

# 9 Data Analysis

## 9.1 Mean, Median, Mode and Range

In Unit 8, you were looking at ways of collecting and representing data. In this unit, 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.
- 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

$$\begin{aligned} & \frac{5+6+2+4+7+8+3+5+6+6}{10} \\ &= \frac{52}{10} \\ &= 5.2. \end{aligned}$$

- (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,

$$\begin{aligned} \text{median} &= \frac{5+6}{2} \\ &= \frac{11}{2} \\ &= 5.5. \end{aligned}$$

- (c) From the list above 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

$$\begin{aligned} & \frac{3+4+4+5+7}{5} \\ &= \frac{23}{5} \\ &= 4.6. \end{aligned}$$

- (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

$$\begin{aligned} \text{range} &= 7 - 3 \\ &= 4. \end{aligned}$$



## 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.

<i>Worker</i>	<i>Mon</i>	<i>Tue</i>	<i>Wed</i>	<i>Thu</i>	<i>Fri</i>
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.

<i>Company A</i>	20	5	20	20	20	6	20	20	20	8
<i>Company B</i>	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.

<i>City</i>	<i>Maximum</i>	<i>Minimum</i>
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?
- (SEG)

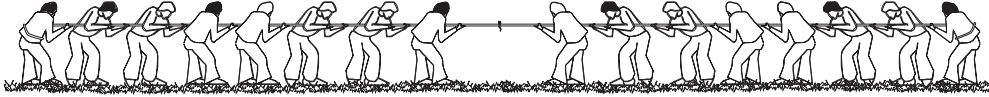
13. Here are the number of goals scored by a school football team in their matches this term.

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.

<i>Kim</i>	28	24	21	27	24	26
<i>Pat</i>	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)



## Information

*The study of statistics was begun by an English mathematician, John Graunt (1620–1674). He collected and studied the death records in various cities in Britain and, despite the fact that people die randomly, he was fascinated by the patterns he found.*

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)

## 9.2 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.

<i>No. of Goals</i>	<i>Frequency</i>
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.

The mean can now be calculated.

$$\begin{aligned} \text{Mean} &= \frac{73}{40} \\ &= 1.825. \end{aligned}$$

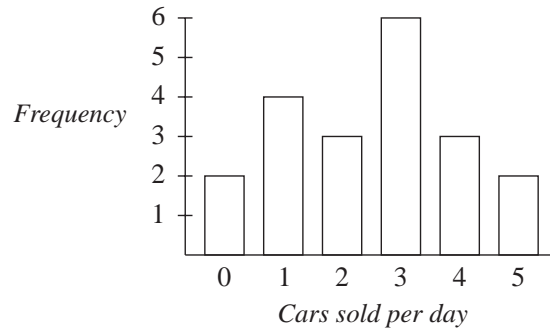
<i>No. of Goals</i>	<i>Frequency</i>	<i>No. of Goals × Frequency</i>
0	8	$0 \times 8 = 0$
1	10	$1 \times 10 = 10$
2	12	$2 \times 12 = 24$
3	3	$3 \times 3 = 9$
4	5	$4 \times 5 = 20$
5	2	$5 \times 2 = 10$
<b>TOTALS</b>	<b>40</b>	<b>73</b>

(Total matches) (Total goals)



## 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.



## Solution

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

<i>Cars sold daily</i>	<i>Frequency</i>	<i>Cars sold × Frequency</i>
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	$5 \times 2 = 10$
<b>TOTALS</b>	<b>20</b>	<b>50</b>

(Total days)                      (Total number of cars sold)

$$\begin{aligned} \text{Mean} &= \frac{50}{20} \\ &= 2.5 \end{aligned}$$



## 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.

<i>Number of Accidents</i>	<i>Tally</i>	<i>Frequency</i>	<i>No. of Accidents × Frequency</i>
0		4	$0 \times 4 = 0$
1		10	$1 \times 10 = 10$
2		22	$2 \times 22 = 44$
3		23	$3 \times 23 = 69$
4		16	$4 \times 16 = 64$
5		17	$5 \times 17 = 85$
6		6	$6 \times 6 = 36$
7		1	$7 \times 1 = 7$
8		1	$8 \times 1 = 8$
	<b>TOTALS</b>	<b>100</b>	<b>323</b>

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



## Exercises

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

<i>No. of Cars</i>	<i>Frequency</i>
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.

<i>No. of TV Sets</i>	<i>Frequency</i>
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.

<i>Number of calls per day</i>	0	1	2	3	4	5	6	7	8
<i>Frequency</i>	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.

<i>No. of Runs</i>	<i>Frequency</i>
0	3
1	2
2	1
3	6
4	5
5	4
6	2
7	1
8	1

Calculate the mean number of runs per over.

5. A class conduct an experiment in biology. They place a number of 1m by 1m square grids on the playing field and count the number of worms which appear when they pour water on the ground. The results obtained are given below.

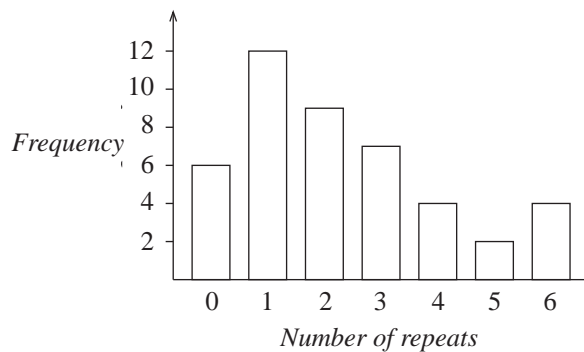
6	3	2	1	3	2	1	3	0	1
0	3	2	1	1	4	0	1	2	0
1	1	2	2	2	4	3	1	1	1
2	3	3	1	2	2	2	1	7	1

- (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 drew this bar chart to show the number of repeated cards she got when she opened packets of football stickers.



Calculate the mean number of repeats per packet.

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

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

- (a) In how many matches did they score 2 goals?  
 (b) Calculate the mean number of goals per match.
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	●	●
<b>TOTALS</b>	<b>26</b>	<b>72</b>

Fill in the missing numbers and calculate the mean.

10. Here are the weights, in kg, of 30 students.

45, 52, 56, 65, 34, 45, 67, 65, 34, 45, 65, 87, 45, 34, 56,  
54, 45, 67, 84, 45, 67, 45, 56, 76, 57, 84, 35, 64, 58, 60

- (a) Copy and complete the frequency table below using a class interval of 10 and starting at 30.

<i>Weight Range (w)</i>	<i>Tally</i>	<i>Frequency</i>
$30 \leq w < 40$		

- (b) Which class interval has the highest frequency?

(LON)

11. 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)

## 9.3 Calculations with the Mean

This section considers calculations concerned with the mean, which is usually taken to be the most important measure of the average of a set of data.



### 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

$$\begin{aligned}\text{Total of original numbers} &= 6 \times 3.2 \\ &= 19.2\end{aligned}$$

$$\begin{aligned}\text{New total} &= 19.2 + 3.9 \\ &= 23.1\end{aligned}$$

$$\begin{aligned}\text{New mean} &= \frac{23.1}{7} \\ &= 3.3\end{aligned}$$



## 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

$$\begin{aligned}\text{Total of the original numbers} &= 5 \times 12.7 \\ &= 63.5\end{aligned}$$

$$\begin{aligned}\text{Total of the new numbers} &= 6 \times 13.1 \\ &= 78.6\end{aligned}$$

$$\begin{aligned}\text{Difference} &= 78.6 - 63.5 \\ &= 15.1\end{aligned}$$

So the extra number is 15.1.



## Exercises

- 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?
- 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?
- 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?
- 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.
- 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?
- 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?
- 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?

## 9.4 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*.



### Worked Example 1

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

<i>Class Interval</i>	$140 \leq h < 150$	$150 \leq h < 160$	$160 \leq h < 170$	$170 \leq h < 180$
<i>Frequency</i>	6	16	21	8

- (a) Estimate the mean height.
- (b) Estimate the median height.
- (c) Find the modal class.



### Solution

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

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

$$\begin{aligned} \text{Mean} &= \frac{8215}{51} \\ &= 161 \text{ (to the nearest cm)} \end{aligned}$$

- (b) The median is the 26th value. In this case it lies in the  $160 \leq h < 170$  class interval. The 4th value in the interval is needed. It is estimated as

$$160 + \frac{4}{21} \times 10 = 162 \text{ (to the nearest cm)}$$

- (c) The modal class is  $160 \leq h < 170$  as it contains the most values.

Also note that 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 in 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.

<i>Age</i>	5 – 6	7 – 8	9 – 10
<i>Frequency</i>	29	40	38

- (a) Estimate the mean.      (b) Estimate the median.      (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

$$5 \leq \text{age} < 7,$$

the mid-point is taken as 6.

<i>Class Interval</i>	<i>Mid-point</i>	<i>Frequency</i>	<i>Mid-point</i> $\times$ <i>Frequency</i>
5 – 6	6	29	$6 \times 29 = 174$
7 – 8	8	40	$8 \times 40 = 320$
9 – 10	10	38	$10 \times 38 = 380$
<b>Totals</b>		<b>107</b>	<b>874</b>

$$\begin{aligned} \text{Mean} &= \frac{874}{107} \\ &= 8.2 \text{ (to 1 decimal place)} \end{aligned}$$

- (b) The median is given by the 54th value, which we have to estimate. There are 29 values in the first interval, so we need to estimate the 25th value in the second interval. As there are 40 values in the second interval, the median is estimated as being

$$\frac{25}{40}$$

of the way along the second interval. This has width  $9 - 7 = 2$  years, so the median is estimated by

$$\frac{25}{40} \times 2 = 1.25$$

from the start of the interval. Therefore the median is estimated as

$$7 + 1.25 = 8.25 \text{ years.}$$

- (c) The modal age is the 7 – 8 age group.

Worked Example 1 uses what are called *continuous data*, since height can be of any value. (Other examples of continuous data are weight, temperature, area, volume and time.)

The next example uses *discrete data*, that is, data which can take only a particular value, such as the integers 1, 2, 3, 4, . . . in this case.

The calculations for mean and mode are not affected but estimation of the median requires replacing the *discrete* grouped data with an approximate *continuous* interval.



### Worked Example 3

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

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

- (a) Estimate the mean      (b) Estimate the median.      (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.

<i>Class Interval</i>	<i>Mid-point</i>	<i>Frequency</i>	<i>Mid-point × Frequency</i>
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	$23 \times 3 = 69$
<b>Totals</b>		<b>40</b>	<b>395</b>

$$\begin{aligned} \text{Mean} &= \frac{395}{40} \\ &= 9.925 \text{ days.} \end{aligned}$$

- (b) As there are 40 pupils, we need to consider the mean of the 20th and 21st values. These both lie in the 6–10 class interval, which is really the 5.5–10.5 class interval, so this interval contains the median.

As there are 12 values in the first class interval, the median is found by considering the 8th and 9th values of the second interval.

As there are 11 values in the second interval, the median is estimated as being

$$\frac{8.5}{11}$$

of the way along the second interval.

But the length of the second interval is  $10.5 - 5.5 = 5$ , so the median is estimated by

$$\frac{8.5}{11} \times 5 = 3.86$$

from the start of this interval. Therefore the median is estimated as

$$5.5 + 3.86 = 9.36.$$

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



## Exercises

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

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

- (a) Estimate the mean number of homes visited.  
 (b) Estimate the median.  
 (c) What is the modal class?
2. The weights of a number of students were recorded in kg.

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

- (a) Estimate the mean weight.                      (b) Estimate the median.  
 (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.

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

- (a) Is the median in the modal class?            (b) Estimate the mean.  
 (c) Estimate the median.  
 (d) Is the median greater or less than the mean?
4. The distances that children in a year group travelled to school is recorded.

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

- (a) Does the modal class contain the median?  
 (b) Estimate the median and the mean.  
 (c) Which is the largest, the median or the mean?



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

<i>Age (years)</i>	6 – 8	9 – 11	12 – 14	15 – 17
<i>Frequency</i>	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.

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

Estimate (a) the mean (b) the median length of a leaf.

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

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

- (a) Estimate the mean. (b) Estimate the median.  
 (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.

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

- (i) Estimate the mean. (ii) Estimate the median.  
 (iii) What is the modal class?
- (b) Another class took the same test. Their results are given below.

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

- (i) Estimate the mean. (ii) Estimate the median.  
 (iii) What is the modal class?
- (c) How do the results for the two classes compare?



## Information

A **quartile** is one of 3 values (lower quartile, median and upper quartile) which divides data into 4 equal groups.

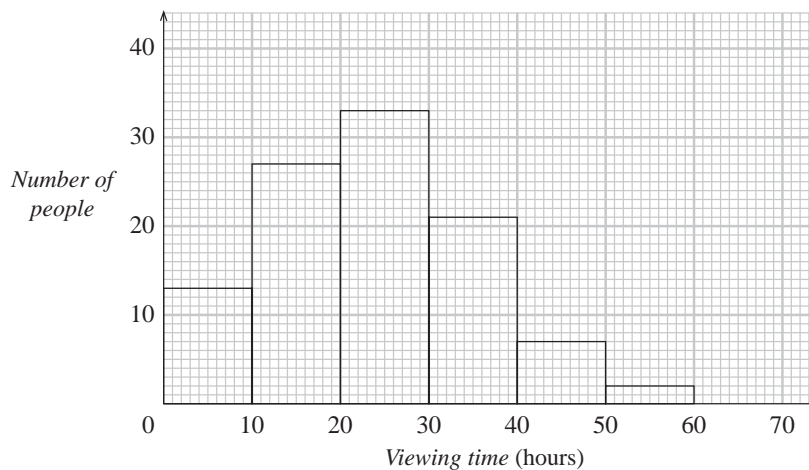
A **percentile** is one of 99 values which divides data into 100 equal groups.

The **lower quartile** corresponds to the 25th percentile. The **median** corresponds to the 50th percentile. The **upper quartile** corresponds to the 75th percentile.

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

<i>Pocket money</i> £	<i>Frequency</i> <i>f</i>
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.

<i>Viewing time</i> ( <i>h</i> hours)	<i>Number of</i> <i>people</i>
$0 \leq h < 10$	13
$10 \leq h < 20$	27
$20 \leq h < 30$	33
$30 \leq h < 40$	
$40 \leq h < 50$	
$50 \leq h < 60$	

- (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.

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

<i>Length (cm)</i>	20	21	22	23	24	25	26	27	28	29
<i>Frequency</i>	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.

<i>Temperature, T</i>		<i>Frequency</i>
$40 < T \leq 50$		
$50 < T \leq 54$		
$54 < T \leq 58$		
$58 < T \leq 62$		
$62 < T \leq 70$		

- (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)

## 9.5 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 describes how spread out the data are. This is illustrated in Example 2.



### Worked Example 1

For the data given in the table, draw up a cumulative frequency table and then draw a cumulative frequency graph.



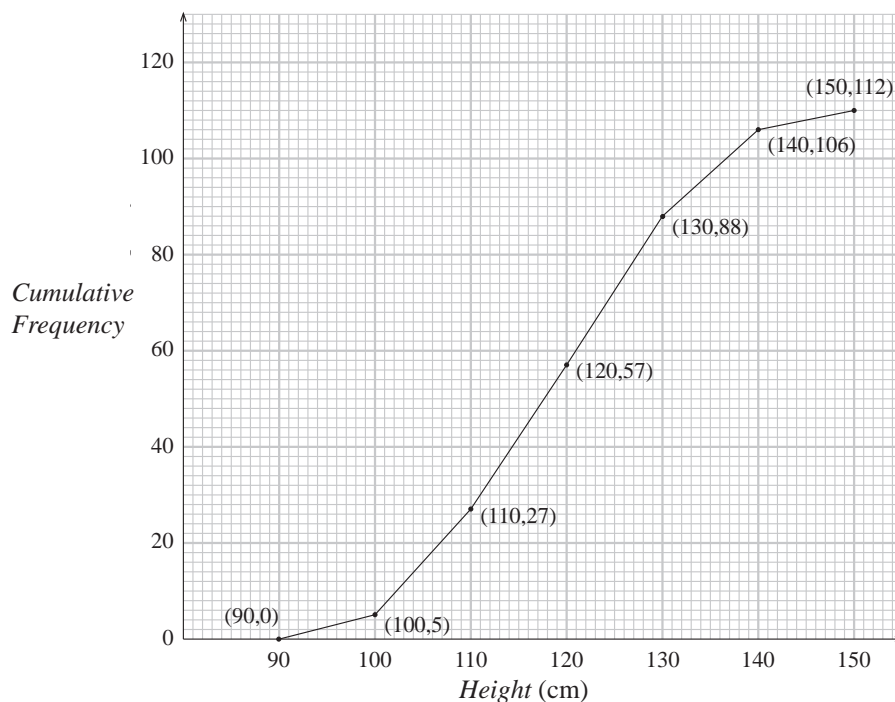
### Solution

The table below shows how to calculate the cumulative frequencies.

<i>Height (cm)</i>	<i>Frequency</i>
$90 < h \leq 100$	5
$100 < h \leq 110$	22
$110 < h \leq 120$	30
$120 < h \leq 130$	31
$130 < h \leq 140$	18
$140 < h \leq 150$	6

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

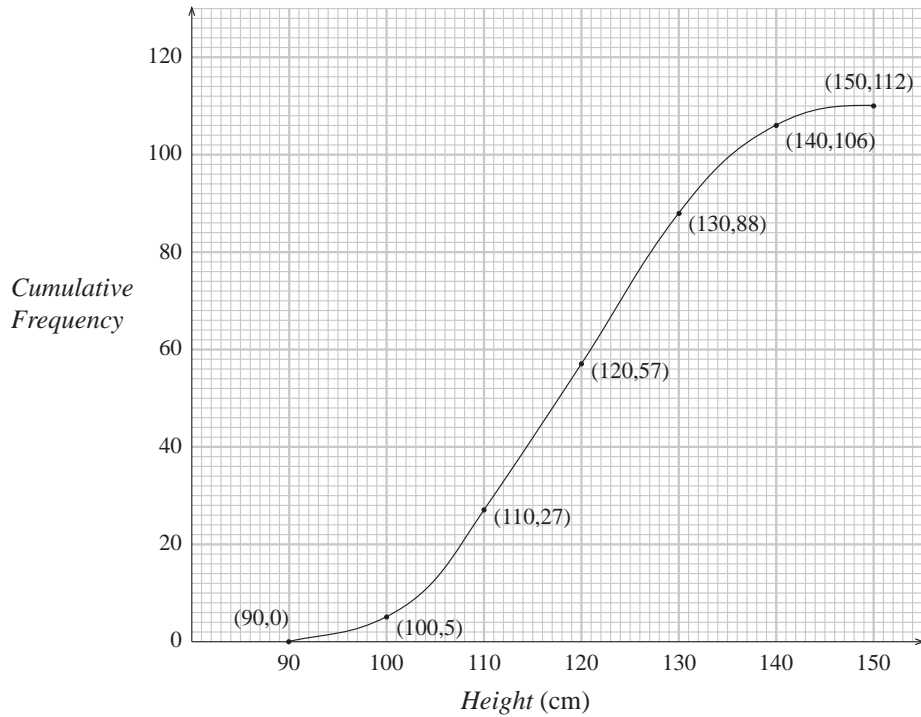
A graph can then be plotted using points as shown below.





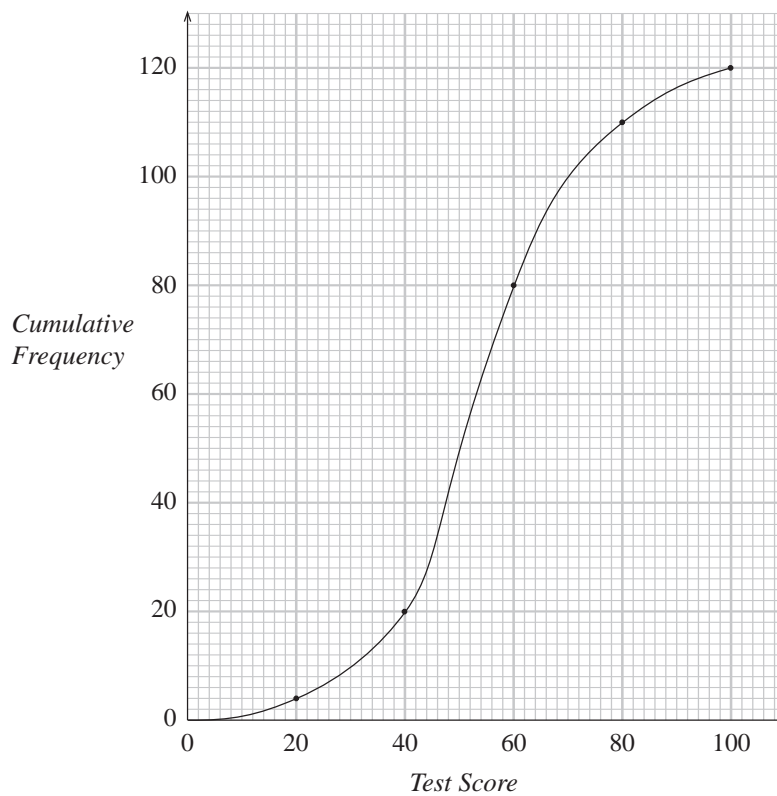
## Note

A more accurate graph is found by drawing a smooth curve through the points, rather than using straight line segments.



## Worked Example 2

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



Use the graph to find:

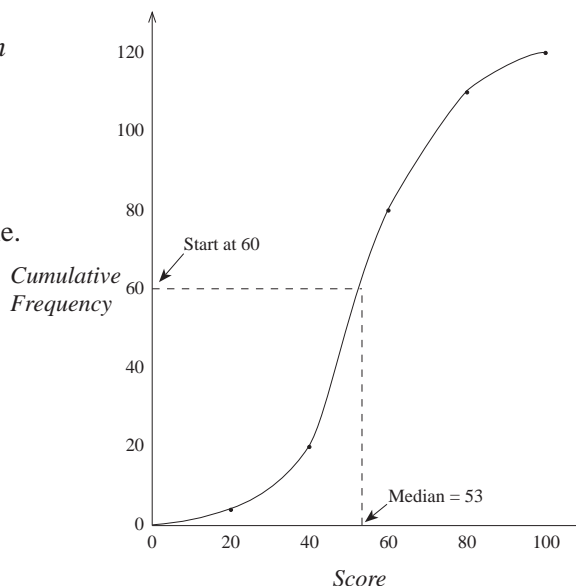
- (a) the median score, (b) the inter-quartile range,  
 (c) the mark which was attained by only 10% of the students,  
 (d) the number of students who scored more than 75 on the test.



## 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.

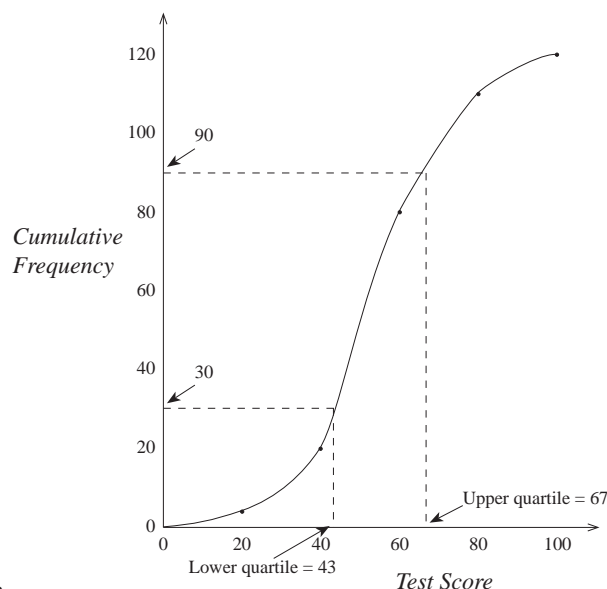
This gives

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

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

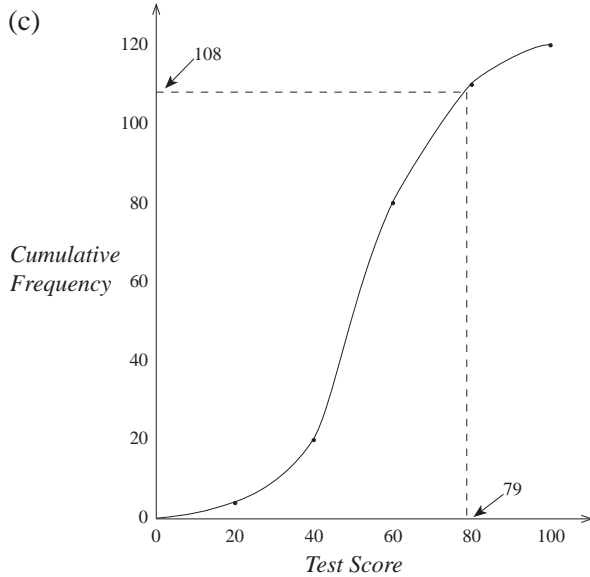
This gives

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



The *inter-quartile range* is then

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



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.

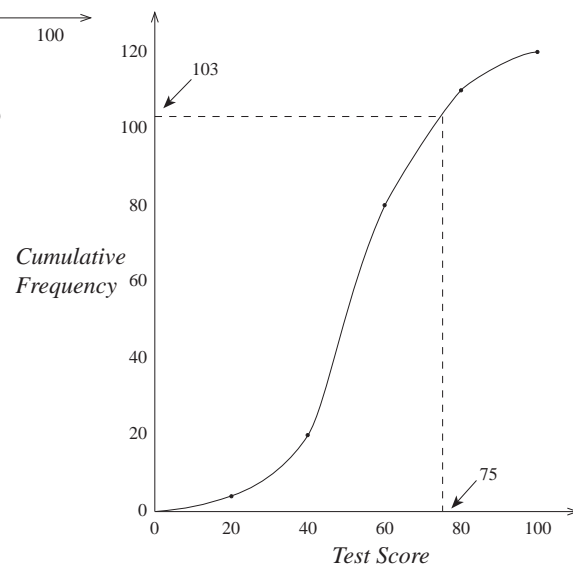
This gives a mark of 79.

- (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

$$120 - 103 = 17.$$



As in Worked Example 1, a more accurate estimate for the median and inter-quartile range is obtained if you draw a smooth curve through the data points.



## 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)	$60 < w \leq 80$	$80 < w \leq 100$	$100 < w \leq 120$	$120 < w \leq 140$	$140 < w \leq 160$
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)	$0 < d \leq 1$	$1 < d \leq 2$	$2 < d \leq 3$	$3 < d \leq 4$	$4 < d \leq 5$	$5 < d \leq 6$
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.

<i>Length of jump (m)</i>	$1 < d \leq 2$	$2 < d \leq 3$	$3 < d \leq 4$	$4 < d \leq 5$	$5 < d \leq 6$
<i>Frequency</i>	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.

<i>Mass of grain (g)</i>	$0 < m \leq 5$	$5 < m \leq 10$	$10 < m \leq 15$	$15 < m \leq 20$	$20 < m \leq 25$	$25 < m \leq 30$
<i>Frequency Field A</i>	3	8	22	10	4	3
<i>Frequency Field B</i>	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.

<i>Lifetime (hours)</i>	$2 < l \leq 3$	$3 < l \leq 4$	$4 < l \leq 5$	$5 < l \leq 6$	$6 < l \leq 7$	$7 < l \leq 8$
<i>Frequency Type A</i>	1	3	10	22	8	4
<i>Frequency Type B</i>	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.

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

A doctor wishes to be able to classify children as:

<i>Category</i>	<i>Percentage of Population</i>
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.

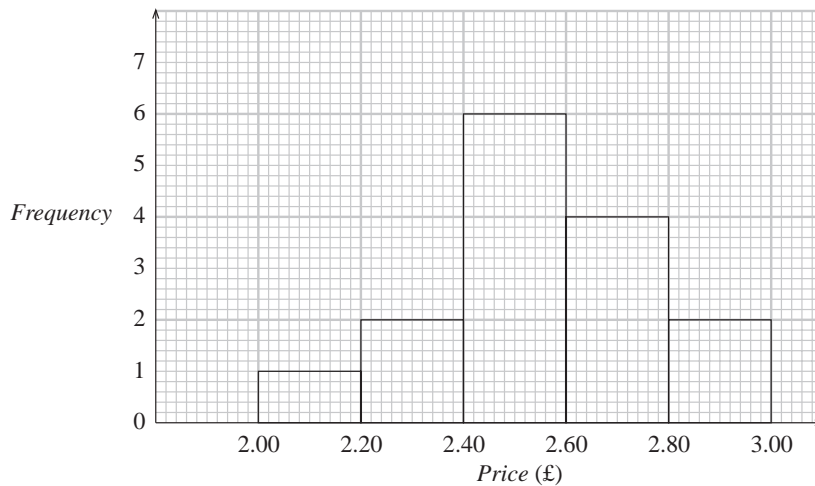
<i>Bonus</i>	<i>Awarded to</i>
£500	Best 10% of salesmen
£250	Middle 70% of salesmen
£ 50	Bottom 20% of salesmen

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

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

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

6. The histogram shows the cost of buying a particular toy in a number of different shops.

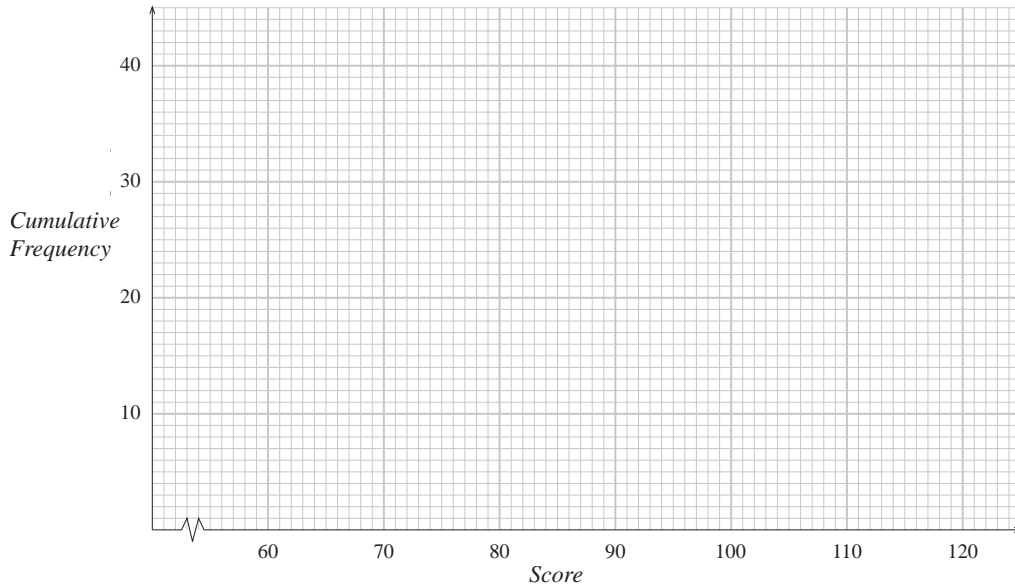


- (a) Draw a cumulative frequency graph and use it to answer the following questions.
- How many shops charged more than £2.65?
  - What is the median price?
  - How many shops charged less than £2.30?
  - How many shops charged between £2.20 and £2.60?
  - How many shops charged between £2.00 and £2.50?
- (b) Comment on which of your answers are exact and which are estimates.

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

Scores ( $x$ )	$70 < x \leq 80$	$80 < x \leq 90$	$90 < x \leq 100$	$100 < x \leq 110$	$110 < x \leq 120$
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.

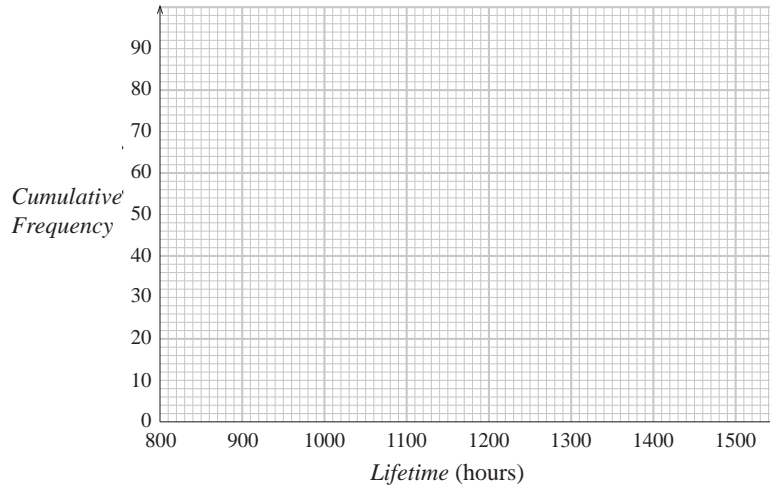


- (b) Making your method clear, use your graph to find
- (i) Laura's median score,
  - (ii) the inter-quartile range of her scores.
- (c) Joy's median score was 103. The inter-quartile range of her scores was 6.
- (i) Who was the more consistent player?  
Give a reason for your choice.
  - (ii) The winner of a game of golf is the one with the lowest score.  
Who won most of these 40 games? Give a reason for your choice.
- (NEAB)

8. 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 below.



- (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)

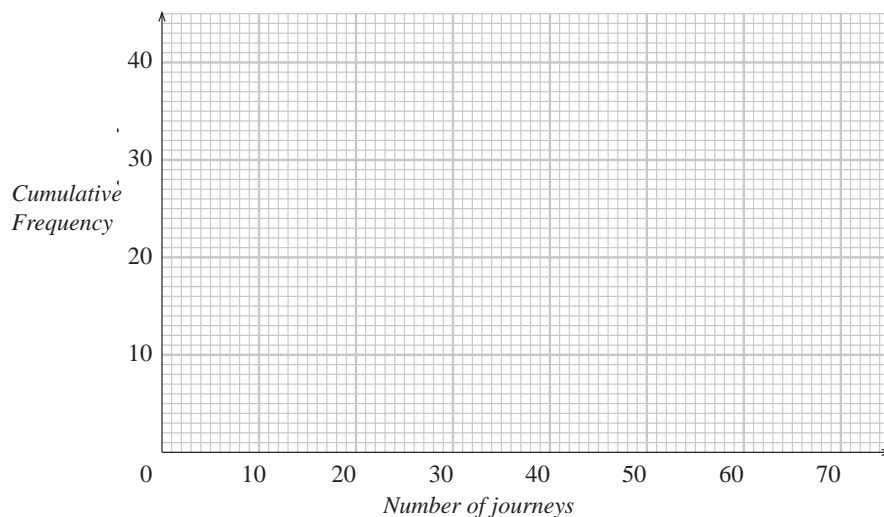
9. The numbers of journeys made by a group of people using public transport in one month are summarised in the table.

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

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

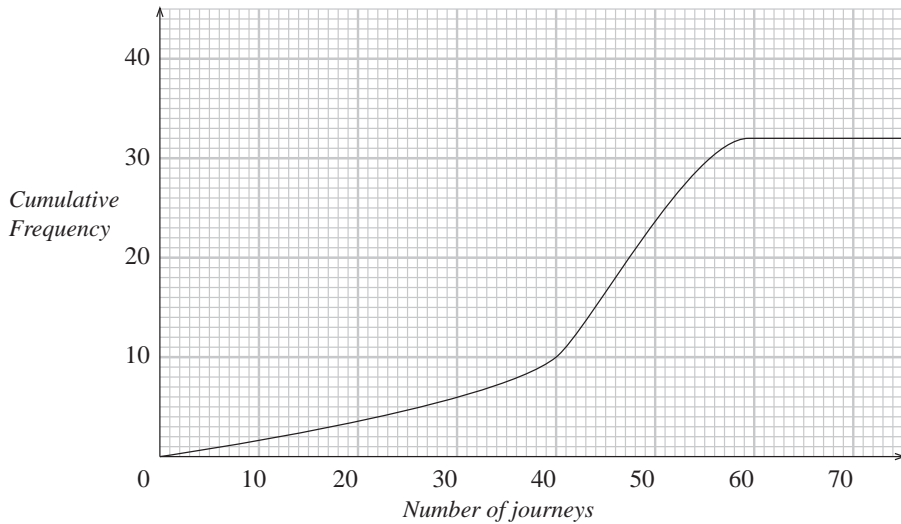
<i>Number of journeys</i>	$\leq 10$	$\leq 20$	$\leq 30$	$\leq 40$	$\leq 50$	$\leq 60$	$\leq 70$
<i>Cumulative frequency</i>							

- (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.

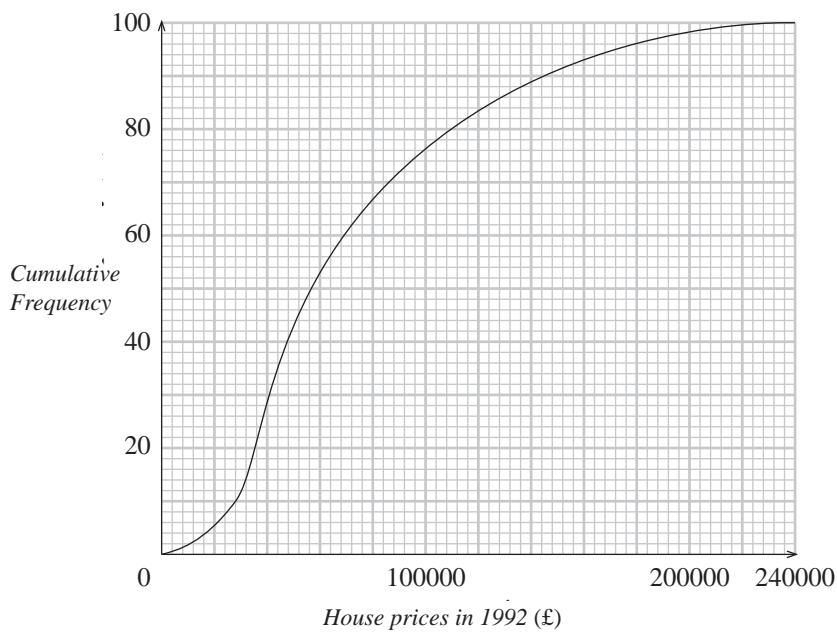
(c) The numbers of journeys made using public transport in one month, by another group of people, are shown in the graph.



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

(SEG)

10. The cumulative frequency graph below gives information on the house prices in 1992. The cumulative frequency is given as a percentage of all houses in England.



This grouped frequency table gives the percentage distribution of house prices ( $p$ ) in England in 1993.

<i>House prices (p) in pounds 1993</i>	<i>Percentage of houses in this class interval</i>
$0 \leq p < 40\,000$	26
$40\,000 \leq p < 52\,000$	19
$52\,000 \leq p < 68\,000$	22
$68\,000 \leq p < 88\,000$	15
$88\,000 \leq p < 120\,000$	9
$120\,000 \leq p < 160\,000$	5
$160\,000 \leq p < 220\,000$	4

- (a) Use the data above to complete the cumulative frequency table below.

<i>House prices (p) in pounds 1993</i>	<i>Cumulative Frequency (%)</i>
$0 \leq p < 40\,000$	
$0 \leq p < 52\,000$	
$0 \leq p < 68\,000$	
$0 \leq p < 88\,000$	
$0 \leq p < 120\,000$	
$0 \leq p < 160\,000$	
$0 \leq p < 220\,000$	

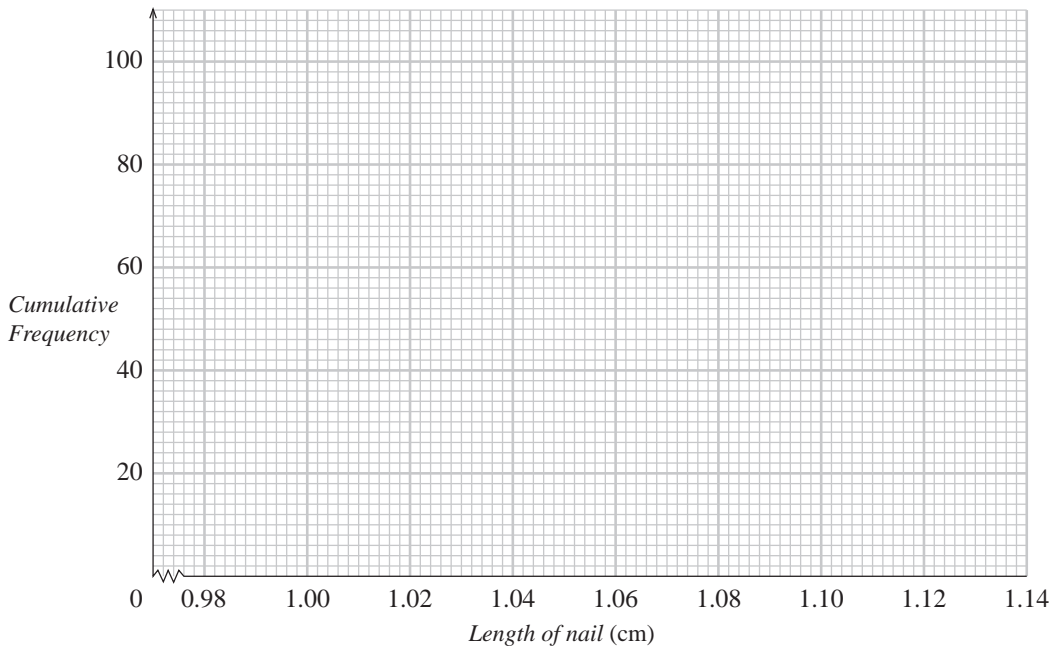
- (b) Trace or photocopy the grid for 1992, and on it construct a cumulative frequency graph for your table for 1993.
- (c) In 1992 the price of a house was £100 000. Use both cumulative frequency graphs to estimate the price of this house in 1993. Make your method clear.

(LON)

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

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

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

<i>Distance</i> ( $d$ miles)	$0 < d \leq 10$	$10 < d \leq 20$	$20 < d \leq 30$	$30 < d \leq 50$	$50 < d \leq 100$	$100 < d \leq 140$
<i>Number of</i> <i>guests</i>	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.

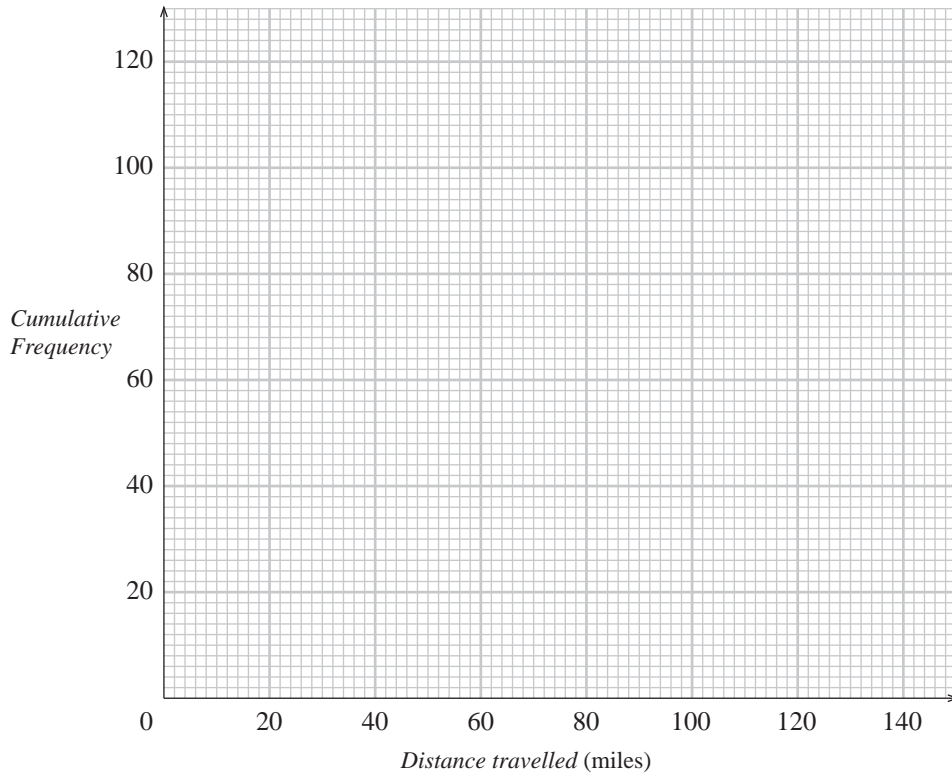
<i>Distance</i> ( $d$ miles)	$d \leq 10$	$d \leq 20$	$d \leq 30$	$d \leq 50$	$d \leq 100$	$d \leq 140$
<i>Number of</i> <i>guests</i>						120

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



### Just for Fun

Comment on this statement. "Students who smoke perform worse academically than non-smokers, so smoking will make you stupid."

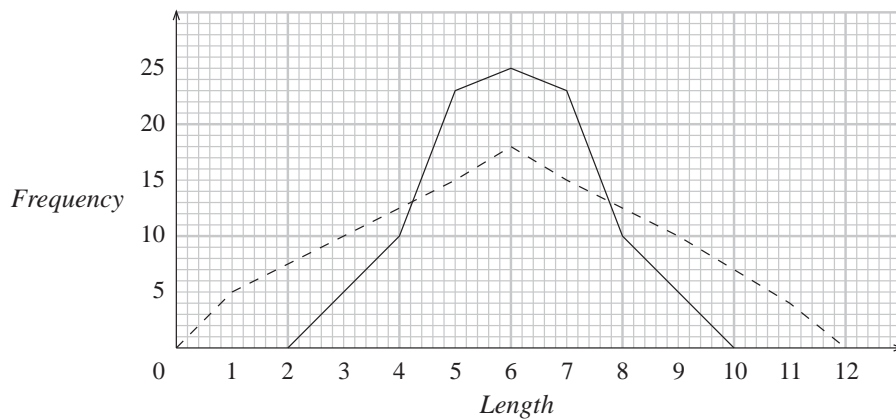


- (c) (i) Use the cumulative frequency curve to estimate the median distance travelled by the guests.
- (ii) Give a reason for the large difference between the mean distance and the median distance.

(MEG)

## 9.6 Standard Deviation

The two *frequency polygons* drawn on the graph below show samples which have the same mean, but the data in one are much more spread out than in the other.



The *range* (highest value – lowest value) gives a simple measure of how much the data are spread out.

*Standard deviation* (s.d.) is a much more useful measure and is given by the formula:

$$\text{s.d.} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

where  $x_i$  represents each datapoint ( $x_1, x_2, \dots, x_n$ )  
 $\bar{x}$  is the mean,  
 $n$  is the number of values.

Then  $(x_i - \bar{x})^2$  gives the square of the difference between each value and the mean (squaring exaggerates the effect of data points far from the mean and gets rid of negative values), and

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

sums up all these squared differences.

The expression

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

gives an average value to these differences. If all the data were the same, then each  $x_i$  would equal  $\bar{x}$  and the expression would be zero.

Finally we take the square root of the expression so that the dimensions of the standard deviation are the same as those of the data.

So standard deviation is a measure of the spread of the data. The greater its value, the more spread out the data are. This is illustrated by the two frequency polygons shown above. Although both sets of data have the same mean, the data represented by the 'dotted' frequency polygon will have a greater standard deviation than the other.



### Worked Example 1

Find the mean and standard deviation of the numbers,

6, 7, 8, 5, 9.



### Solution

The mean,  $\bar{x}$ , is given by,

$$\begin{aligned} \bar{x} &= \frac{6 + 7 + 8 + 5 + 9}{5} \\ &= \frac{35}{5} \\ &= 7. \end{aligned}$$



Now the standard deviation can be calculated.

$$\begin{aligned}
 \text{s.d.} &= \sqrt{\frac{(6-7)^2 + (7-7)^2 + (8-7)^2 + (5-7)^2 + (9-7)^2}{5}} \\
 &= \sqrt{\frac{1 + 0 + 1 + 4 + 4}{5}} \\
 &= \sqrt{\frac{10}{5}} \\
 &= \sqrt{2} \\
 &= 1.414 \quad (\text{to 3 decimal places})
 \end{aligned}$$

An alternative formula for standard deviation is

$$\text{s.d.} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}$$

This expression is much more convenient for calculations done without a calculator. The proof of the equivalence of this formula is given below although it is beyond the scope of the GCSE syllabus.



## Proof

You can see the proof of the equivalence of the two formulae by noting that

$$\begin{aligned}
 \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\
 &= \sum_{i=1}^n x_i^2 - \sum_{i=1}^n (2x_i \bar{x}) + \sum_{i=1}^n \bar{x}^2 \\
 &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x}^2 \sum_{i=1}^n 1
 \end{aligned}$$

(since the expressions  $2\bar{x}$  and  $\bar{x}^2$  are common for each term in the summation).

But  $\sum_{i=1}^n 1 = n$ , since you are summing  $\underbrace{1 + 1 + \dots + 1}_{n \text{ terms}} = n$ , and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , by definition, thus

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x}^2 n \right) \quad (\text{substituting } \sum_{i=1}^n 1 = n)$$

$$\begin{aligned}
 &= \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{x} \left( \frac{\sum_{i=1}^n x_i}{n} \right) + \bar{x}^2 \quad (\text{dividing by } n) \\
 &= \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 \quad (\text{substituting } \bar{x} \text{ for } \frac{\sum_{i=1}^n x_i}{n}) \\
 &= \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2
 \end{aligned}$$

and the result follows.



### Worked Example 2

Find the mean and standard deviation of each of the following sets of numbers.

- (a) 10, 11, 12, 13, 14                      (b) 5, 6, 12, 18, 19



### Solution

- (a) The mean,  $\bar{x}$ , is given by

$$\begin{aligned}
 \bar{x} &= \frac{10 + 11 + 12 + 13 + 14}{5} \\
 &= \frac{60}{5} \\
 &= 12
 \end{aligned}$$

The standard deviation can now be calculated using the alternative formula.

$$\begin{aligned}
 \text{s.d.} &= \sqrt{\left( \frac{10^2 + 11^2 + 12^2 + 13^2 + 14^2}{5} \right) - 12^2} \\
 &= \sqrt{146 - 144} \\
 &= \sqrt{32} \\
 &= 1.414 \quad (\text{to 3 decimal places}).
 \end{aligned}$$

- (b) The mean,  $\bar{x}$ , is given by

$$\begin{aligned}
 \bar{x} &= \frac{5 + 6 + 12 + 18 + 19}{5} \\
 &= 12 \quad (\text{as in part (a)}).
 \end{aligned}$$

The standard deviation is given by

$$\begin{aligned} \text{s.d.} &= \sqrt{\left(\frac{5^2 + 6^2 + 12^2 + 18^2 + 19^2}{5}\right) - 12^2} \\ &= \sqrt{178 - 144} \\ &= 5.831 \quad (\text{to 3 decimal places}). \end{aligned}$$

Note that both sets of numbers have the same mean value, but that set (b) has a much larger standard deviation. This is expected, as the spread in set (b) is clearly far more than in set (a).



### Worked Example 3

The table below gives the number of road traffic accidents per day in a small town.

<i>Accidents per day</i>	0	1	2	3	4	5	6
<i>Frequency</i>	5	8	6	3	2	1	1

Find the mean and standard deviation of this data.



### Solution

The necessary calculations for each datapoint,  $x_i$ , are set out below.

<i>Accidents per day</i> ( $x_i$ )	<i>Frequency</i> ( $f_i$ )	$x_i^2$	$x_i f_i$	$x_i^2 f_i$
0	5	0	0	0
1	8	1	8	8
2	6	4	12	24
3	3	9	9	27
4	2	16	8	32
5	0	25	0	0
6	1	36	6	36
<b>TOTALS</b>	<b>25</b>		<b>43</b>	<b>127</b>

From the totals,

$$n = 25, \quad \sum_{i=1}^n x_i f_i = 43, \quad \sum_{i=1}^n x_i^2 f_i = 127.$$

The mean,  $\bar{x}$ , is now given by

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^n x_i f_i}{n} \\ &= \frac{43}{25} \\ &= 1.72. \end{aligned}$$

The standard deviation is now given by

$$\begin{aligned} \text{s.d.} &= \sqrt{\frac{\sum_{i=1}^n x_i^2 f_i}{n} - \bar{x}^2} \\ &= \sqrt{\frac{127}{25} - 1.72^2} \\ &= 1.457. \end{aligned}$$

Most scientific calculators have statistical functions which will calculate the mean and standard deviation of a set of data.



## Exercises

1. (a) Find the mean and standard deviation of each set of data given below.

A	51	56	51	49	53	62
B	71	76	71	69	73	82
C	102	112	102	98	106	124

- (b) Describe the relationship between each set of numbers and also the relationship between their means and standard deviations.
2. Two machines, A and B, fill empty packets with soap powder. A sample of boxes was taken from each machine and the weight of powder (in kg) was recorded.

A	2.27	2.31	2.18	2.2	2.26	2.24			
B	2.78	2.62	2.61	2.51	2.59	2.67	2.62	2.68	2.70

- (a) Find the mean and standard deviation for each machine.
- (b) Which machine is most consistent?
3. Two groups of students were trying to find the acceleration due to gravity. Each group conducted 5 experiments.

Group A	9.4	9.6	10.2	10.8	10.1
Group B	9.5	9.7	9.6	9.4	9.8

Find the mean and standard deviation for each group, and comment on their results.

4. The number of matches per box was counted for 100 boxes of matches. The results are given in the table below.

<i>Number of Matches</i>	<i>Frequency</i>
44	28
45	31
46	14
47	15
48	8
49	2
50	2

Find the mean and standard deviation of this data.

5. When two dice were thrown 50 times the total scores shown below were obtained.

<i>Score</i>	<i>Frequency</i>
2	1
3	0
4	4
5	8
6	12
7	9
8	7
9	5
10	3
11	1
12	0

Find the mean and standard deviation of these scores.

6. The length of telephone calls from an office was recorded. The results are given in the table below.

<i>Length of call (mins)</i>	$0 < t \leq 0.5$	$0.5 < t \leq 1.0$	$1.0 < t \leq 2.0$	$2.0 < t \leq 5.0$
<i>Frequency</i>	8	10	12	4

Estimate the mean and standard deviation using this table.

7. The charges (to the nearest £) made by a garage for repair work on cars in one week are given in the table below.

<i>Charge (£)</i>	20 – 29	30 – 49	50 – 99	100 – 149	150 – 199	200 – 300
<i>Frequency</i>	10	22	6	2	4	1

Use this table to estimate the mean and standard deviation.

8. Thirty families were selected at random in two different countries. They were asked how many children there were in each family.

Country A			Country B		
1	2	1	2	2	1
2	0	1	4	1	1
2	2	3	2	3	5
2	1	1	3	5	1
2	2	2	9	4	2
3	1	0	6	5	1
5	1	4	4	2	2
0	2	1	4	4	5
1	1	0	3	2	2
2	0	2	4	7	0

Find the mean and standard deviation for each country and comment on the results.

9. (a) Calculate the standard deviation of the numbers  
3, 4, 5, 6, 7.
- (b) Show that the standard deviation of *every* set of five consecutive integers is the same as the answer to part (a).  
(LON)
10. Ten students sat a test in Mathematics, marked out of 50. The results are shown below for each student.  
25, 27, 35, 4, 49, 10, 12, 45, 45, 48
- (a) Calculate the mean and standard deviation of the data.
- The same students also sat an English test, marked out of 50. The mean and standard deviation are given by  
mean = 30, standard deviation = 3.6.
- (b) Comment on and contrast the results in Mathematics and English.  
(SEG)
11. Ten boys sat a test which was marked out of 50. Their marks were  
28, 42, 35, 17, 49, 12, 48, 38, 24 and 27.
- (a) Calculate
- the mean of the marks,
  - the standard deviation of the marks.
- Ten girls sat the same test. Their marks had a mean of 30 and a standard deviation of 6.5.
- (b) Compare the performances of the boys and girls.  
(NEAB)

12. There are twenty pupils in class *A* and twenty pupils in class *B*. All the pupils in class *A* were given an I.Q. test. Their scores on the test are given below.

100, 104, 106, 107, 109, 110, 113, 114, 116, 117,  
118, 119, 119, 121, 124, 125, 127, 127, 130, 134.

- (a) The mean of their scores is 117. Calculate the standard deviation.  
 (b) Class *B* takes the same I.Q. test. They obtain a mean of 110 and a standard deviation of 21. Compare the data for class *A* and class *B*.  
 (c) Class *C* has only 5 pupils. When they take the I.Q. test they all score 105. What is the value of the standard deviation for class *C*?

(SEG)

13. The following are the scores in a test for a set of 15 students.

5	4	8	7	3
6	5	9	6	10
7	8	6	4	2

- (a) (i) Calculate the mean score.  
 (ii) Calculate the standard deviation of the scores.

A set of 10 different students took the same test. Their scores are listed below.

5	6	6	7	7
4	7	8	3	7

- (b) After making any necessary calculations for the second set, compare the two sets of scores. Your answer should be understandable to someone who does not study Statistics.

(MEG)

14. In a survey on examination qualifications, 50 people were asked,  
*How many subjects are listed on your GCSE certificate?*

The frequency distribution of their responses is recorded in the table below.

<i>Number of subjects</i>	1	2	3	4	5	6	7
<i>Number of people</i>	5	3	7	8	9	10	8

- (a) Calculate the mean and standard deviation of the distribution.  
 (b) *A Normal Distribution has approximately 68% of its data values within one standard deviation of the mean.*

Use your answers to part (a) to check if the given distribution satisfies this property of a Normal Distribution. Show your working clearly.

(MEG)