerned with the mean.

Worked Example 1

The mean of a sample of 6 numbers is 3.2. An extra value of 3.9 is included in the sample. What is the new mean?

Solution

$$\begin{aligned}
\text{Total of original numbers} &= 6 \times 3.2 \\
&= 19.2 \\
\text{New total} &= 19.2 + 3.9 \\
&= 23.1 \\
\text{New mean} &= \frac{23.1}{7} \\
&= 3.3
\end{aligned}$$

Worked Example 2

The mean number of a set of 5 numbers is 12.7. What extra number must be added to bring the mean up to 13.1?

Solution

$$\begin{aligned}
\text{Total of the original numbers} &= 5 \times 12.7 \\
&= 63.5 \\
\text{Total of the new numbers} &= 6 \times 13.1 \\
&= 78.6 \\
\text{Difference} &= 78.6 - 63.5 \\
&= 15.1
\end{aligned}$$

So the extra number is 15.1.

3.4

Exercises

1. The mean height of a class of 28 students is 162 cm. A new girl of height 149 cm joins the class. What is the mean height of the class now?

2. After 5 matches the mean number of goals scored by a football team per match is 1.8. If they score 3 goals in their 6th match, what is the mean after the 6th match?

3. The mean number of children ill at a school is 3.8 per day, for the first 20 school days of a term. On the 21st day 8 children are ill. What is the mean after 21 days?

4. The mean weight of 25 children in a class is 58 kg. The mean weight of a second class of 29 children is 62 kg. Find the mean weight of all the children.

5. A salesman sells a mean of 4.6 conservatories per day for 5 days. How many must he sell on the sixth day to increase his mean to 5 sales per day?

6. Adrian's mean score for four tests is 64%. He wants to increase his mean to 68% after the fifth test. What does he need to score in the fifth test?

7. The mean salary of the 8 people who work for a small company is £15 000. When an extra worker is taken on this mean drops to £14 000. How much does the new worker earn?

8. The mean of 6 numbers is 12.3. When an extra number is added, the mean changes to 11.9. What is the extra number?

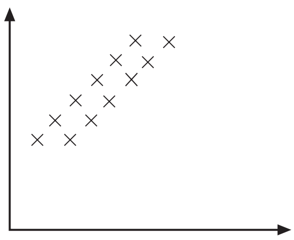
9. When 5 is added to a set of 3 numbers the mean increases to 4.6. What was the mean of the original 3 numbers?

10. Three numbers have a mean of 64. When a fourth number is included the mean is doubled. What is the fourth number?

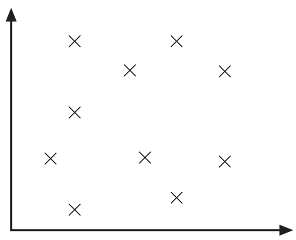
3.5 Scatter Plots and Lines of Best Fit

When there might be a connection between two different quantities, a *scatter plot* can be used. If there does appear to be a connection, a *line of best fit* can be drawn.

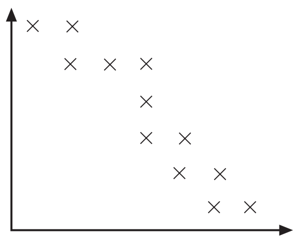
The following diagrams show 3 different scatter plots.



Positive correlation



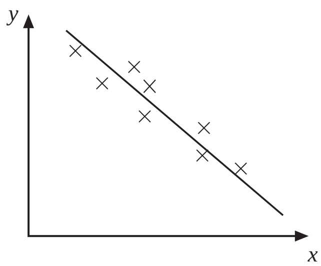
No correlation

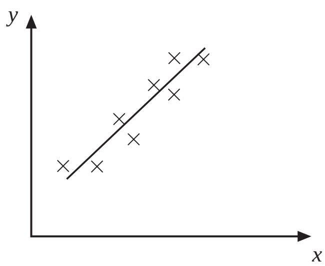


Negative correlation

If there is a relationship between the two quantities, there is said to be a *correlation* between the two quantities. This may be *positive* or *negative,* as shown in the examples above. When there is a positive correlation, one variable increases as the other increases. When there is a negative correlation, as one variable gets bigger the other gets smaller.

There is little value in attempting to draw lines of best fit unless there is either strong positive or strong negative correlation between the points plotted, as shown in the following diagrams.





Also note that the line of best fit should always pass through the point representing the mean values of the data points, i.e. through the point $(\bar{x}, \bar{y})$.

Note

A line of best fit can be used to estimate values using *interpolation* and *extrapolation.*

Interpolation involves finding a value within the range of the plotted points.

Extrapolation looks for values outside the range of the values given. Interpolation is generally more reliable than extrapolation. Extrapolation should be used with caution, and only to make estimates for values that are just outside the range of original data.

Note

Sometimes variables correlate in a non-linear way. This is shown using *curves of best fit*.

Worked Example 1

A salesman records, for each working day, how much petrol his car uses and how far he travels. The table shows his figures for 10 days.

Day	1	2	3	4	5	6	7	8	9	10
Petrol used (litres)	24	13	29	19	21	35	10	40	44	18
Distance travelled (miles)	200	150	320	180	190	280	120	360	400	160

3.5

(a) Plot a scatter graph and describe any connection that is present.

(b) Calculate (i) the mean amount of petrol used,

(ii) the mean distance travelled.

Plot the mean point.

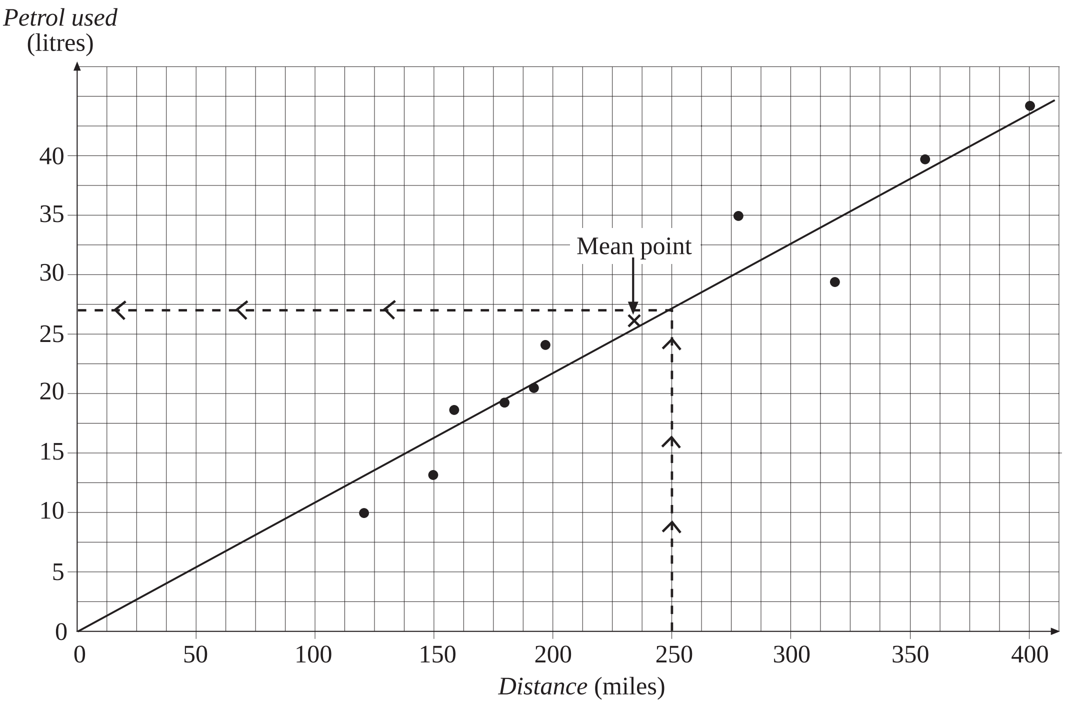
(c) Explain why it is sensible for a line of best fit to go through (0, 0).

(d) Draw a line of best fit.

(e) Estimate how much petrol would be used on a journey of 250 miles.

Solution

(a) Each point has been plotted on the graph below. This is an example of *positive correlation.* This means that the longer the journey, the more petrol is used.



(b) (i) Mean amount of petrol used

$$= \frac{24 + 13 + 29 + 19 + 21 + 35 + 10 + 40 + 44 + 18}{10}$$

$$= \frac{253}{10}$$

$$= 25.3 \text{ litres}$$

(ii) Mean distance travelled

$$= \frac{200 + 150 + 320 + 180 + 190 + 280 + 120 + 360 + 400 + 160}{10}$$

$$= \frac{2360}{10}$$

$$= 236 \text{ miles}$$

The mean point (236, 25.3) has been plotted with a cross on the scatter diagram.

3.5

(c) This is sensible because a car will not use any petrol if it is not used.

(d) A line of best fit has been drawn through the mean point and the origin. There are approximately the same number of points above and below the line.

(e) The dashed lines on the graph predict that approximately 27 litres of petrol are needed for a journey of 250 miles.

Note

Lines of best fit should not be drawn through the origin unless there is a sensible reason for doing so.

Worked Example 2

A sample of 8 U.S. companies showed the following sales and profit levels for the year ending April 1994.

Sales Turnover ($m) ($s$)	22	36	26	14	25	34	6	18
Profit ($m) ($p$)	1.8	4.9	0.8	0.9	3.2	3.7	0.5	2.1

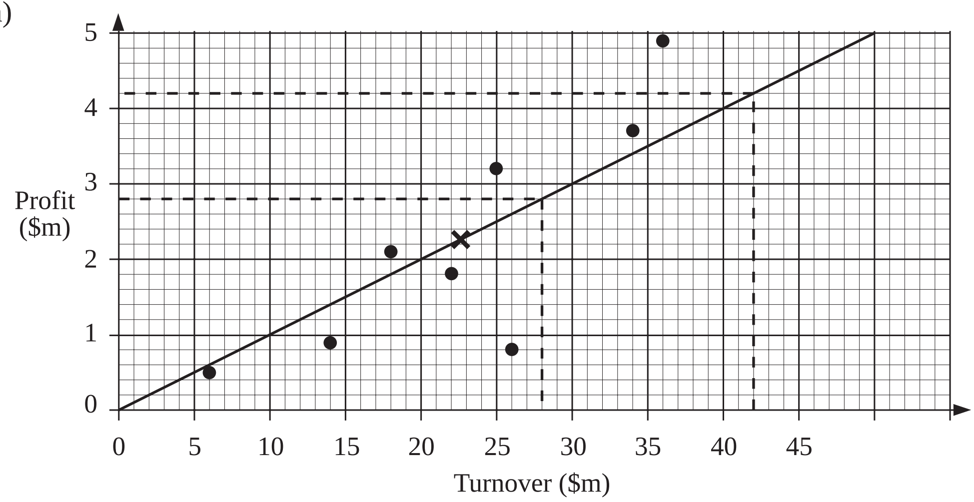
(a) Draw a scatter diagram of this information.

(b) After making suitable calculations draw in a line of best fit and use this to estimate Profit levels for *two* companies with annual turnovers respectively of $28m and $42m.

(c) State briefly which of the estimates in (b) is likely to be more accurate. Justify your choice. *(NEAB)*

Solution

(a)



(b) The mean values are calculated as $\bar{s} = 2.6$, $\bar{p} = 2.24$, and shown on the scatter diagram. (The line of best fit will pass through this point.)

For $s = \$28\text{m}$ the estimate of the profit is $2.8m, and for $s = \$42\text{m}$, the estimate is $4.2m.

(c) The estimate for a turnover of $28m is likely to be more accurate than for $42m, as the latter is outside the range of data on which the line of best fit is based.

Note The line of best fit in Worked Example 2 has been drawn through the origin on the assumption that a company with no turnover will make no profit.

3.5

Worked Example 3

Brunel plc is keen to set up a forecasting system which will enable them to estimate maintenance for delivery vehicles of various ages.

The following table summarises the age in months (x) and maintenance cost (y) for a sample of ten such vehicles.

Vehicle	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Age, months (x)	63	13	34	80	51	14	45	74	24	82
Maintenance Cost £, (y)	141	14	43	170	95	21	72	152	31	171

(a) Draw a scatter diagram of this data on graph paper.

(b) Find the mean value of the ages (x) and maintenance cost (y).

(c) Use your results from (b) and the fact that the line of best fit for the data passes through the point (20, 24.5) to draw this line on the graph.

(d) Estimate from your line the maintenance cost for a vehicle aged

(i) 85 months (ii) 5 months (iii) 60 months.

(e) Order these forecasts in terms of their reliability, listing the most reliable first. Justify your choice. *(NEAB)*

Solution

(a) See the following scatter diagram.

(b) $\bar{x} = 48$, $\bar{y} = 91$

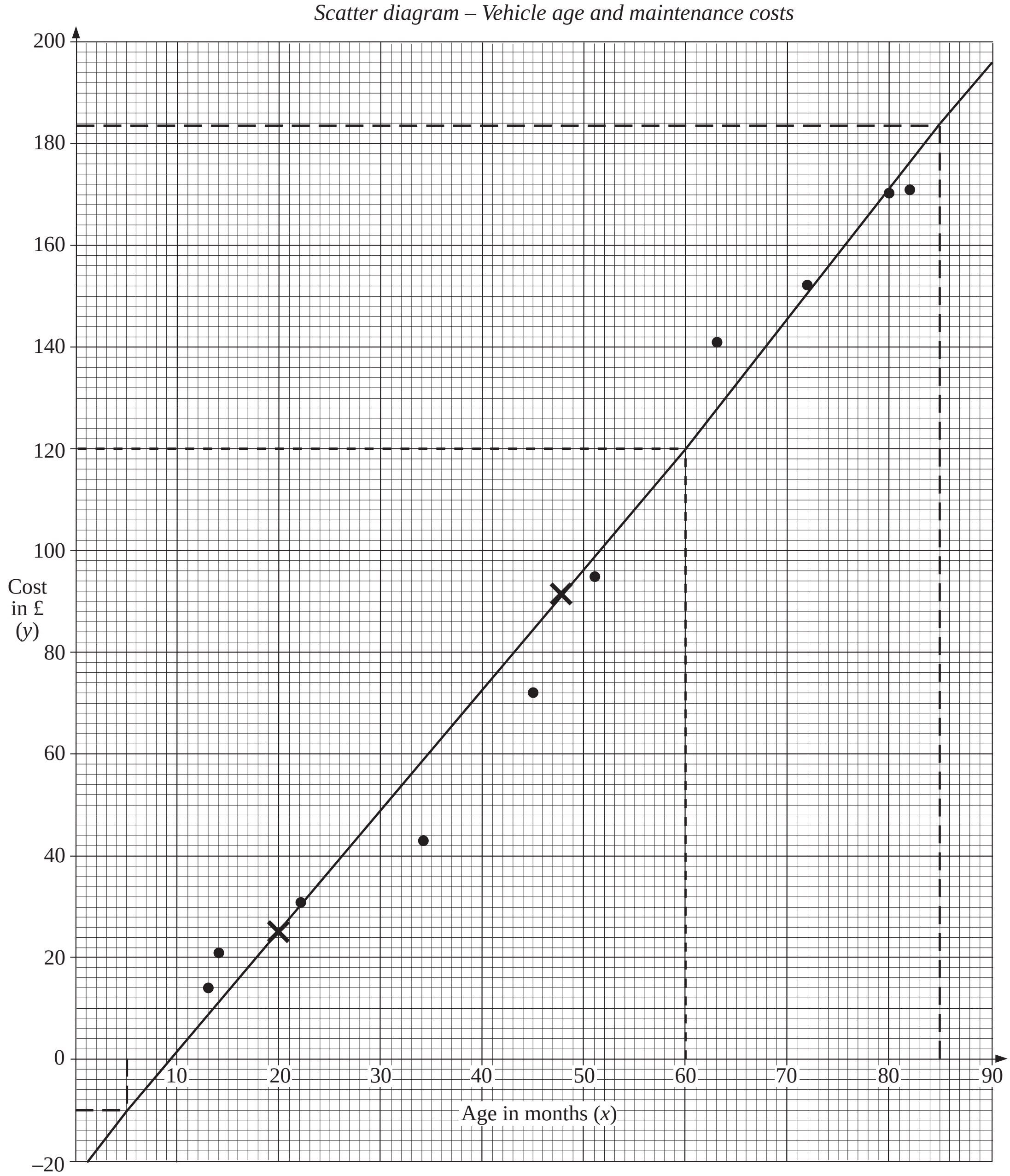
(c) See diagram.

(d) (i) £184 (ii) $-$£10 (iii) £120

(e) £120 (in middle of data range), £185 (just outside data range), $-$£10 (makes no sense!)

3.5

Scatter diagram – Vehicle age and maintenance costs



3.5

Exercises

1. The table gives the scores obtained by 10 students on three different tests.

Maths Test	7	19	7	18	10	18	11	14	11	16
Science Test	8	19	9	17	10	17	11	13	12	13
French Test	11	10	14	16	7	13	19	5	16	10

(a) Draw a scatter graph for maths against science.

(b) Draw a scatter graph for maths against French.

(c) Which set of points lie closer to a straight line?

(d) Would it be reasonable to draw a line of best fit in both cases? Explain your decision.

2. A firm records how long it takes a driver to make deliveries at warehouses at different distances from the factory.

Distance (miles)	155	65	80	145	100	95	50	120	90
Time taken (hours)	4.8	1.8	2.9	3.5	3.0	2.2	1.0	3.5	2.6

(a) Draw a scatter graph of time taken against distance.

(b) Describe the correlation and what it means.

(c) Calculate the mean distance and the mean time taken.

(d) Plot the mean point and use it to draw a line of best fit.

(e) A delivery takes 2 hours. Use your line to estimate how far the driver has travelled.

(f) How long would you expect a delivery to take if the driver has to travel 140 miles?

(g) Explain why it would not be sensible to use the line of best fit to estimate the time taken for a 300-mile journey.

3. The table shows the flying time and costs for holidays in some popular resorts.

Destination	*Flying time (hours)*	*Cost of holiday (£)*
Algarve	2.0	194
Benidorm	2.5	139
Gambia	6.0	357
Majorca	2.5	148
Morocco	3.0	237
Mombasa	8.5	523
Tenerife	4.5	238
Torremolinos	2.5	146
Tunisia	3.0	129

3.5

(a) Draw a scatter graph of cost against time.

(b) Describe the correlation and what it means.

(c) Calculate the mean flying time and the mean holiday cost.

(d) Plot the mean point and use it to draw a line of best fit.

(e) Estimate the cost of a holiday with a flying time of 5 hours.

(f) Estimate the flying time for a holiday that costs £400.

(g) Why would it not be sensible to use the line of best fit for a holiday costing £1000 ?

4. Ten children were weighed and then had their heights measured. The results are in the table.

Weight (kg)	47	60	47	49	50	59	46	54	57	53
Height (cm)	82	110	95	101	88	121	79	98	105	100

(a) Draw a scatter graph of height against weight.

(b) Describe the correlation and what it means.

(c) Calculate the mean height and the mean weight.

(d) Plot the mean point and draw a line of best fit.

(e) Comment on how well the line of best fit can be applied to the data.

(f) Estimate the height of a boy who weighs 60 kg.

(g) Estimate the weight of a girl who is 110 cm tall.

5. A group of children were tested on their tables. The time in seconds taken to do a test on the 7 times table and a test on the 3 times table were recorded.

Time for 3 times table	6	12	10	13	12	14	12	9	13	10	9
Time for 7 times table	9	20	15	20	19	22	18	15	17	17	13

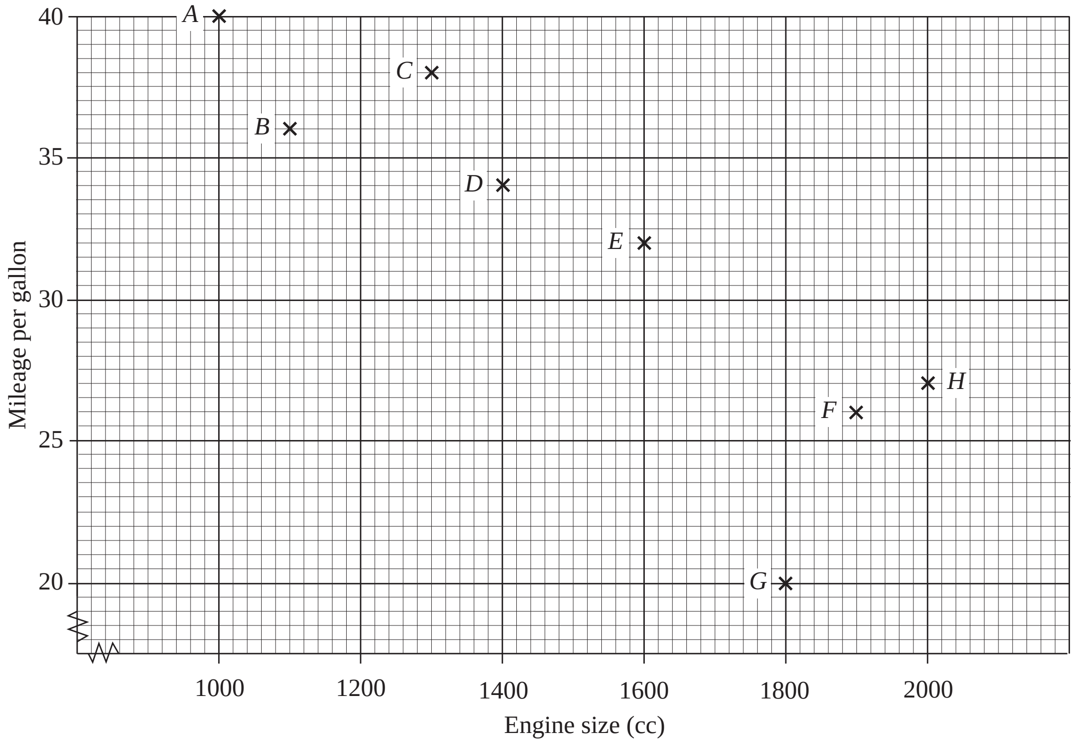
(a) Draw a scatter graph and a line of best fit.

(b) Ben missed the test for the 7 times table but took 11 seconds for the 3 times table test. Estimate how long he would have taken for the 7 times table test.

(c) Emma completed her 3 times table test in 5 seconds. She missed the test for the 7 times table. How long do you estimate that she would have taken for this test?

(d) Explain why you might get better estimates for Ben and Emma if there were more pupils included in the table.

6. A guide to used cars shows the engine size in cc and the mileage per gallon.



(a) Complete a copy of the table below.

Car	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Engine size (cc)	1000	1100	1300	1400		1900	1800	
Mileage per gallon	40	36	38	34	32		20	

(b) Another car, with engine size 1600 cc, was tested and its mileage was 30 per gallon. Plot this on the diagram, labelling it I.

(c) One car does not appear to follow the trend. Which one is it? Give a reason for your answer.

(NEAB)

7. Describe two quantities that you would expect to have

(a) a positive correlation,

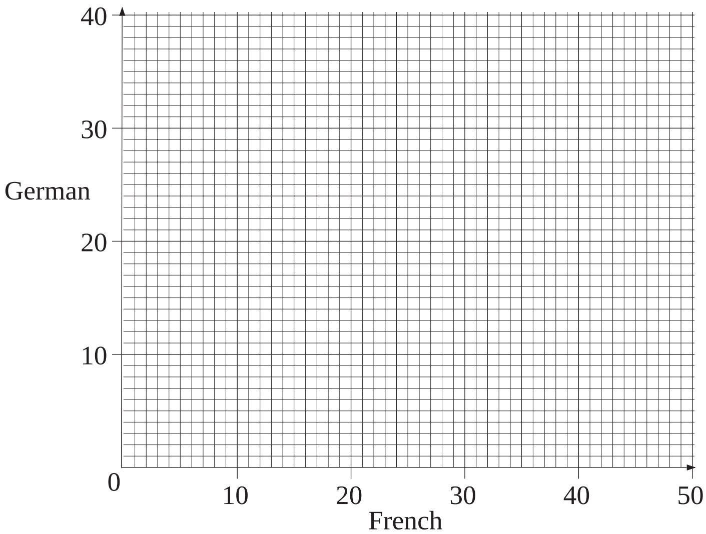
(b) no correlation,

(c) a negative correlation.

8. The table gives you the marks scored by pupils in a French test and in a German test.

French	15	35	34	23	35	27	36	34	23	24	30	40	25	35	20
German	20	37	35	25	33	30	39	36	27	20	33	35	27	32	28

(a) On a copy of the following grid, draw a scatter graph of the marks scored in the French and German tests.



(b) Describe the correlation between the marks scored in the two tests. *(LON)*

9. The table gives information about the age and value of a number of cars of the same type.

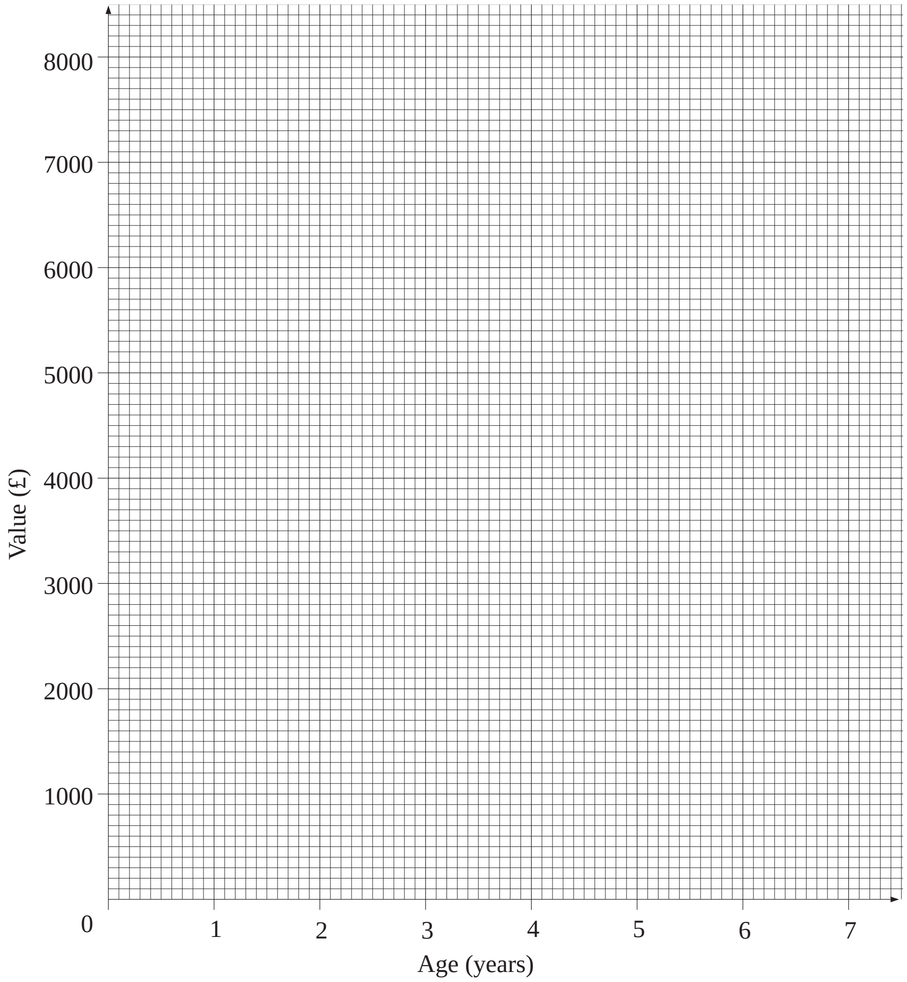
Age (years)	1	3	$4\frac{1}{2}$	6	3	5	2	$5\frac{1}{2}$	4	7
Value (£)	8200	5900	4900	3800	6200	4500	7600	2200	5200	3200

(a) Use this information to draw a scatter graph on a copy of the grid on the following page.

(b) What does the graph tell you about the value of these cars as they get older?

(c) Which car does not follow the general trend?

(d) Give a reason that might explain why the car in part (c) does not follow the general trend.



(SEG)

10. Ten people entered a craft competition.

Their displays of work were awarded marks by two different judges.

Competitor	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
First judge	90	35	60	15	95	25	5	100	70	45
Second judge	75	30	55	20	75	30	10	85	65	40

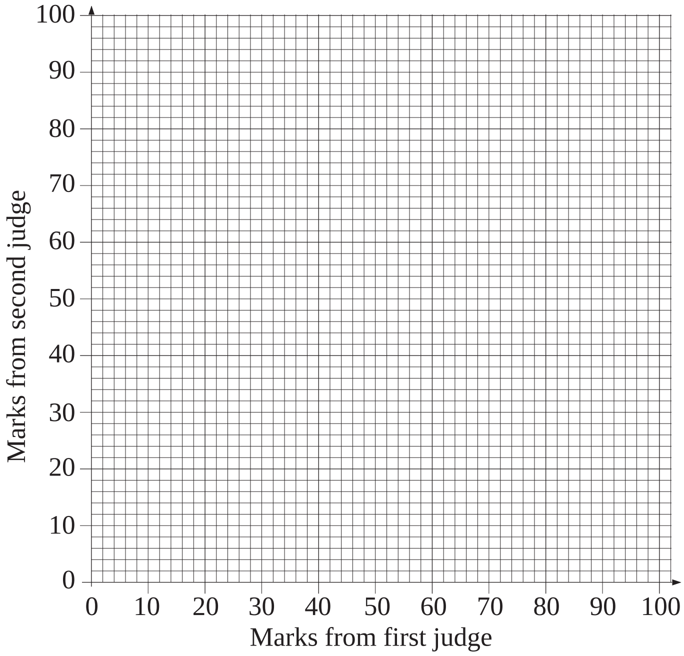
The table shows the marks that the two judges gave to each of the competitors.

(a) (i) On a copy of the grid on the following page, draw a scatter diagram to show this information.

(ii) Draw a line of best fit.

(b) A late entry was given 75 marks by the first judge.

Use your scatter diagram to estimate the mark that might have been given by the second judge. (Show how you found your answer.)

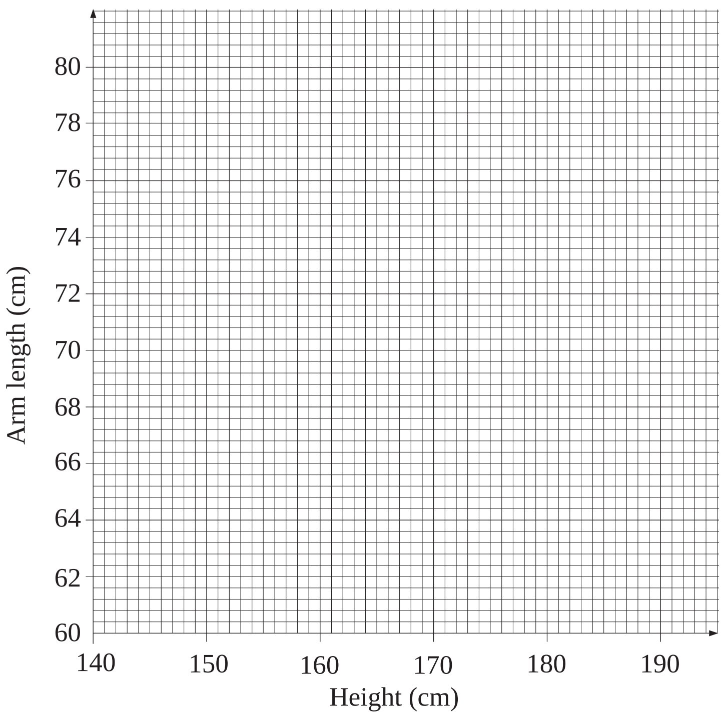


(NEAB)

11. The height and arm length for each of eight pupils are shown in the table.

Height (cm)	169	176	182	157	166	188	154	190
Arm length (cm)	67	73	70	63	69	74	62	77

(a) On a copy of the following grid, plot a scatter graph for these data.



(b) (i) Peter gives his height as 171 cm.
Use the scatter graph to estimate Peter's arm length.

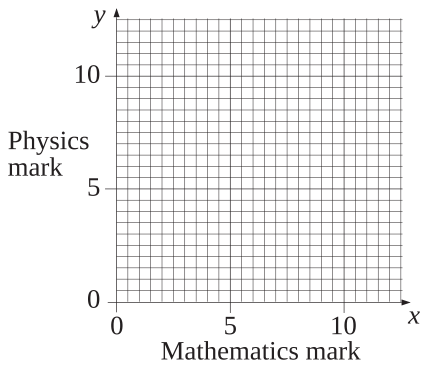
(ii) Explain why your answer can only be an estimate.

(SEG)

3.5

12. A group of schoolchildren took a Mathematics test and a Physics test. The results for 12 children are shown in the table.

Mathematics mark	1	1.5	3	4	5	5	6	8	9	9	10	10
Physics mark	3.5	1.5	5	4	2	7	5	6	7	9	7	9



(a) On a copy of the grid, draw a scatter diagram for the information in the table.

(b) Does the scatter diagram show the results you would expect? Explain your answer.

(c) (i) Add a line of best fit, by inspection, to the scatter diagram.

(ii) One pupil scored 7 marks for Mathematics but missed the Physics test.

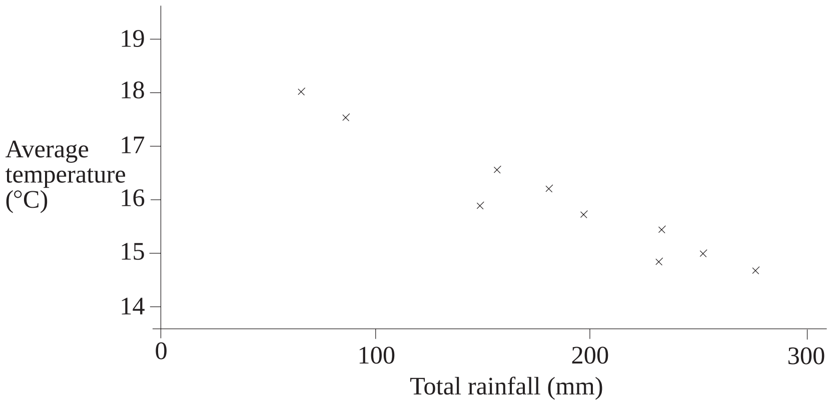
Use the line of best fit to estimate the mark she might have score for Physics.

(iii) One pupil was awarded the prize for the best overall performance in Mathematics and Physics.

Put a ring around the cross representing that pupil on the scatter diagram.

13. Megan wanted to find out if there is a connection between the average temperature and the total rainfall in the month of August.

She obtained weather records for the last 10 years and plotted a scatter graph.



(a) What does the graph show about a possible link between temperature and rainfall in August?

(b) Estimate the total rainfall in August when the average temperature is 17 °C.

3.6 Equations of Lines of Best Fit

Worked Example 1

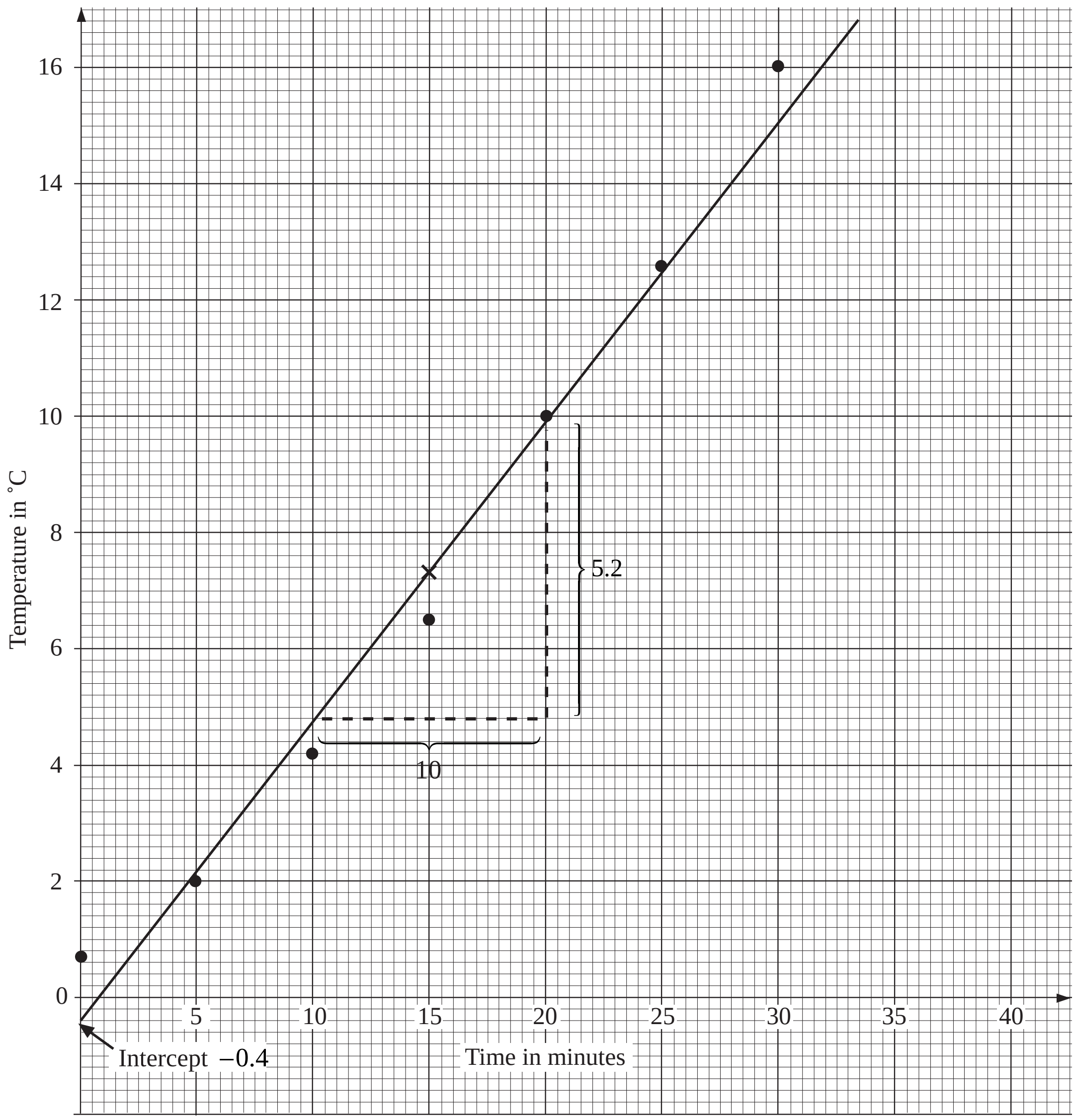
An electric heater was switched on in a cold room and the temperature of the room was taken at 5 minute intervals. The results were recorded and plotted on the following graph.

(a) Given that $\bar{x} = 15$ and $\bar{y} = 7.4$, draw a line of best fit for these data.

(b) Obtain the equation of this line of best fit in the form $y = mx + c$, stating clearly your values of m and c.

(c) Use your equation to predict the temperature of the room 40 minutes after switching on the fire,

(d) Give *two* reasons why this result may not be reliable.



Solution

(a) See diagram above.

(b) Intercept $c = -0.4$ and slope, $m = \dfrac{10 - 4.8}{10}$, (see triangle drawn on graph)

i.e. $m = \dfrac{5.2}{10} = 0.52$,

so $\boxed{y = 0.52x - 0.4}$

An alternative approach would be to note the intercept $c = -0.4$ from the diagram, so that

$$y = mx - 0.4$$

To pass through the point $\bar{x} = 15$, $\bar{y} = 7.4$ means that

$$\begin{aligned} 7.4 &= 15m - 0.4 \\ 15m &= 7.4 + 0.4 \\ m &= \frac{7.8}{15} = 0.52 \end{aligned}$$

giving the equation

$$\boxed{y = 0.52x - 0.4}$$

(c) Predicted temperature $= 0.52 \times 40 - 0.4$

$= 20.4\ ^\circ\text{C}$

(d) The value of 40 minutes is outside the range of values on which the line of regression is based; the heater may not continue to raise the temperature if it has a thermostat on it.

Worked Example 2

Mr Bean often travels by taxi and has to keep details of the journeys in order to complete his claim form at the end of the week. Details for journeys made during a week are:

Distance travelled (miles)	2	7	$8\frac{1}{2}$	11	6	3	$4\frac{1}{2}$
Cost (£)	3.00	5.40	6.10	7.40	5.00	3.20	4.20

(a) On graph paper, plot the above points.

(b) Calculate the mean point of these data and use this line to draw the line of best fit on your graph.

(c) Obtain the equation of your line of best fit in the form $y = mx + c$.

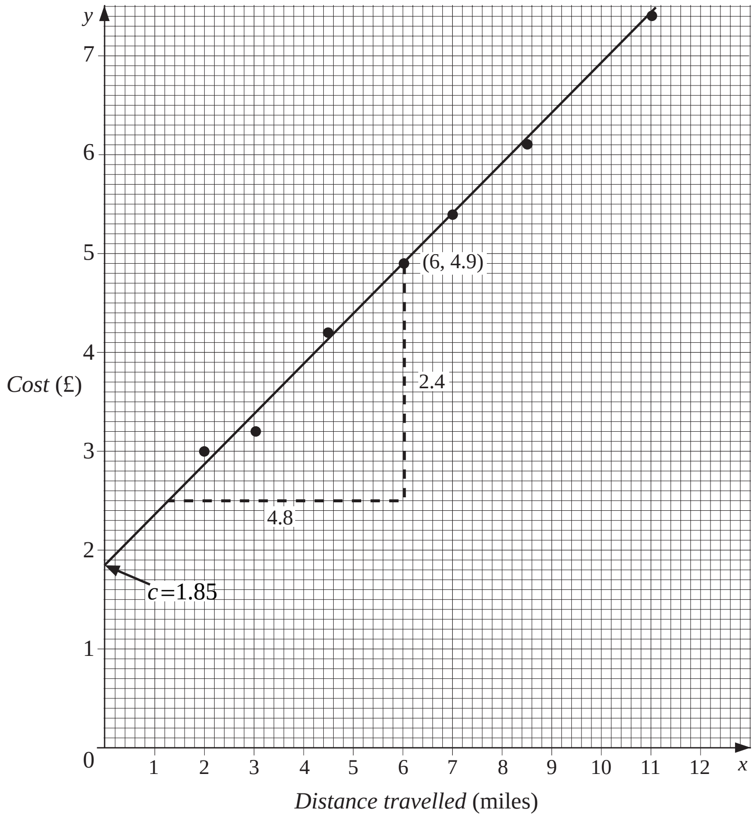
(d) Give an interpretation for the values of c and m in your calculation.

(SEG)

3.6

Solution

(a)



(b) Distance, $\bar{x} = 6$; cost, $\bar{y} = 4.9$

(c) From the intercept, $c = 1.85$, and from the construction, the gradient

$$m = \frac{2.4}{4.8} = 0.5.$$

Thus

$$y = 0.50\,x + 1.85$$

(d) *c* is the standard charge (£1.85) for using the taxi.

m is the amount (50p) charged per mile.

3.6

Exercises

1. The following data relate to the age and weight of ten randomly chosen children in Bedway Primary School.

Age (years)	7.8	8.1	6.4	5.2	7.0	9.9	8.4	6.0	7.2	10.0
Weight (kg)	29	28	26	20	24	35	30	22	25	36

(a) Draw a scatter diagram to show this information.

The mean age of this group of children is 7.6 years.

(b) Calculate the mean weight of this group.

(c) On the graph, draw the line of best fit.

(d) Use your graph to find the equation of this line of best fit, in the form $y = mx + c$.

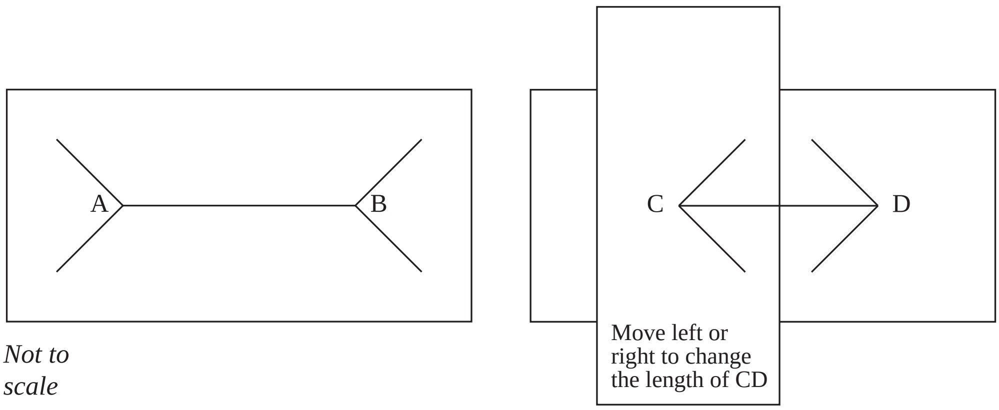
Jane is a pupil at Bedway Primary School and her age is 8.0 years.

(e) Use your answer to (d) to estimate Jane's weight.

(f) Give *one* reason why a prediction of the weight of a twelve year old from your graph might not be reliable.

(SEG)

2.



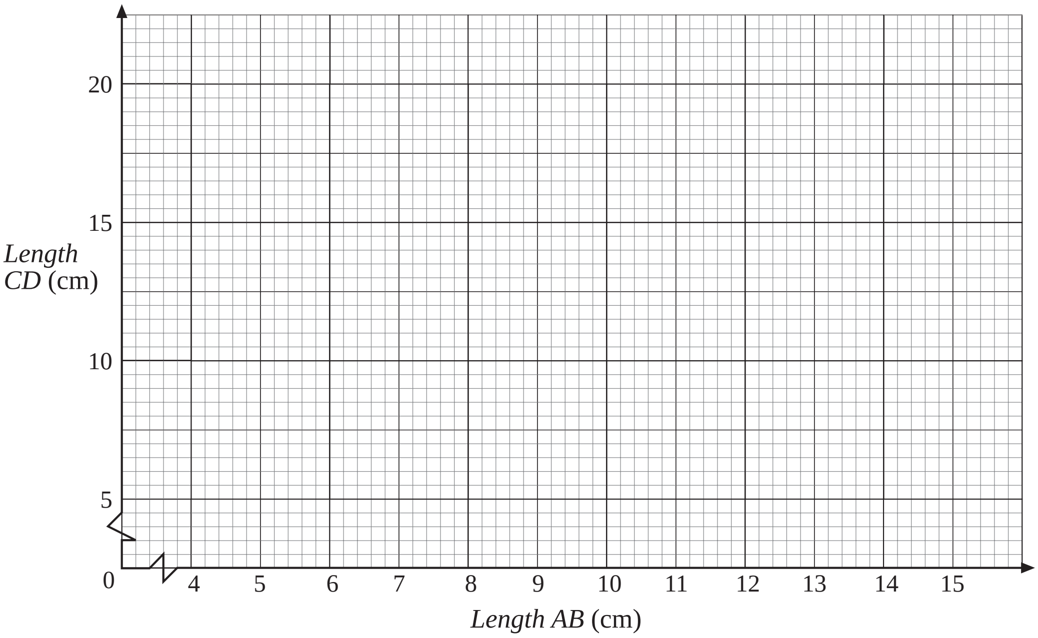
A student was shown the line AB and asked to adjust the length of the line CD so that the length of CD appeared to be the same as the length of AB.

The experiment was repeated using different lengths for AB and the results are given in the table. For example, when the line AB was 8 cm long the student made the length of CD equal to 10.3 cm.

AB cm	5	6	7	8	9	10	11	12
CD cm	6.5	7.8	8.6	10.3	11.4	12.7	14.2	15.1

3.6

(a) On a copy of the diagram below, plot a scatter graph to illustrate these results.



The mean length of CD is 10.8 cm.

(b) Calculate the mean length of AB.

(c) Draw a line of best fit on the scatter graph.

(d) Calculate the equation of your line of best fit.

(e) Use your equation to estimate the length of CD when the length of AB is 6.5 cm.

3.7 Moving Averages

Time series analysis is a method of using past data to predict future values. It is based on the assumption that there is an underlying trend to the data; for example, constant growth or *constant* decay, although, due to fluctuations, the actual data will not follow an exact pattern. To predict future values, we use the concept of a *moving average.* This will be illustrated in the following examples.

Worked Example 1

Below is a table showing the quarterly electricity bills paid for a school during a period of three years. The Headteacher, in a drive to save money, wishes to display these data in order to show the rising cost over the years.

		Year		
		1987	1988	1989
Quarter	1st	£9000	£9550	£9990
	2nd	£5000	£5450	£6100
	3rd	£4000	£4850	£5350
	4th	£8900	£9850	£10400

(a) On graph paper, plot the above figures joining the points with straight lines.

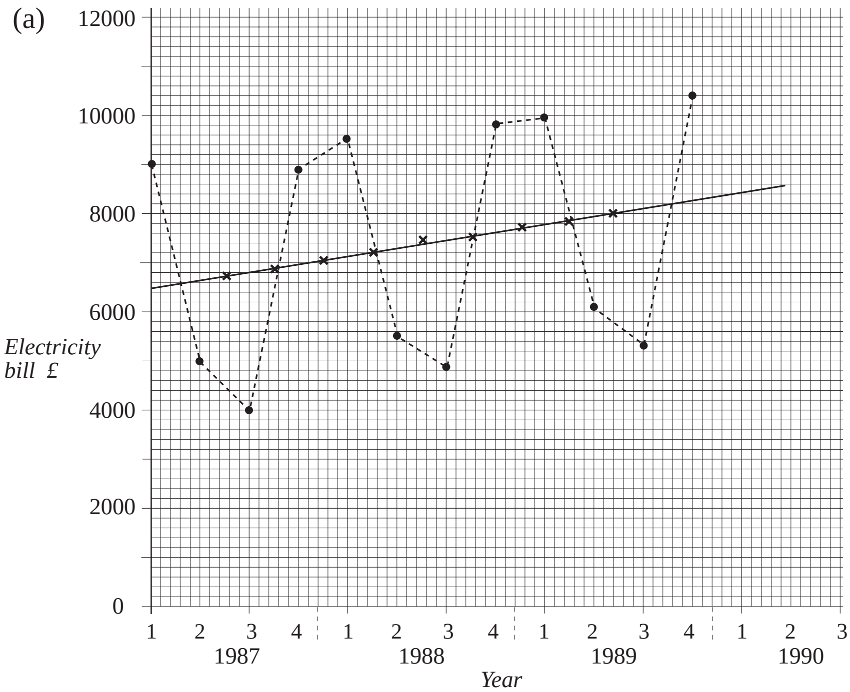
(b) Suggest a reason for the seasonal variation shown by your graph.

(c) (i) Calculate suitable moving averages for these data and plot these values on your graph.

(ii) What is the purpose of plotting moving averages? *(SEG)*

Solution

(a)



This diagram shows a *trendline.* This is a line of best fit for the moving average data.

3.7

(b) Cold, dull winter weather – more electricity used for heating and lighting.
Warmer, brighter weather – less lighting and heating needed so less electricity used.

(c) (i) 4-point moving averages are calculated below.

Year	*Quarter*	*Bill paid* £	*FOUR-point moving average*
1987	1	9000	
	2	5000	
			6725.0
	3	4000	
			6862.5
	4	8900	
			6975.0
1988	1	9550	
			7187.5
	2	5450	
			7425.0
	3	4850	
			7512.5
	4	9850	
			7675.0
1989	1	9900	
			7800.0
	2	6100	
			7937.5
	3	5350	
	4	10400	

(ii) Plotting moving averages allows you to identify the underlying trend. In this case, even though there are seasonal variations, the underlying trend shows a steady increase in the school's electricity bill year on year.

Worked Example 2

The following data give the quarterly sales (in thousands) of the Koala model of a car over a period of 4 years.

	1st	2nd	3rd	4th
1990	30	28	27	20
1991	27	26	26	18
1992	24	22	22	15
1993	19	19	18	10

(a) Plot these values on a graph.

(b) Suggest a reason for the seasonal variation shown by your graph.

(c) Calculate appropriate moving averages for these data and plot these values on your graph.

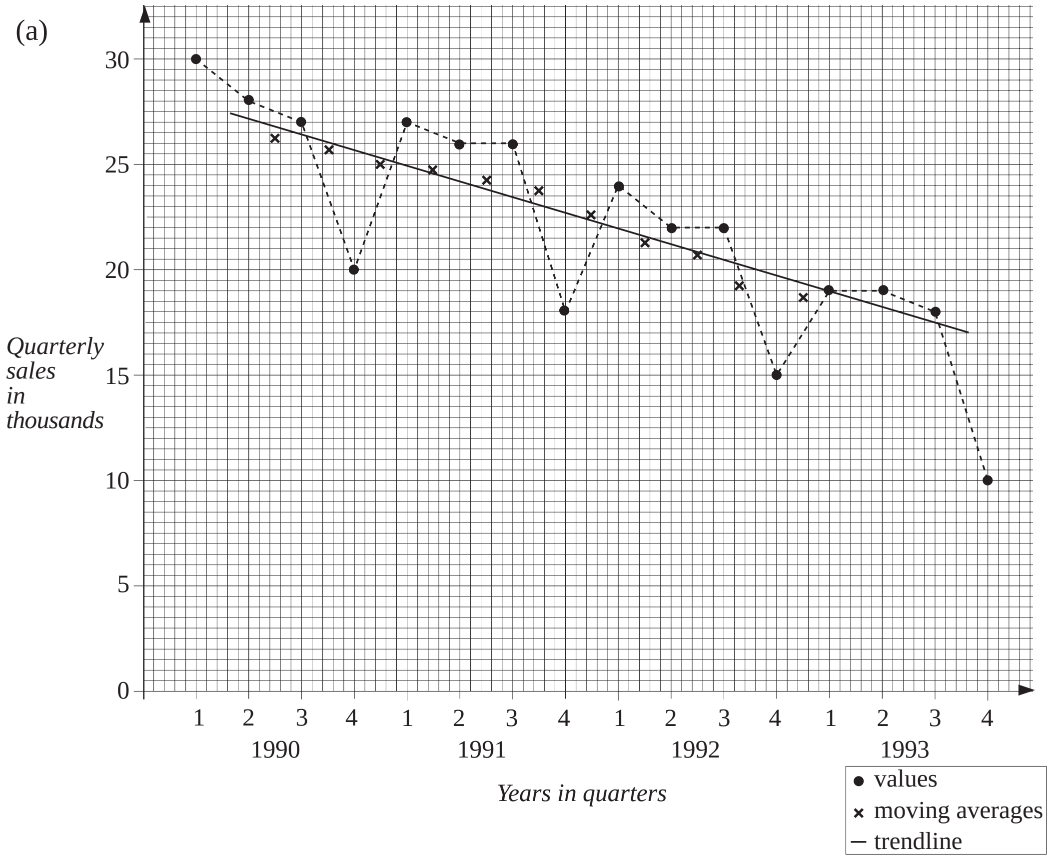
(d) Comment on the underlying trend of the data.

(SEG)

3.7

Solution

(a)



(b) Fewer people buy cars near the Christmas period.

(c) Using 4-point moving averages, we obtain the data shown opposite.

	Quarter	Sales	Moving Average
1990	1	30	
	2	28	
			26.25
	3	27	
			25.50
	4	20	
			25.00
1991	1	27	
			24.75
	2	26	
			24.25
	3	26	
			23.50
	4	18	
			22.50
1992	1	24	
			21.50
	2	22	
			20.75
	3	22	
			19.50
	4	15	
			18.75
1993	1	19	
			17.75
	2	19	
			16.50
	3	18	
	4	10	

(d) Even though the sales have fluctuated, the underlying trend shows that, over the four year period, there has been a decline in sales of the Koala car.

3.7

Exercises

1. A school canteen manager notes the quarterly turnover as follows.

	Jan-Mar	**Apr-June**	**July-Sept**	**Oct-Dec**
1992	£7500	£5500	£2000	£8200
1993	£6700	£4300	£1600	£8200
1994	£5100	£3900	£1200	£7500

(a) Which year has the biggest turnover?

(b) Give a reason why the third quarter is the lowest in each year.

(c) Using graph paper, plot the above figures for the school canteen, joining the dots with straight lines.

(d) (i) Calculate the appropriate 4-point moving averages for these data on a copy of the following table.

Quarter	**Turnover**	**Moving average**
1	7500	
2	5500	
3	2000	
4	8200	
1		
2		
3		
4		
1		
2		
3		
4		

(ii) Plot these values on your graph and add a trend line by eye.

(iii) What does the trend line tell you about the canteen's turnover?

(SEG)

2. The following data give the quarterly sales, in £10 000's, of gardening equipment at the Green Fingers Garden Centre over a period of four years.

	Quarter			
	1st	2nd	3rd	4th
1992	20	26	24	18
1993	24	30	27	23
1994	26	34	31	25
1995	30	36	35	29

(a) Plot these values on a graph joining the points with straight lines.

(b) Suggest a reason for the seasonal variation shown by your graph.

(c) Calculate the four-point moving averages for these data and enter these values on a copy of the table.

Year	**Quarter**	**Sales £10 000's**	**Four-point moving average**
1992	1	20	
	2	26	
	3	24	
	4	18	
1993	1	24	
	2	30	
	3	27	
	4	23	
1994	1	26	
	2	34	
	3	31	
	4	25	
1995	1	30	
	2	36	
	3	35	
	4	29	

(d) Plot these moving averages on the graph.

(e) On your graph, draw a trend line by eye.

(f) Use your graph to estimate the sales during the first quarter of 1996.

(SEG)

3. The table gives the number of passengers per quarter on all United Kingdom airline services in 1991 and 1992. The numbers are in millions.

Year	Quarter	Passengers (millions)
1991	1	3.0
	2	4.0
	3	5.2
	4	3.8
1992	1	3.4
	2	4.4
	3	5.3
	4	4.0

(a) Draw a carefully labelled time series graph of this information.

(b) Describe three seasonal variations in numbers of passengers shown by your graph.

(c) A newspaper reported that in July 1993 there were 1 800 000 passengers on all United Kingdom airline services.

Explain why you think this figure may be wrong.

(NEAB)

4. The table shows the amount, in £1000s, taken by a department store over a 15-day period during the run-up to Christmas one year, beginning on Monday 6 December.

	Mon	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*
Week beginning 6/12	52	46	47	59	52	61
Week Beginning 13/12	55	49	51	62	57	63
Week beginning 20/12	57	53	54			

(a) Plot a graph showing the daily amount taken by the store.

(b) Explain why you think the data shows that the store may be staying open for late-night shopping on Thursdays during this period.

(c) Calculate the 6-day moving average for this data.

(d) Explain why a 6-point moving average is appropriate in this case.

(e) Plot the moving averages on your graph and draw a trend line.

(f) Describe what is shown by the graph of the moving average.

(g) Estimate the amount taken that year on Thursday 23 December.

(h) Explain why you think it may be unwise to use moving averages to predict the amounts taken by the store on

(i) Friday 24 December (ii) Saturday 25 December.

3.8 Cumulative Frequency

Cumulative frequencies are useful if more detailed information is required about a set of data. In particular, they can be used to find the median and inter-quartile range.

The *inter-quartile range* contains the middle 50% of the sample and is a measure of how spread out the data are. This is illustrated in Worked Example 2.

Worked Example 1

The heights of 112 children aged 5 to 11 are recorded in this table. Draw up a cumulative frequency table and then draw a cumulative frequency polygon.

Height (cm)	*Frequency*
$90 < h \le 100$	5
$100 < h \le 110$	22
$110 < h \le 120$	30
$120 < h \le 130$	31
$130 < h \le 140$	18
$140 < h \le 150$	6

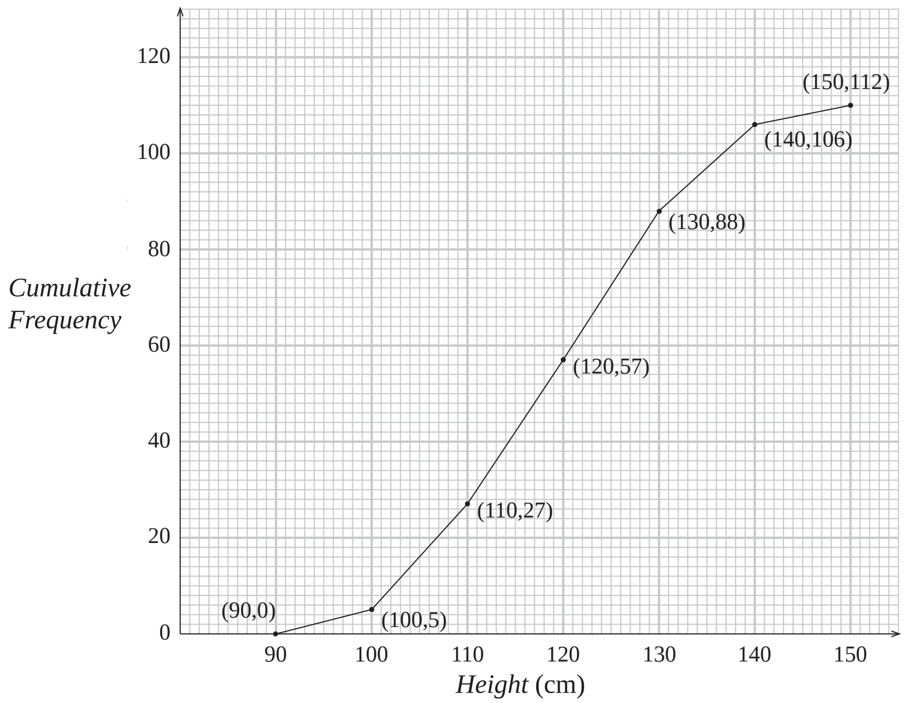
Solution

The table below shows how to calculate the cumulative frequencies.

Height (cm)	*Frequency*	*Cumulative Frequency*
$90 < h \le 100$	5	5
$100 < h \le 110$	22	$5 + 22 = 27$
$110 < h \le 120$	30	$27 + 30 = 57$
$120 < h \le 130$	31	$57 + 31 = 88$
$130 < h \le 140$	18	$88 + 18 = 106$
$140 < h \le 150$	6	$106 + 6 = 112$

Note
Cumulative frequency 27 means that there are 27 values in the interval $90 < h \le 110$. In the same way, 57 values are at most 120 cm, so we plot 57 against 120 cm on the graph.

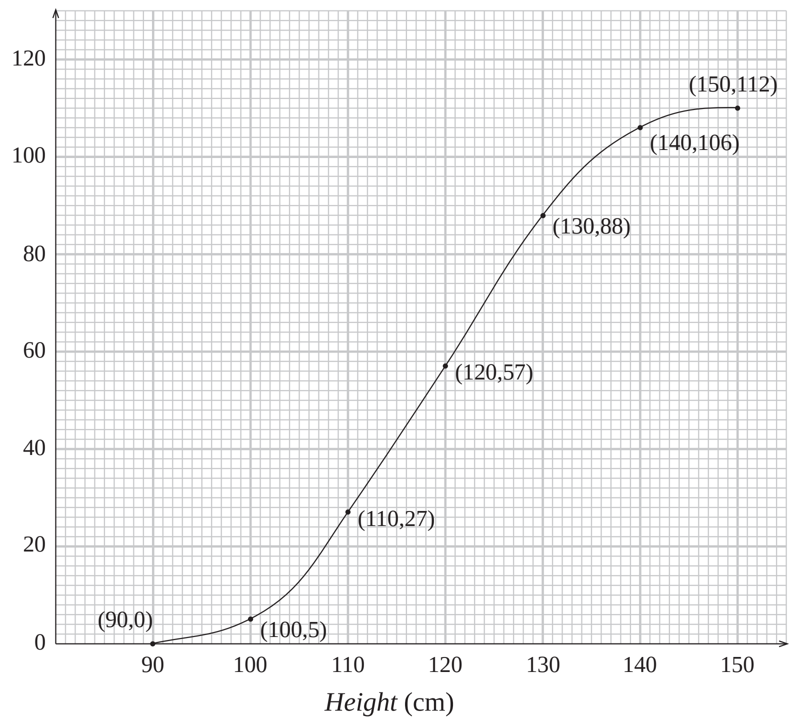
A graph can then be plotted using points as shown below.



3.8

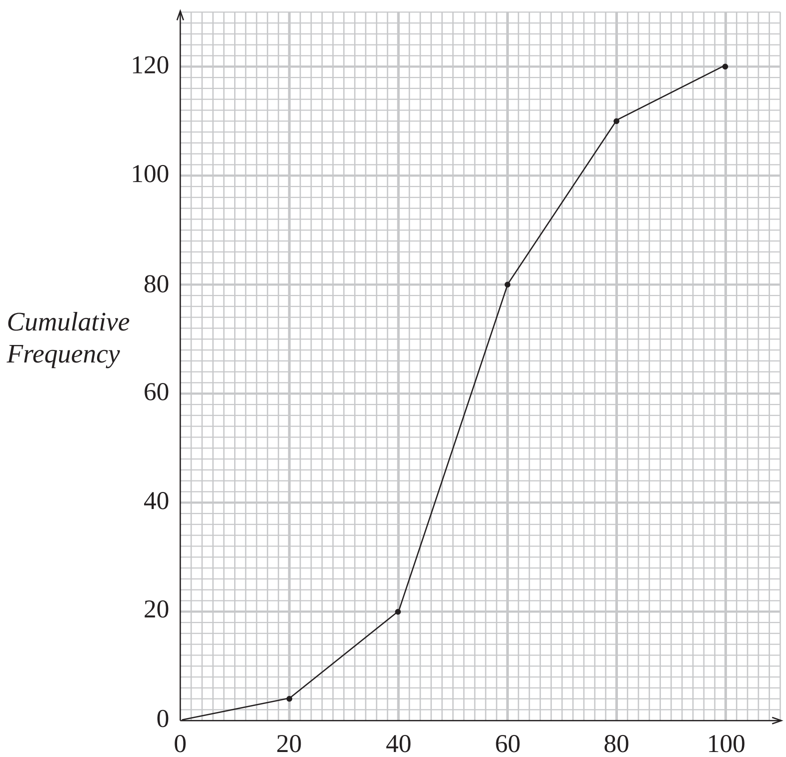
Note

If the table in Worked Example 1 gives the frequencies *in thousands* for each height group in a particular town, then we represent this by a *cumulative frequency curve*, showing the distribution for that whole population.



Worked Example 2

The cumulative frequency graph below gives the results of 120 students on a test.



3.8

Use the graph to find:

(a) the median score,

(b) the inter-quartile range,

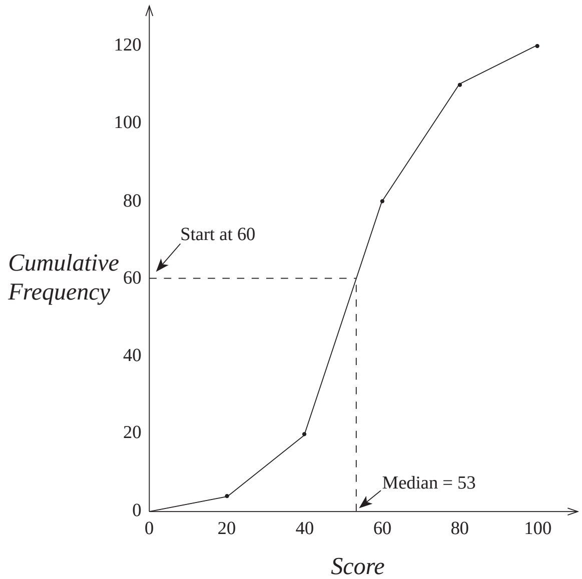
(c) the mark which was attained by only 10% of the students,

(d) the number of students who scored more than 75 on the test.

Solution

(a) Since $\frac{1}{2}$ of 120 is 60, the *median* can be found by starting at 60 on the vertical scale, moving horizontally to the graph line and then moving vertically down to meet the horizontal scale.

In this case the median is 53.



(b) To find out the *inter-quartile range*, we must consider the middle 50% of the students.

To find the *lower quartile*, start at $\frac{1}{4}$ of 120, which is 30.

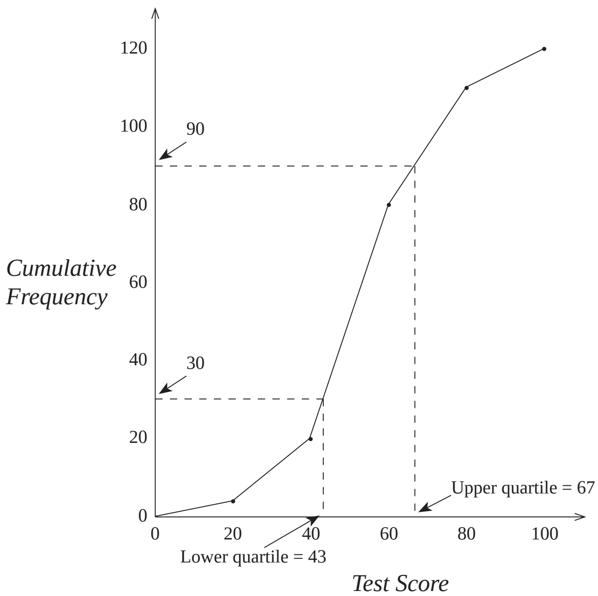
This gives

$$\text{Lower Quartile} = 43$$

To find the *upper quartile*, start at $\frac{3}{4}$ of 120, which is 90.

This gives

$$\text{Upper Quartile} = 67$$

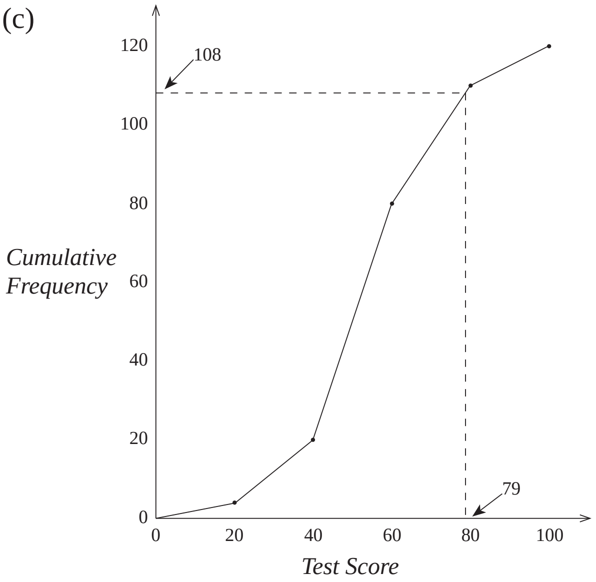


The *inter-quartile range* is then

$$\begin{aligned}\text{Inter-quartile Range} &= \text{Upper Quartile} - \text{Lower Quartile}\\ &= 67 - 43\\ &= 24\end{aligned}$$

3.8

(c)



Here the mark which was attained by the top 10% is required.

$$10\% \text{ of } 120 = 12$$

so start at 108 on the cumulative frequency scale.

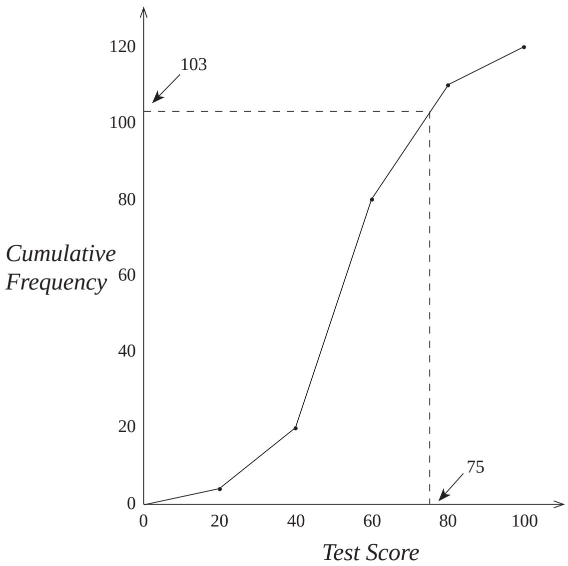
This gives a mark of 79, so the top 10% scored 79 or higher on this test.

(d) To find the number of students who scored more than 75, start at 75 on the horizontal axis.

This gives a cumulative frequency of 103.

So the number of students with a score greater than 75 is

$$120 - 103 = 17$$



To find the median in a small sample we should use the $\dfrac{n+1}{2}$ rule (see Section 3.3 on Mean, Median and Mode). However, for this example the difference between using the $60\frac{1}{2}$th value and the 60th value is minimal and so the method shown here is appropriate for this and all other large samples. (This also applies to the finding of quartiles, percentiles, deciles, etc.)

3.8

Exercises

1. Make a cumulative frequency table for each set of data given below. Then draw a cumulative frequency graph and use it to find the median and inter-quartile range.

   (a) John weighed each apple in a large box. His results are given in this table.

Weight of apple (g)	$60 < w \le 80$	$80 < w \le 100$	$100 < w \le 120$	$120 < w \le 140$	$140 < w \le 160$
Frequency	4	28	33	27	8

   (b) Pasi asked the students in his class how far they travelled to school each day. His results are given below.

Distance (km)	$0 < d \le 1$	$1 < d \le 2$	$2 < d \le 3$	$3 < d \le 4$	$4 < d \le 5$	$5 < d \le 6$
Frequency	5	12	5	6	5	3

   (c) A P.E. teacher recorded the distances children could reach in the long jump event. His records are summarised in the table below.

Length of jump (m)	$1 < d \le 2$	$2 < d \le 3$	$3 < d \le 4$	$4 < d \le 5$	$5 < d \le 6$
Frequency	5	12	5	6	5

2. A farmer grows a type of wheat in two different fields. He takes a sample of 50 heads of corn from each field at random and weighs the grains he obtains.

Mass of grain (g)	$0<m\le5$	$5<m\le10$	$10<m\le15$	$15<m\le20$	$20<m\le25$	$25<m\le30$
Frequency Field A	3	8	22	10	4	3
Frequency Field B	0	11	34	4	1	0

   (a) Draw cumulative frequency graphs for each field.

   (b) Find the median and inter-quartile range for each field.

   (c) Comment on your results.

3. A consumer group tests two types of batteries using a personal stereo.

Lifetime (hours)	$2 < l \le 3$	$3 < l \le 4$	$4 < l \le 5$	$5 < l \le 6$	$6 < l \le 7$	$7 < l \le 8$
Frequency Type A	1	3	10	22	8	4
Frequency Type B	0	2	2	38	6	0

(a) Use cumulative frequency graphs to find the median and inter-quartile range for each type of battery.

(b) Which type of battery would you recommend and why?

4. The table below shows how the height of girls of a certain age vary. The data was gathered using a large-scale survey.

Height (cm)	$50 < h \le 55$	$55 < h \le 60$	$60 < h \le 65$	$65 < h \le 70$	$70 < h \le 75$	$75 < h \le 80$	$80 < h \le 85$
Frequency	100	300	2400	1300	700	150	50

A doctor wishes to be able to classify children as:

Category	*Percentage of Population*
Very Tall	5%
Tall	15%
Normal	60%
Short	15%
Very short	5%

Use a cumulative frequency graph to find the heights of children in each category.

5. The manager of a double glazing company employs 30 salesmen. Each year he awards bonuses to his salesmen.

Bonus	*Awarded to*
£500	Best 10% of salesmen
£250	Middle 70% of salesmen
£ 50	Bottom 20% of salesmen

The sales made during 1995 and 1996 are shown in the table below.

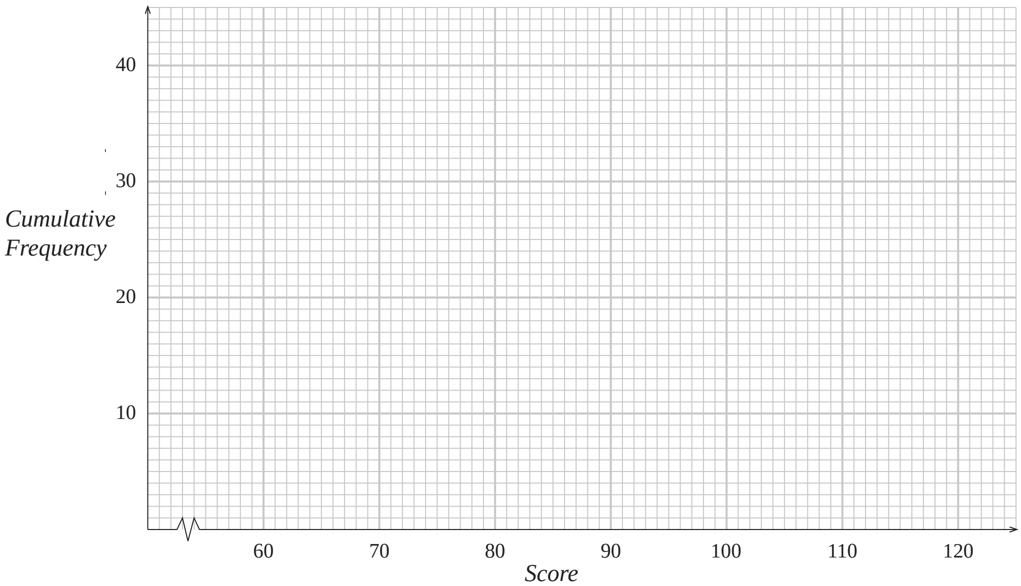
Value of sales (£1000)	$0 < V \le 100$	$100 < V \le 200$	$200 < V \le 300$	$300 < V \le 400$	$400 < V \le 500$
Frequency 1995	2	8	18	2	0
Frequency 1996	0	2	15	10	3

Use cumulative frequency graphs to find the values of sales needed to obtain each bonus in the years 1995 and 1996.

6. Laura and Joy played 40 games of golf together. The table below shows Laura's scores.

Scores (x)	$70 < x \le 80$	$80 < x \le 90$	$90 < x \le 100$	$100 < x \le 110$	$110 < x \le 120$
Frequency	1	4	15	17	3

(a) On a grid similar to the one below, draw a cumulative frequency diagram to show Laura's scores.



(b) Making your method clear, use your graph to find

(i) Laura's median score,

(ii) the inter-quartile range of her scores.

(c) Joy's median score was 103. The inter-quartile range of her scores was 6.

(i) Who was the more consistent player?
Give a reason for your choice.

(ii) The winner of a game of golf is the one with the lowest score.
Who was the better player? Give a reason for your choice.

(NEAB)

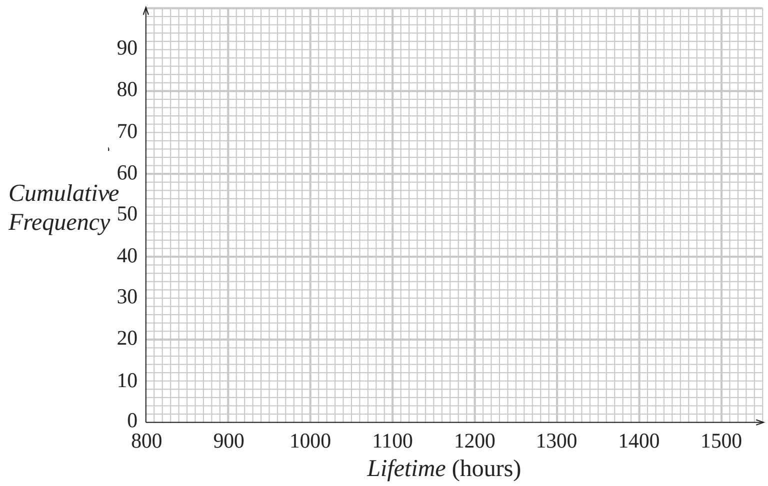
3.8

7. A sample of 80 electric light bulbs was taken. The lifetime of each light bulb was recorded. The results are shown below.

Lifetime (hours)	800–	900–	1000–	1100–	1200–	1300–	1400–
Frequency	4	13	17	22	20	4	0
Cumulative Frequency	4	17					

(a) Copy and complete the table of values for the cumulative frequency.

(b) Draw the cumulative frequency curve, using a grid as shown.



(c) Use your graph to estimate the number of light bulbs which lasted more than 1030 hours.

(d) Use your graph to estimate the inter-quartile range of the lifetimes of the light bulbs.

(e) A second sample of 80 light bulbs has the same median lifetime as the first sample. Its inter-quartile range is 90 hours. What does this tell you about the difference between the two samples?

(SEG)

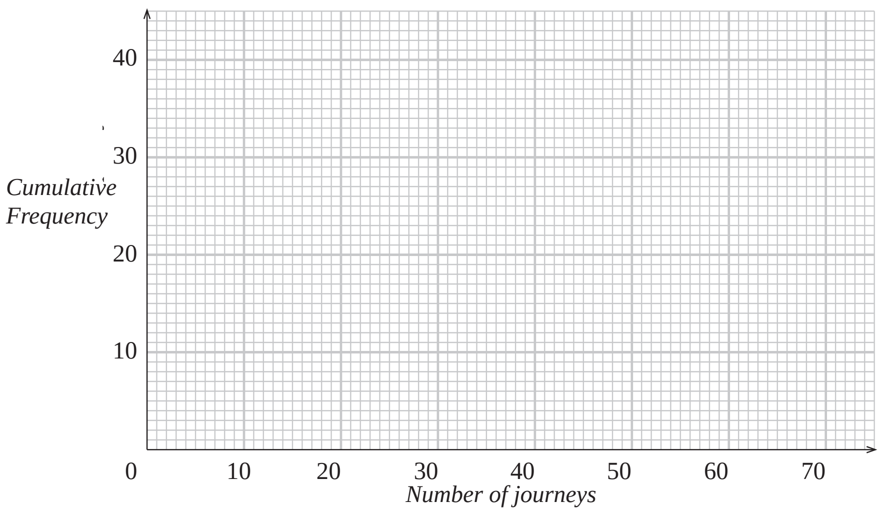
8. The numbers of journeys made by a group of people using public transport in one month are summarised in the table.

Number of journeys	0–10	11–20	21–30	31–40	41–50	51–60	61–70
Number of people	4	7	8	6	3	4	0

(a) Copy and complete the cumulative frequency table below.

Number of journeys	≤ 10	≤ 20	≤ 30	≤ 40	≤ 50	≤ 60	≤ 70
Cumulative frequency							

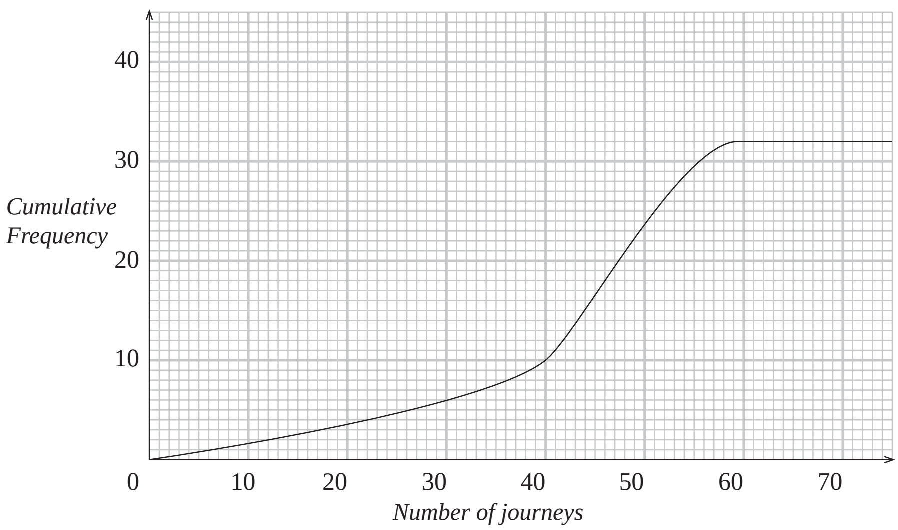
(b) (i) Draw the cumulative frequency graph, using a grid as below.



(ii) Use your graph to estimate the median number of journeys.

(iii) Use your graph to estimate the number of people who made more than 44 journeys in the month.

(c) The numbers of journeys made using public transport in one month, by another group of people, are shown in the graph.



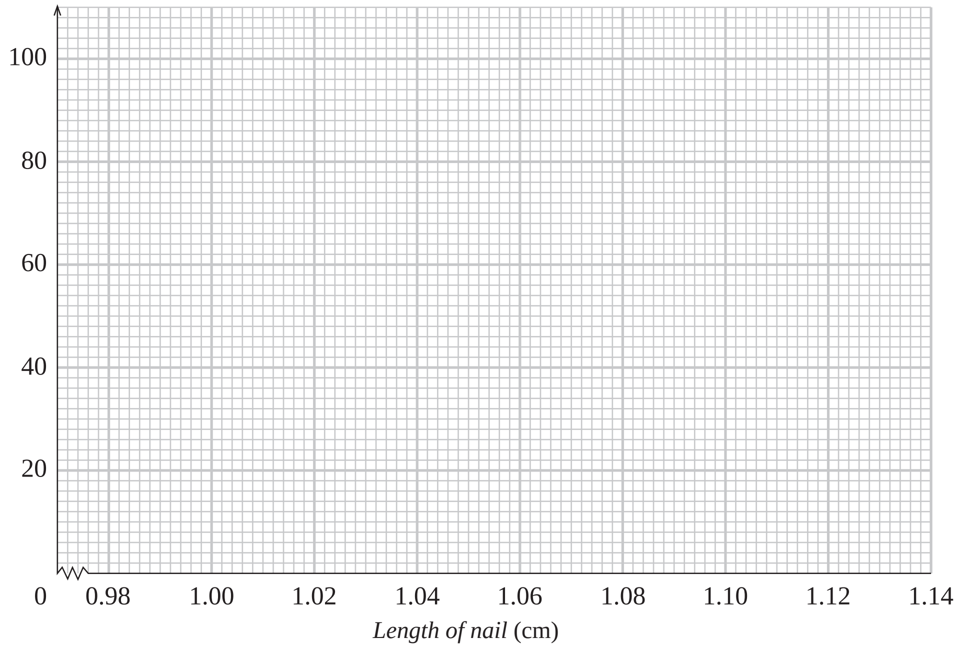
Make **one** comparison between the numbers of journeys made by these two groups. *(SEG)*

9. The lengths of a number of nails were measured to the nearest 0.01 cm, and the following frequency distribution was obtained.

Length of nail (x cm)	*Number of nails*	*Cumulative Frequency*
$0.98 \le x < 1.00$	2	
$1.00 \le x < 1.02$	4	
$1.02 \le x < 1.04$	10	
$1.04 \le x < 1.06$	24	
$1.06 \le x < 1.08$	32	
$1.08 \le x < 1.10$	17	
$1.10 \le x < 1.12$	7	
$1.12 \le x < 1.14$	4	

(a) Complete the cumulative frequency column.

(b) Draw a cumulative frequency diagram on a grid similar to the one below.



Use your graph to estimate

(i) the median length of the nails (ii) the inter-quartile range.

(NEAB)

10. A wedding was attended by 120 guests. The distance, d miles, that each guest travelled was recorded in the frequency table below.

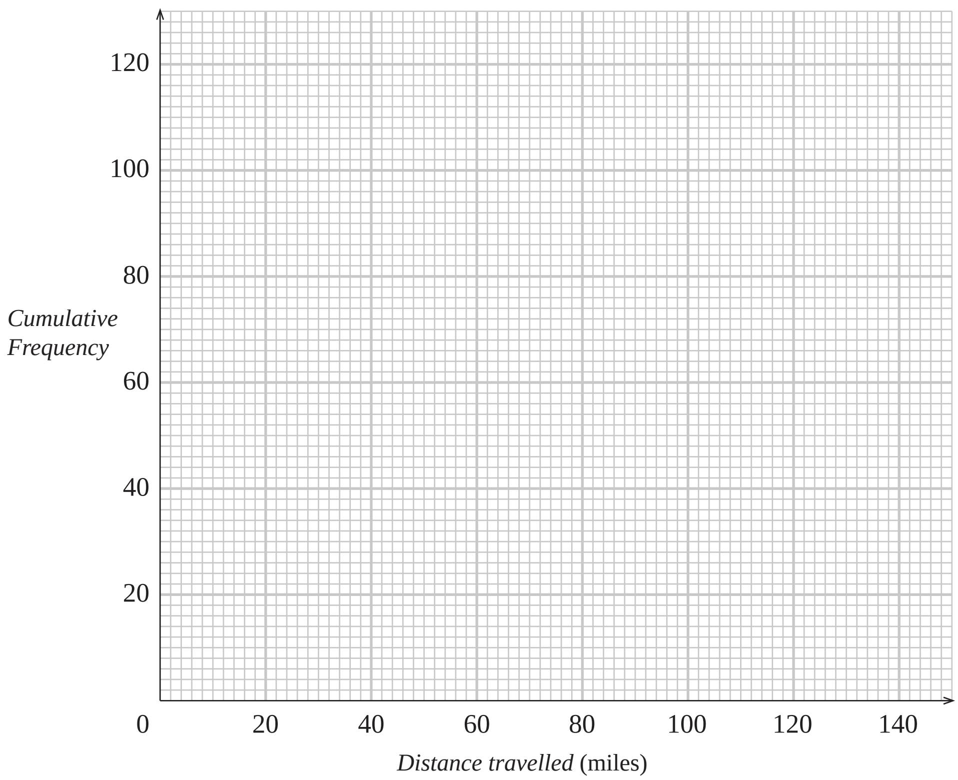
Distance (d miles)	$0 < d \leq 10$	$10 < d \leq 20$	$20 < d \leq 30$	$30 < d \leq 50$	$50 < d \leq 100$	$100 < d \leq 140$
Number of guests	26	38	20	20	12	4

(a) Using the mid-interval values, calculate an estimate of the mean distance travelled.

(b) (i) Copy and complete the cumulative frequency table below.

Distance (d miles)	$d \leq 10$	$d \leq 20$	$d \leq 30$	$d \leq 50$	$d \leq 100$	$d \leq 140$
Number of guests						120

(ii) On a grid similar to that shown below, draw a cumulative frequency curve to represent the information in the table.



(c) (i) Use the cumulative frequency curve to estimate the median distance travelled by the guests.

(ii) Give a reason for the large difference between the mean distance and the median distance.

(MEG)

3.9 Box and Whisker Plots

By *location*, we mean a measure which represents the *average* value – you have already met three key measures of location, namely

mean, *mode* and *median.*

They are all important and you should be familiar with their calculations and uses. You will need to revise these key topics because some of the concepts will be used in the following exercises.

As well as measures of location, you have also met measures of spread, that is, measures of how close the data is to the average value; for example,

range and *interquartile range.*

Again, you need to be familiar with finding and using these measures of spread, and you will need the range, interquartile range and median when illustrating the data with a *box and whisker* plot, which is the focus of this section. These diagrams are also referred to as *box plots.*

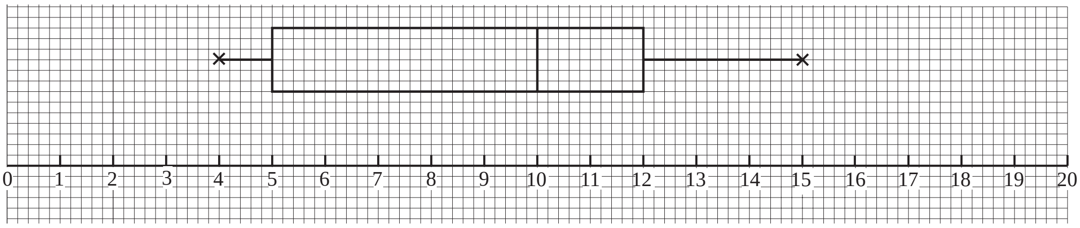
For example, for the data set below with 23 values, we can easily find the median and quartiles.

4, 4, 4, 4, 5, 5, 7, 9, 10, 10, 10, 10, 10, 11, 11, 11, 12, 12, 13, 14, 15, 15, 15

↑	↑	↑
lower quartile	*median*	*upper quartile*
6th value	12th value	18th value

The *box i*s formed by the two quartiles, with the median marked by a line, whilst the *whiskers are* fixed by the two extreme values, 4 and 15.

The plot is shown below, relative to a scale.



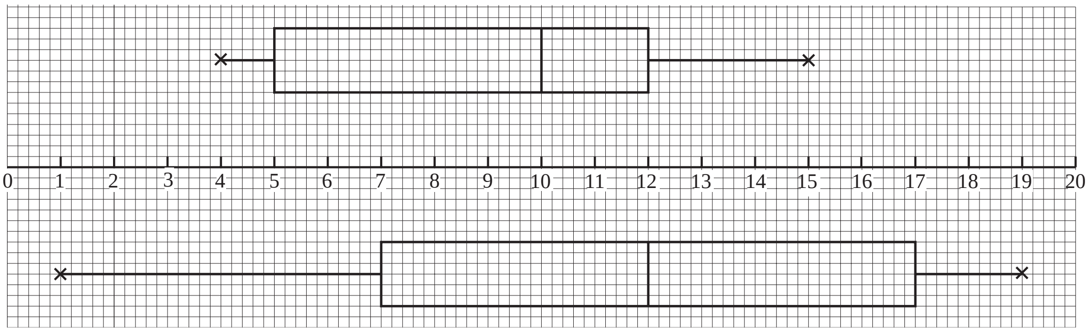
Box and whisker plots are particularly useful when comparing and contrasting two sets of data.

For example, if you wish to compare the data set above with the following data set with 15 values,

1, 1, 6, 7, 8, 9, 10, 12, 12, 14, 17, 17, 18, 19, 19

↑	↑	↑
lower quartile	*median*	*upper quartile*
4th value	8th value	12th value

then you can illustrate the two plots together. This is shown on the following diagram – you can immediately see that the data in the second set is much more spread out than that in the first set.



Worked Example 1

The number of goals scored by the 11 members of a hockey team in 1993 were as follows:

6 0 8 12 2 1 2 9 1 0 11

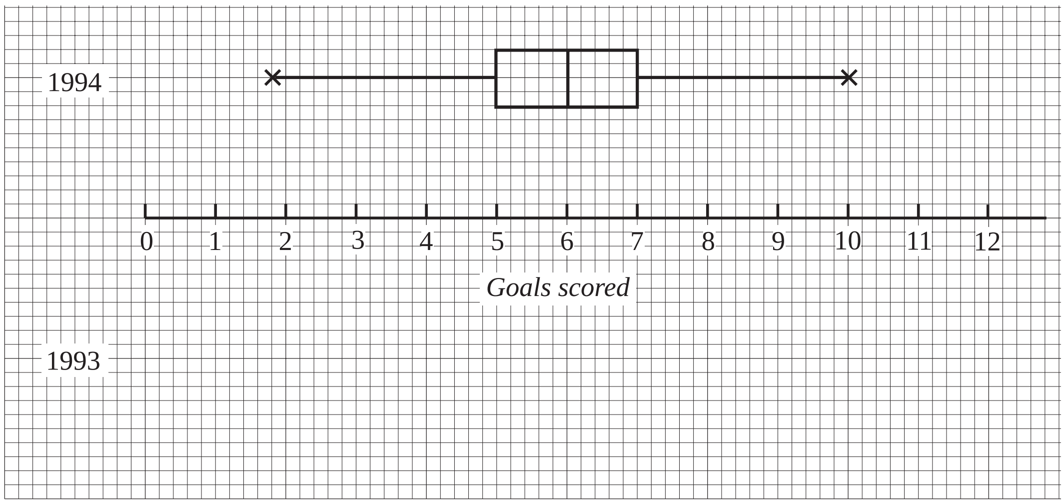
(a) Find the median.

(b) Find the upper and lower quartiles.

(c) Find the interquartile range.

(d) Explain why, for this data, the interquartile range is a more appropriate measure for spread than the range.

(e) The goals scored by the 11 members of the hockey team in 1994 are summarised in the box plot below.



(i) On a copy of the diagram above, summarise the results for 1993 in the same way.

(ii) Compare the team's goal scoring records for 1993 and 1994.

Solution

We first put the number of goals in increasing order, i.e:

0 0 1 1 2 2 6 8 9 11 12

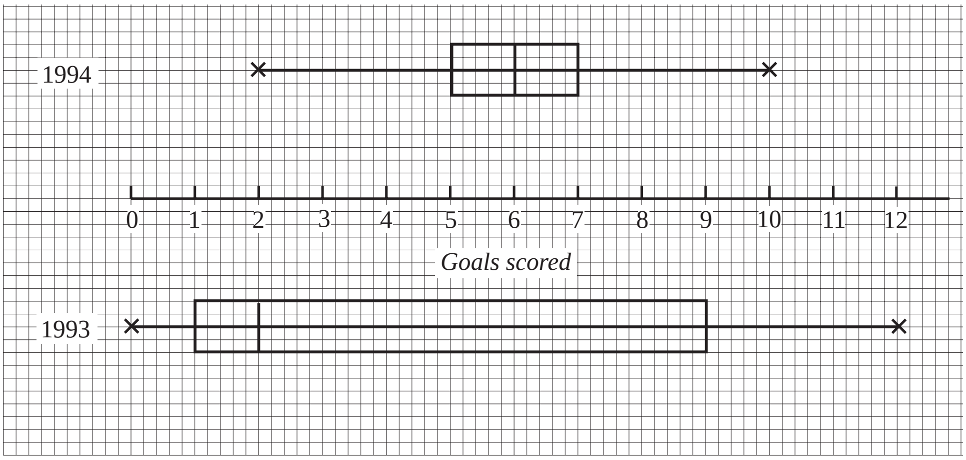
(a) There are 11 data points, so the median is the $\left(\dfrac{11+1}{2}\right)$th value, i.e. the 6th value, which is 2.

(b) The lower quartile is the $\left(\dfrac{11+1}{4}\right)$th value, i.e. the 3rd value, which is 1; the upper quartile is the 9th value, i.e. 9.

(c) Interquartile range $= 9 - 1 = 8$.

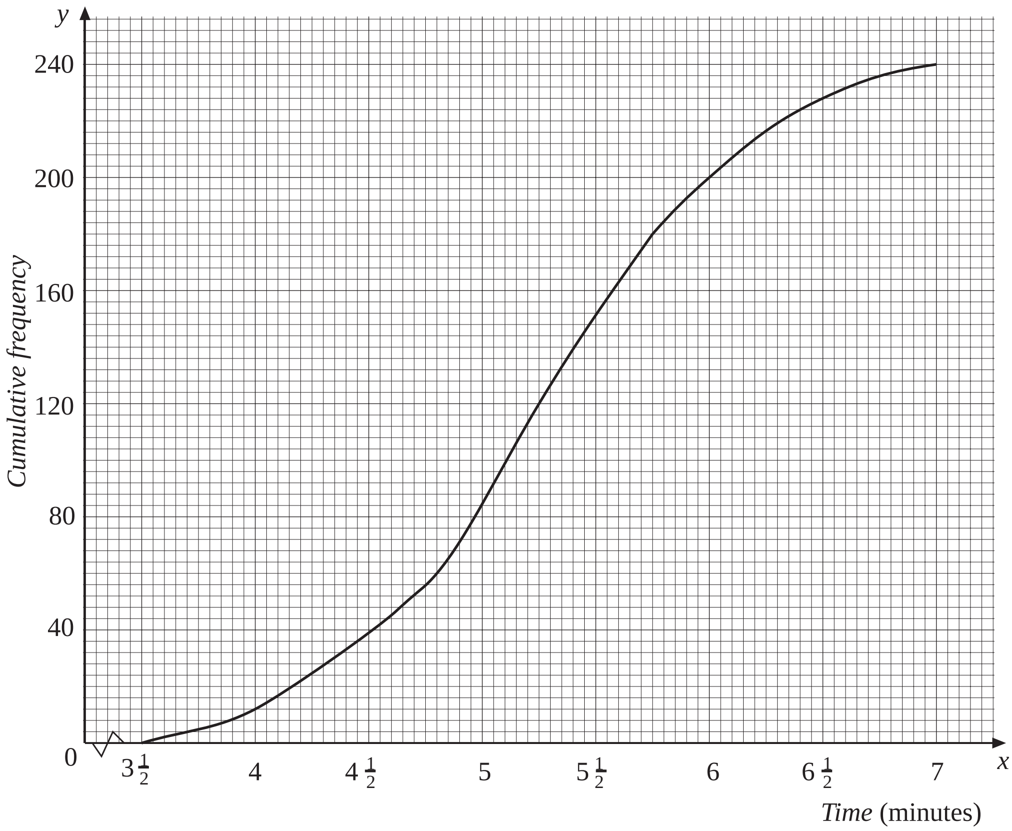
(d) The interquartile range is a better measure to represent the 'average' spread than the range because it excludes the outlying values. Values at the ends of the data range are often unrepresentative.

(e)



Comparing the median values (2 in 1993, 6 in 1994) shows that the players, on average, scored more goals in 1994. However, the box plot shows a far greater variation of scoring in 1993, with some players scoring more goals in 1993 than the highest number scored in 1994.

Worked Example 2



3.9

The cumulative frequency curve shown represents the times taken to run 1500 metres by each of the 240 members of the athletics club, Weston Harriers.

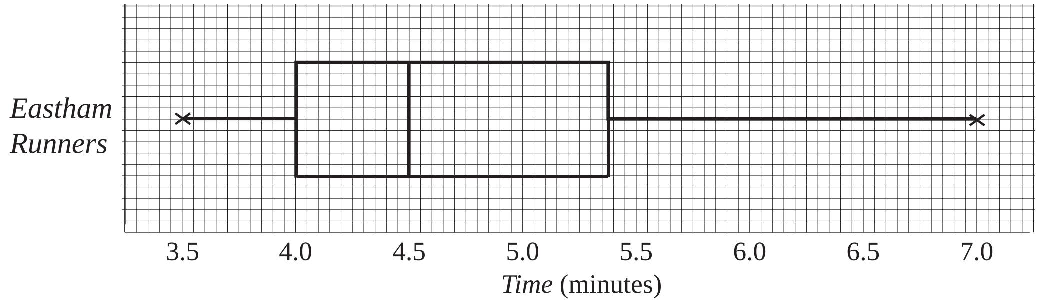
(a) From the graph, find:

(i) the median time; (ii) the upper quartile and the lower quartile.

(b) Draw a box and whisker diagram to illustrate the data.

(c) Use your box and whisker diagram to make *one* comment about the distribution of the data for Weston Harriers.

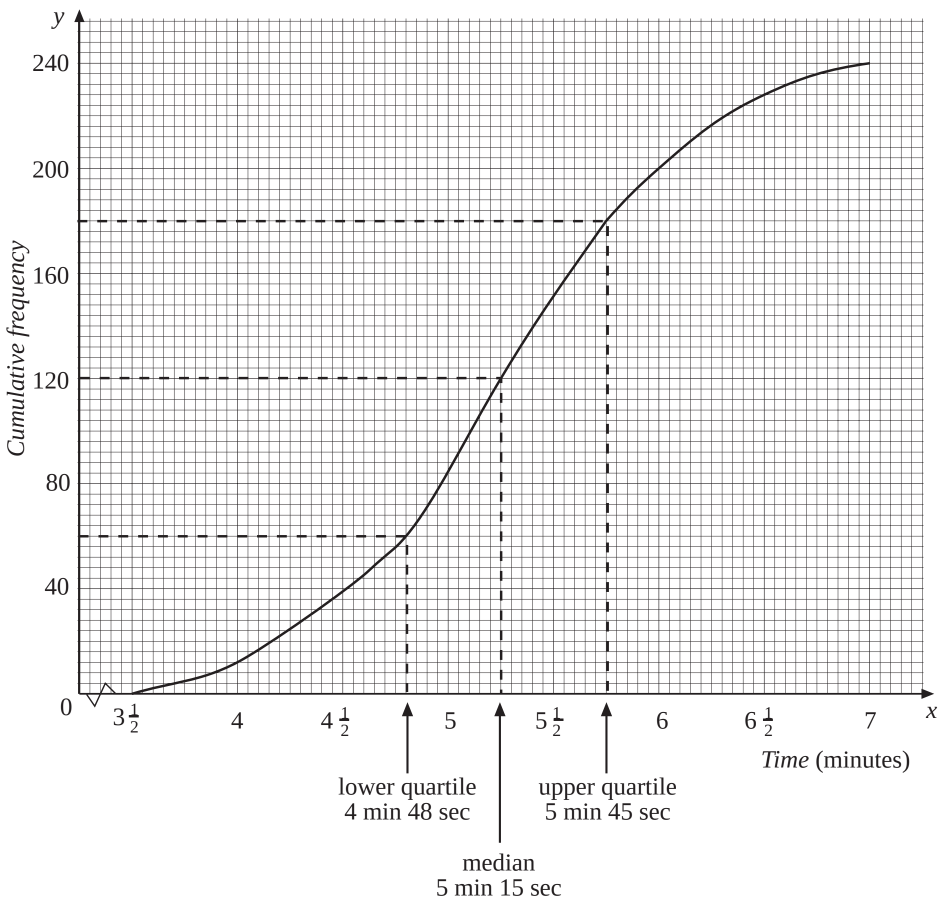
A rival athletics club, Eastham Runners, also has 240 members. The time taken by each member to run 1500 metres is recorded and these data are shown in the following box and whisker diagram.



(d) Use this diagram to make *one* comment about the data for Eastham Runners as compared to that for Weston Harriers.

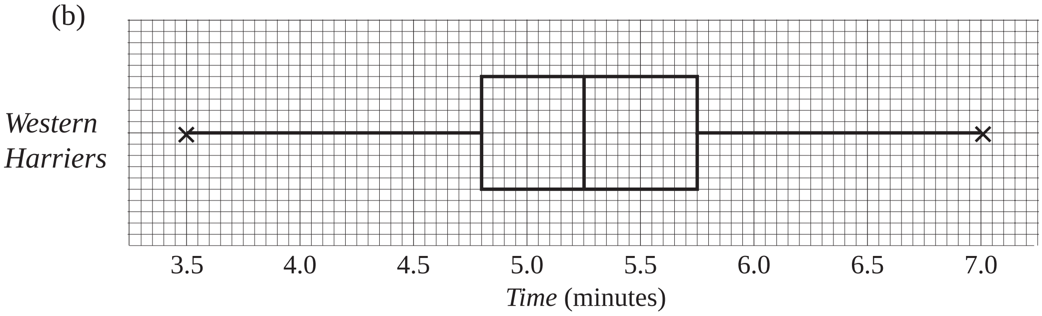
Solution

(a) (i) From the graph, the median is $5\frac{1}{4}$ mins or 5 min 15 sec.



(ii) Similarly: upper quartile is $5\frac{3}{4}$ mins or 5 mins 45 sec, lower quartile is 4 min 48 sec (each small square is 3 seconds in width)

(b)



(c) The data is almost symmetrical about the median.

(d) The data for Eastham Runners is skewed to the left with a lower median time. Their results are generally significantly better than those for Western Harriers, as there are relatively more athletes with faster times.

3.9

Exercises

1. (a) Explain the difference between a discrete and a continuous variable. Give an example of each.

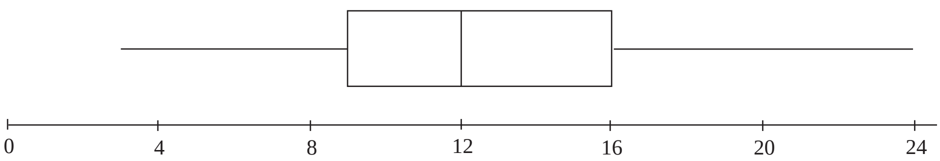
Listed below are the number of times per month that a particular photocopier was unfit for use.

10	15	17	3	9	22	16	11	10
7	9	9	12	16	20	13	24	14
10	9	5	10	21	8	23	15	13

(b) Construct a stem and leaf display for these data.

(c) State two advantages associated with such displays.

The box-plot for the above data is illustrated below.



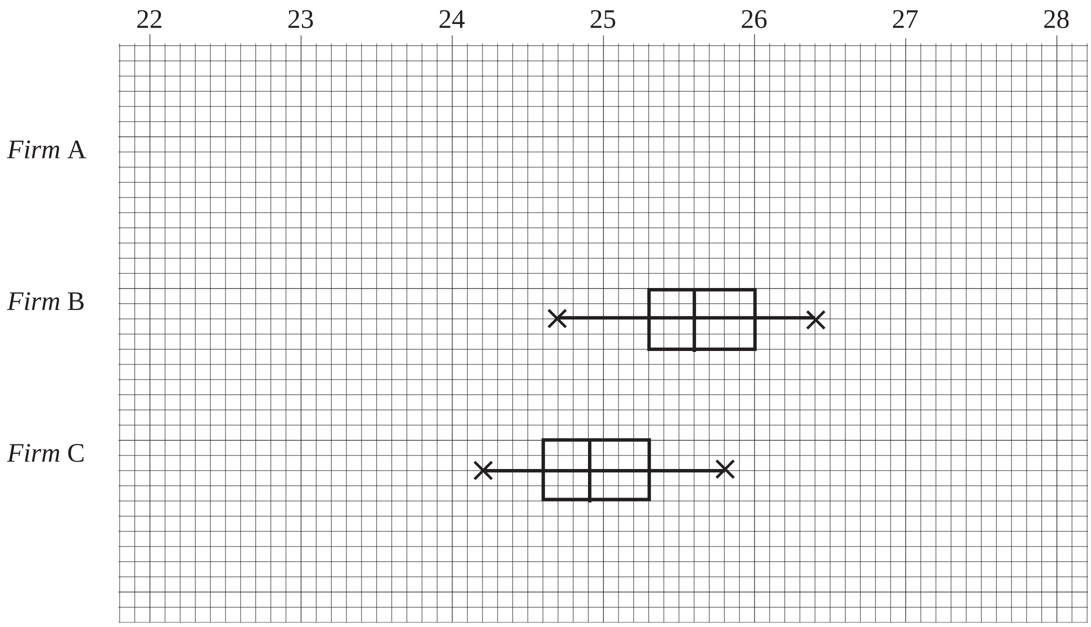
(d) Write down the important numerical values featured in this box-plot.

(LON)

2. A manufacturing company needs to place a regular order for components. The manager investigates components produced by three different firms and measures the diameters of a sample of 25 components from each firm.

The results of the measurements for the samples of components from Firm B and Firm C are illustrated in the two box plots shown below.

Diameter of component (mm)



(a) (i) Find the range of the sample of measurement for Firm B.

(ii) Find the interquartile range of the sample of measurement for Firm C.

(b) The results of the measurements for the sample from Firm A are summarised as follows. Median = 25.0 mm, lower quartile = 23.4 mm, upper quartile = 26.5 mm, lowest value = 22.5 mm, highest value = 27.3 mm.

Draw, on a copy of the grid above, a box plot to illustrate the sample results for Firm A.

(c) The manager studies the three box plots to decide which firm's components he should use. The components he requires should have a diameter of 25 mm, but some variation above and below this measurement is bound to happen and is acceptable. Any components with diameters below 24 mm or above 26 mm will have to be thrown away. State which firm's components you think the manager should choose. Explain carefully why you think he should choose this firm rather than the other two.

(NEAB)

3. A random sample of 32 people were asked to record the number of miles they travelled by car in a given week. The distances, to the nearest mile, are shown below.

67	76	85	42	92	48	93	46
52	72	77	53	41	48	86	78
56	80	70	70	66	62	54	85
60	58	43	58	74	44	52	74

(a) Construct a stem and leaf diagram to represent these data.

(b) Find the median and the quartiles of this distribution.

(c) Draw a box plot to represent these data.

(d) Give one advantage of using (i) a stem and leaf diagram, (ii) a box plot, to illustrate data such as that given above.

4. The following cumulative frequency polygon shows the times taken to travel to a city centre school by a group of children.

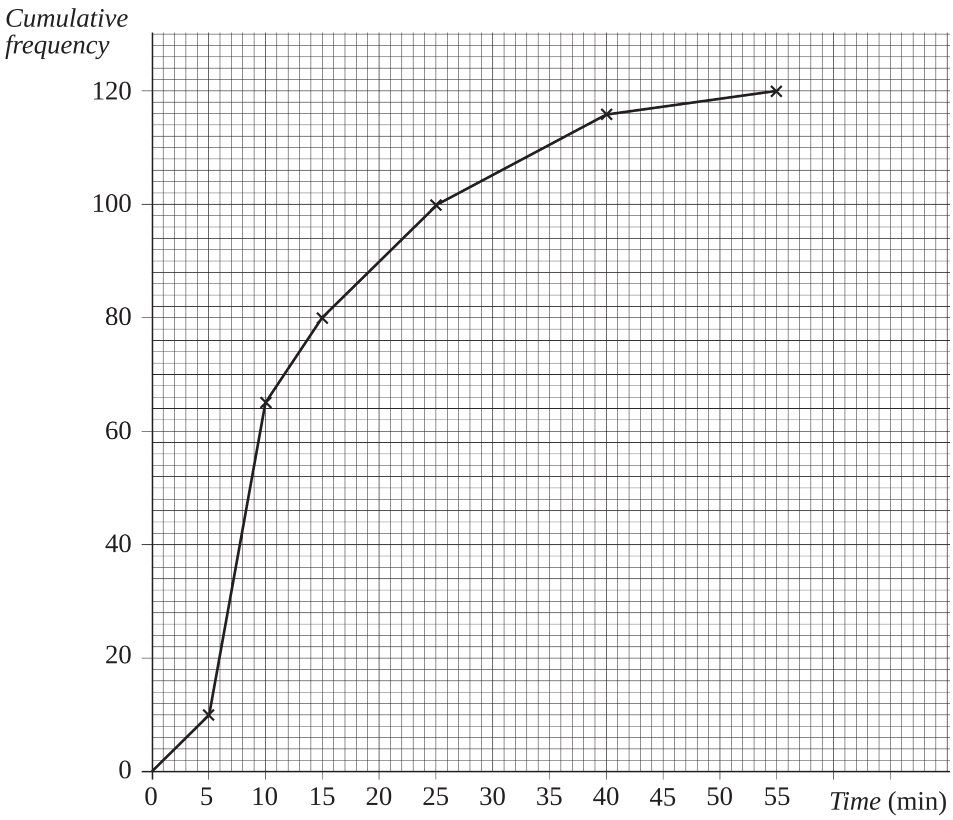
(a) Estimate from the graph:

(i) the median;

(ii) the interquartile range;

(iii) the percentage of children taking more than 35 minutes to reach school.

3.9



(b) A school of equivalent size in a rural area showed the following distribution of times taken to travel to school.

Time taken (min)	*No. of pupils*	*Cumulative frequency table*	
		Time	*No. of pupils*
0 and under 5	8	<5	8
5 and under 10	44	<10	
10 and under 15	15	<15	
15 and under 25	9	<25	
25 and under 40	7	<40	
40 and under 55	37	<55	

(i) Complete a copy of the cumulative frequency table for the data shown.

(ii) Draw the cumulative frequency polygon for this school.

(iii) Estimate from this polygon:

the median,

the interquartile range.

(c) Construct box and whisker plots for each set of data and comment on the main differences that are apparent between the two distributions.

5. In a village a record was kept of the ages of those people who died in 1992. The data are shown on the stem and leaf diagram.

0	6				
1	1				
2	3	5			
3	0	0	1		
4	4	5	6	9	9
5	3	7	8	9	
6	7	8	9		

Key: 2|3 denotes 23 years

(a) How many people died in this village in 1992 ?

At the start of the year there were 750 people in the village.

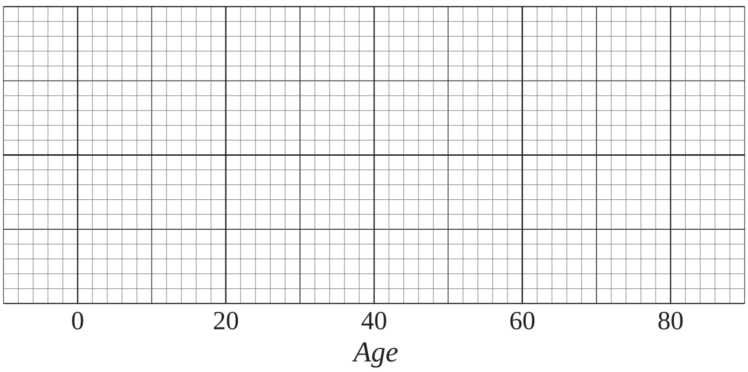
(b) What percentage of the population of the village died that year?

(c) Use the stem and leaf diagram to obtain values for:

(i) the median,

(ii) the lower and upper quartile.

(d) On a copy of the diagram below, draw a box and whisker diagram to illustrate the data.



(e) Do you think this village has a large population of elderly residents? Justify your answer by using the box and whisker plot.