6 Dealing with variability

Key words: variability, random error, true value, uncertainty, population, sample, distribution, histogram, batch, class interval, frequency, average, mean, arithmetic mean, median, mode, spread, range, quartile, interquartile range, box plot, outlier, anomaly, probability, independent events, combined events, risk.

Variability in a set of data relates to how spread out or how close together the values are (the distribution of values). Although variability can arise for various reasons, the questions of interest, and the mathematical techniques used to answer them, are similar. These questions are:

- How big is a typical value?
- How much do the values vary?
- Are there any unusual values?

6.1 Where does variability come from?

A familiar example of ***variability*** is the way that people are different to each other. For example, some people run faster than others. In a 100 metres race, we would not expect all the runners to get exactly the same time – there would be a range of values. For a random sample of people, there would be lot of variability, i.e. a wide spread of values. For a race with elite runners, the times would be faster but also much closer to each other, and thus less variability.

A very different example is rolling a marble down a slope. If all the conditions were kept exactly the same, using identical marbles down identical slopes, then in principle we might expect to get exactly the same time for every run. However, in practice, it is likely that there will be some variability in the values. This variability does not result from differences between what is being measured, but in the act of measurement itself. Repeated measurements of the same thing may be different from each other because of measurement uncertainty.

The distinction between these is very important to understand. For all measurements there is *uncertainty in the measured values*, so this may always be a source of variability in repeated measurements. Variability can also arise from *differences between the individuals in a population*. Examples of these two sources of variability are:

- *Measurement uncertainty (for repeated measurements of the same thing)*
  - time of travel of a marble rolling down a slope
  - height of a bouncing ball dropped from a particular height
  - time for a paper parachute to fall
  - volume of a solution used in a titration.

- *Differences between individuals (for measurements of a sample of different but related things)*
  - time to run 100 metres for different people
  - height of pupils in a year group
  - number of spines on holly leaves
  - air pollution levels in different locations.

Understanding the differences between these two sources of variability is important, for example in appreciating the difference between line graphs and scatter graphs (see Section 3.6 *Line graphs and scatter graphs: two related quantities* on page 29). For '*line graph*' type data, the data points may not all lie on a fitted line because of *measurement uncertainty*; for '*scatter graph*' type data, a fitted line may not pass at all close to many of the data points because of differences between individuals.

6.2 Variability and measurement uncertainty

The reason for making repeated measurements in an experiment is because of ***random error***. The experimental design should aim to minimise these errors but they cannot be eliminated. Thus any single measurement may be different from the ***true value***, and repeated measurements may be different from each other. This ***variability*** is called the ***uncertainty***. It depends both on the nature of the measuring instrument and on what is being measured.

For example, when using a ruler divided into millimetres, it is not difficult to measure the width of a sheet of paper to the nearest millimetre. Using a ruler divided into centimetres makes this harder. An estimate can be made by eye to the nearest millimetre, but with greater uncertainty than using the first ruler. The second ruler has a *lower resolution*, and the uncertainty is due to the nature of the measuring instrument.

By contrast, the uncertainty in measuring by eye the height to which a ball bounces is related more to the nature of what is being measured than to the measuring instrument. One can use a metre rule divided into millimetres but the movement of the ball means that it is really only possible to measure its position to the nearest centimetre.

A metre rule is an example of an *analogue* instrument, for which the resolution is related to the size of the scale divisions. On a *digital* instrument, the resolution is determined by the digits shown on the display.

For example, many digital thermometers read to the nearest 0.1 °C – this is its *resolution*. If the temperature of some warm water is measured as 32.6 °C then this suggests that the true value lies closer to this value than to 32.5 °C or 32.7 °C. However, if a second digital thermometer is used to measure the same temperature, it may display 32.9 °C. It is quite usual for different measuring instruments to give different measurements for the same thing. How close a reading is to the true value is related to the *accuracy* of the instrument. Thus the uncertainty in a measurement may be related to the instrument's accuracy and not its resolution.

The uncertainty in using a digital instrument, as for an analogue instrument, may also be related to the nature of what is being measured. If the temperature of the system being investigated changes very rapidly then it may only be possible to make measurements by eye to the nearest 1 °C even though the thermometer reads to the nearest 0.1 °C.

In 11–16 science, the uncertainty of a measurement relates to the number of significant figures in the value. For example, giving the mass of an object as 12 g suggests a greater

uncertainty in its value than if it is given as 12.39 g. The number of significant figures is a reflection of the precision of the measurement. (See Section 1.2 *Measurement, resolution and significant figures* on page 9.)

In scientific practice, and in post-16 science, measurement uncertainty is indicated *explicitly* using the '±' symbol to show the range within which the true value is likely to lie. For example, the manufacturers of a digital thermometer may state that its accuracy is ±0.3 °C, so a measurement with this thermometer might be written as 32.6 °C ± 0.3 °C. This means that the true value is likely to be found in the range 32.3 °C to 32.9 °C. Note that it does not mean that the true value is *definitely* in this range but that there is a good *probability* that it is (often the probability value used is 95%).

It is important to note that the accuracy of the instrument (±0.3 °C) is not the same as its resolution (0.1 °C). They are, however, related through the choices made in the design of the instrument. It would be possible to make a digital thermometer with an accuracy of ±0.3 °C but with a digital display that could show values to the nearest 0.001 °C. There would be no point to this and it would be confusing. A display that reads to the nearest 0.1 °C is adequate for the accuracy of this particular thermometer.

In summary, measurement uncertainty depends on the resolution of the instrument, the accuracy of the instrument and on the nature of what is being measured. It is a complex and subtle area, where there are no simple rules about rounding and significant figures for measured and calculated values. Attempting to invent artificial rules for 11–16 science is neither desirable nor possible. To develop pupils' understanding, it is better if they think about the nature of each situation and make sensible judgements.

Finding the ***mean*** of a set of repeated measurements can reduce the uncertainty and give a value that is more likely to be closer to the true value than any single measurement.

In secondary school science experiments, a common rule of thumb is to take *three* repeated measurements (unless there is poor agreement between the results, suggesting that further measurement is needed). In scientific work, there is nothing 'special' about three repeated measurements. The choice of how many repeated measurements to take depends on their variability. School activities are usually designed so that random errors are relatively small. Taking more than three repeated measurements is time consuming, while with only two measurements there is a chance that the values may be in agreement but both incorrect. Three measurements is a reasonable compromise for most contexts in school science.

For example, the travel time of a marble rolling down a slope could be measured using a stopwatch. These times would be subject to random errors; for example, the watch might not be started and stopped at exactly the same point on each run.

If the measured times are 4.37 s, 4.72 s and 4.48 s then an answer to the question 'How big is a typical value?' can be found by calculating the mean. Dividing the sum of the values by the number of values using a typical calculator produces a value that may be displayed as 4.523333333. Rounding this value to three significant figures (the same as for the measured times) gives 4.52 s. The stopwatch reads to the nearest 0.01 s but the spread in the values suggests that the random errors in timing by hand are somewhat larger than this. In this case, it may make more sense to round to only two significant figures, i.e. 4.5 s. Thus, for these data, the value of 4.5 s is our best guess of the *true value* of the travel time. (For more about means and significant figures, see Section 2.4 *Calculating means* on page 18.)

For further information about the measurement uncertainty, see the ASE/Nuffield publication *The Language of Measurement*.

6.3 Variability in a population of individuals

Many biological experiments involve making measurements on a ***sample*** of individuals in a ***population***. Here, the ***variability*** is not due to random error but because of differences between the individuals in the sample. The variability relates to what is of interest about the sample itself and not about the way the values are measured. (For further details about sampling, see Section 8.2 *Populations and samples* on page 76.)

For example, suppose a group of four girls aged 15 run a 100 metres race. Their times are measured as 14.5 s, 13.9 s, 15.3 s and 14.8 s. Although there will be random errors in these measurements, the variability here is caused mainly by *differences in the individuals themselves*. Some girls can run faster than others.

An answer to the question 'How big is a typical value?' can again be found by calculating the mean – for these data it is 14.6 s. If the girls are reasonably representative of their age group then this value is our best guess of the typical running time for 100 metres of a 15-year-old girl. Of course, this is a small sample and the girls might not be at all representative of the population. So, our guess might not in fact be a *good* guess for the typical running time but, in the absence of any other data, this is our *best* guess.

Not all populations of individuals are biological. In manufacturing, the objects being made can be thought of as populations. For example, in making pistons for a particular car engine, the sizes of each should in principle be identical. In practice, there will be some variability in this 'population', but the aim of the production process is to keep this variability within acceptable limits.

With only three or four individuals, taking a mean is about as much as can be done. However, experiments and surveys that look at samples of populations often collect data on relatively large numbers of individuals. With larger samples, there is more that can be explored, and it is useful to have techniques for seeing patterns in large data sets.

6.4 Displaying larger sets of values

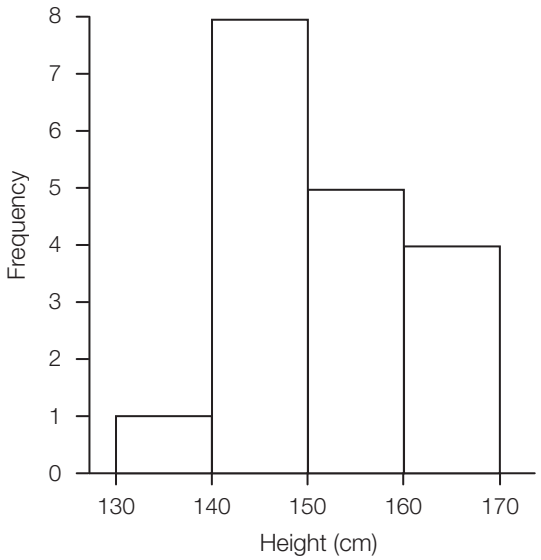
A useful way of displaying the ***distribution*** of a larger number of values is to draw a ***histogram***. For example, the first column in Figure 6.1 shows a set of measurements of the heights of a sample of 12- to 13-year-old pupils (in cm). A data set such as this, which contains a set of values related to a single quantity, is sometimes referred to as a ***batch*** of data. Just looking at the raw data gives a sense of how big the values are, but organising the data enables more to be seen.

The next step here is to order the raw data in order of size. Figure 6.1b shows a column of values going from the highest down to the lowest. (A quick method for ordering sets of values is described in Section 8.3 *Analysing a batch of data* on page 77.) It is now easier to get a sense by eye of the variation in these values.

Figure 6.1 Heights of pupils in a sample (cm)

(a) Raw	**(b)** Ordered	**(c)** Put into classes	
153	168	168	
155	164	164	Frequency = 4
163	163	163	
153	163	163	
164	158	158	
143	155	155	
158	153	153	Frequency = 5
148	153	153	
142	151	151	
168	149	149	
144	149	149	
149	148	148	
163	147	147	
137	144	144	Frequency = 8
149	143	143	
141	142	142	
151	141	141	
147	137	137	Frequency = 1

(d) Histogram



In order to draw a histogram, the next step is to split the set of values into a number of groups or 'classes' – the ***class interval*** is the range of values within each class. The choice of class interval should result in a sensible number of classes, neither too few nor too many. A convenient choice is 130–139 cm, 140–149 cm, 150–159 cm and 160–169 cm.

The data values in each class are counted to give the total number or ***frequency*** for each class as shown in Figure 6.1c. For example, there are four data values in the class interval 160–169 cm (163, 163, 164 and 168), so it has a frequency of 4.

Note that discrete data can also be put into groups like this (for an example, see Section 3.2 *Using tables to process data* on page 24, which discusses the construction of a frequency table from discrete data).

Figure 6.1d shows a histogram drawn from these values. The height of each bar shows the frequency of values (i.e. the number of values) within each class interval (i.e. the range of values of the bar).

This set of data has a fairly small number of values. With a larger number of values, a smaller class interval could be chosen (say 5 cm instead of 10 cm), so that each bar represents a narrower range of values. The histogram would then have more bars and give a better indication of the distribution of values. With only a small number of values, this would not be a good idea since there would not be many values in each of the classes. Choosing an appropriate width for the class interval is entirely a matter of judgement.

It is important not to confuse a histogram with a ***bar chart***, since they are very different. A histogram shows the frequency distribution of a set of values, and the horizontal axis represents a ***quantitative*** (***continuous*** or ***discrete***) variable. To indicate this, the columns are drawn touching each other. By contrast, in a bar chart the horizontal axis represents a ***qualitative*** (***categorical***) variable and the columns are drawn with spaces in between

(see Section 3.4 *Charts showing a quantity categorised by one factor* on page 26 for further details).

Note that there are differences between mathematics and science in the meaning of 'histogram'. For the histogram in Figure 6.1, the width of each column (the class interval) has been chosen to be the same. This is the usual practice in scientific literature and in school science.

In mathematics, pupils learn to draw histograms in which the class intervals (and hence the width of the columns) are different from each other; the vertical axis represents 'frequency density' and not frequency. Frequency density is calculated by dividing the frequency of the class interval by the width of the class interval, and so the frequency for each class is represented by the area of the column and not by its height.

If all the columns are the same width then the shape of the display is the same regardless of whether frequency density or frequency is plotted.

In science, a 'histogram' is generally taken to mean a plot of *frequency*, with all the class intervals being the *same width*. Since this is what pupils will encounter in science textbooks, it is this meaning that is used in this publication. Teachers and pupils need to be aware of this difference in the meaning of histogram in mathematics and science.

6.5 How big is a typical value?

Just by looking at the values for the example above, it is easy to see that a typical value is somewhere between 140 cm and 160 cm. Mathematically, three measures may be used to express this idea of a 'typical value' or 'central tendency' – *mean*, *median* and *mode*:

- ***Mean*** (strictly speaking the ***arithmetic mean***): For this set of values, the sum of the heights is 2728 cm and the number of values is 18, so the mean is 151.6 cm. One problem with a mean is that it can be affected by ***outliers*** (unusually high or low values), since it uses *all* values as part of the calculation.
- ***Median***: This is the middle value of a distribution, and can easily be found once the values have been ordered. If there is an even number of values (as here), the median is the mean of the middle two: these are 149 cm and 151 cm, so the median is 150 cm.
- ***Mode***: In a distribution of ***discrete*** values, the mode is the most common value. Some sets of data may have more than one mode. For the data shown in Figure 6.1 (***continuous*** data, not discrete), the interval with the largest number of values (in this case 140–149 cm) is the *modal interval*.

The mean, the median and the mode are all ways of expressing an ***average***. In everyday language, the word 'average' is generally used as an alternative term for 'arithmetic mean'. In science and mathematics, this usage should be avoided, since an 'average' refers to any measure of a typical value of a distribution.

In summary, three ways of expressing an average for these data are:

- mean = 151.6 cm
- median = 150 cm
- modal interval = 140–149 cm.

The mean is the most familiar and the one most commonly used in school science. An advantage of a median is that, unlike a mean, its value is not affected by *outliers*; a median can also be quicker to find than a mean.

When talking about an average as being a 'typical value', it is important to emphasise that it does not mean 'the most common value'. For example, saying that a typical value for these data is 150 cm does *not* mean that most of these pupils are 150 cm tall. Instead, think of it as just meaning 'roughly how big'. For an alien who did not know whether a 12- to 13-year-old was nearer 10 nanometres or 10 metres tall, saying that a 'typical value' is 150 cm gives a good sense of size.

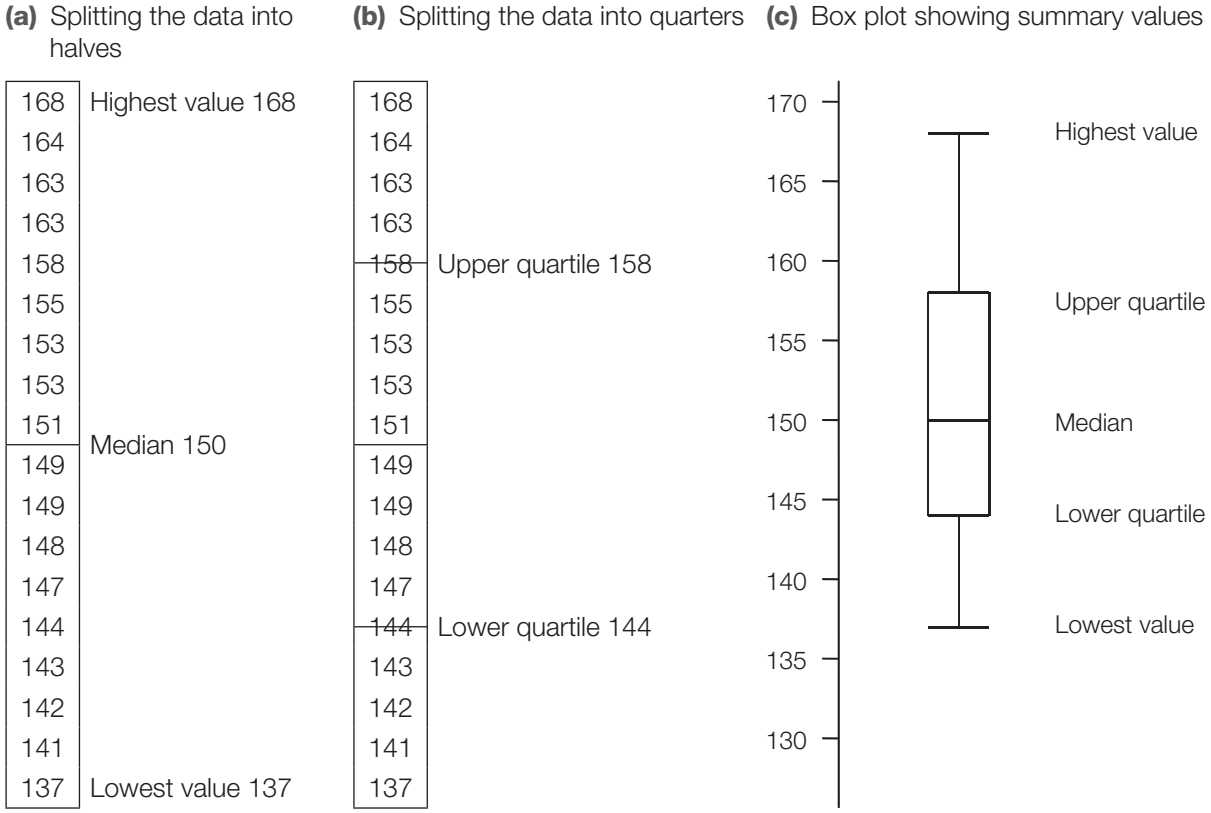
6.6 How much do the values vary?

In addition to having a way of indicating a 'typical value' for a set of data, it is useful to have a measure of how spread out the values are. Two commonly used measures of ***spread*** are the ***range*** and the ***interquartile range***.

The ***range*** is the *difference* between the highest and lowest values. For these data, the range is 31 cm (168 cm − 137 cm). Note that in school science, the 'range' of a set of data is generally used to indicate the lowest and highest values (e.g. the range is 'from 137 cm to 168 cm'). Since this meaning differs from the one used in mathematics and statistics, pupils should be told explicitly what they should do if they are asked to find the range of a set of data (e.g. 'write down the highest and lowest values' or 'calculate the difference between the highest and lowest values'). Although it is easy to calculate, a problem in using the range is that it is affected by extreme values at the high end or low end of the distribution (outliers).

Figure 6.2 shows another way of representing a distribution visually. It is called a ***box plot***, and these types of display are very effective at showing the variation in values. To construct a box plot, the first step is to identify the highest and lowest values, and the median. This is shown in Figure 6.2a.

Figure 6.2 A box plot is another way of displaying a distribution of values



Finding the median can be thought of as splitting the data set into two halves – the median is the value where the data set is split. The next step is to split each of these two halves again in halves. Thus, the data set has now been split into four quarters, and this is represented in Figure 6.2b. The values where the upper and lower halves of the data are split are called the ***quartiles***.

The *upper quartile* is the middle of the top half of the data (in this case, the middle of the largest nine values is 158 cm) and the *lower quartile* is the middle of the bottom half of the data (in this case, 144 cm). These five 'summary' values (highest value, upper quartile, median, lower quartile and lowest value) can now be used to draw the box plot shown in Figure 6.2c. (See Section 8.3 *Analysing a batch of data* on page 77 for further details of how to draw a box plot.)

The central box of a box plot is a better indication of the *spread* of values than the *range*, since it is not distorted by outliers. It represents the ***interquartile range***, which is the difference between the upper quartile and the lower quartile. Here, its value is 14 cm (158 cm – 144 cm).

The line extending above the box represents the upper quarter of the values and the line below represents the lower quarter of the values. Thus, the central box represents *one half of the values in the distribution*: it indicates that the heights of half of the pupils in this sample lie between 144 cm and 158 cm.

Measures of spread are especially useful when comparing two or more data sets, and this is discussed in Section 8.5 *Comparing batches of data* on page 81.

6.7 Comparing shapes of distributions

Box plots and ***histograms*** are both helpful ways of displaying batches of data. The particular strength of the box plot is that when a number of them are drawn side-by-side, the eye can quickly scan across and compare the medians and spreads of different ***distributions***. A histogram shows more detail about a distribution than the five summary values of a box plot but histograms are not as easy to interpret as box plots when comparing two or more batches of data.

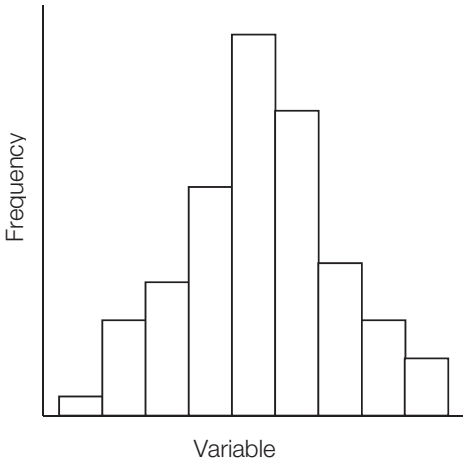
Some distributions may be fairly symmetrical while others may be *skewed* – with values spread out on one side of the middle more than the other. The histogram in Figure 6.3a represents a fairly symmetrical distribution. This kind of distribution is what one might expect when measuring the heights of a sample of pupils of the same age. By contrast, the histogram in Figure 6.3b represents a skewed distribution, with values clustered more at the upper end.

This kind of distribution, with values skewed towards the right, is typical when there is some kind of 'ceiling'. A distribution of the heights of a random sample of people from babies to adults would have a 'ceiling' effect: adults tend not to go over a certain height, while the heights of babies and children would be spread out over lower values.

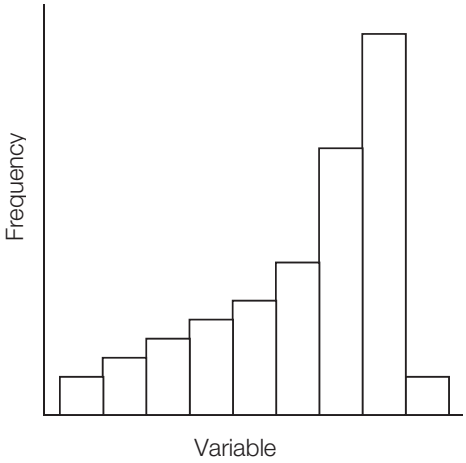
The box plots for these two distributions are shown in Figure 6.3c. The shape of the box plot on the left indicates that it is a roughly symmetrical distribution, since the median is approximately in the middle of the box and the lengths of the lines above and below the box are about equal. The box plot on the right clearly shows a skewed distribution, with the lower part being 'stretched out' and the upper part being 'squashed together'.

Figure 6.3 Comparing distributions

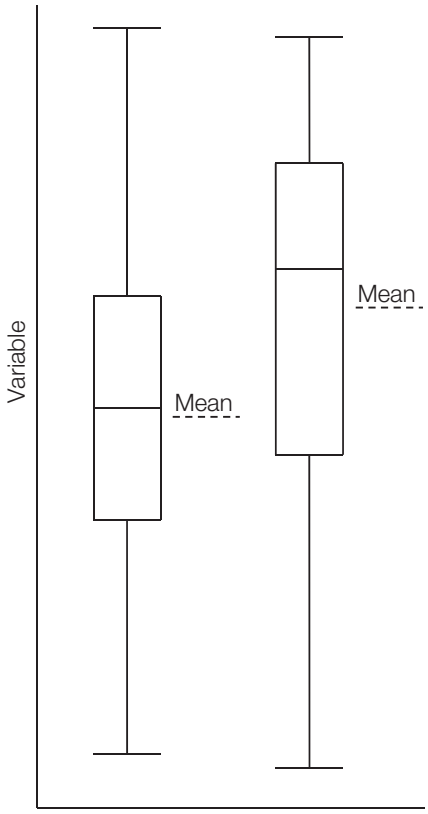
(a) Histogram representing a fairly symmetrical distribution



(b) Histogram representing a skewed distribution



(c) Using box plots to compare the two distributions



It may seem intuitively obvious that, in any distribution, roughly half of the values will be below average, and half above average. However, this is not always the case. If the *median* is taken as the average then it will be true: the median splits the set of data into two, so there are equal numbers of values above and below the median. However, if the *arithmetic mean* is taken as the average then the number of values above and below the mean will depend on the shape of the distribution.

This is illustrated by the box plots in Figure 6.3c. For the first box plot, the distribution is fairly symmetrical and the mean is approximately the same as the median: the numbers of values above and below the mean are about the same. However, for the second box plot, the distribution is *positively skewed* and the value of the mean lies *below* the median: for this distribution there are more values above the mean than below it.

6.8 Are there any unusual values?

An ***outlier*** is a value in a set of data that seems to be unusually large or unusually small in comparison with most of the other values. For repeated measurements, an outlier may be the result of a mistake and is often disregarded. For measurements on a sample of individuals, an outlier may indicate a value that is of particular interest. There are no hard-and-fast rules for how to identify and deal with outliers – it will depend on the context.

Note that 'outlier' is a statistical term that can be used to describe unusual values in *any* kind of distribution. The term ***anomaly*** (or 'anomalous value') is also used in school science, though usually in the context of repeated measurements rather than for individuals in a population. For example, a bowhead whale has a particularly long lifespan for a mammal but it is not generally regarded as being an 'anomaly' (see Section 8.3 *Analysing a batch of data* on page 77).

The heights of pupils in the sample shown in Figure 6.1 ranged from 137 cm to 168 cm. Suppose another value was added to this – a very low one of 48 cm. The value stands out as very different from the rest. One possibility is that it is a mistake – perhaps the actual value was 148 cm and it was written down incorrectly. If a value is a mistake then it should be corrected or removed. Another possibility is that the value is indeed correct but unusual. In this particular example, it is unlikely that there would be a pupil of this height but, in other situations, it is quite possible that there may be an unusual value that genuinely represents a special case. Identifying and displaying such outliers can be useful as there can be interesting and important reasons why they are very different from the rest of the values. (See Section 8.3 *Analysing a batch of data* on page 77 for further details.)

6.9 Basic ideas in probability

A deeper understanding of variability can be gained by using ideas about probability, since randomness is an underlying cause of the variation. Probability is an important topic of study in 11–16 mathematics, though in science at this level it appears very little, except for some basic ideas related to genetics. However, in science itself, probabilistic ideas have a very significant place – both in understanding a wide variety of phenomena and in the design of experiments and handling measurement uncertainty. Pupils will meet these ideas much more in post-16 science.

A simple example of something that produces random outcomes is tossing a coin. The outcome cannot be predicted but there is an equal chance of getting a head or a tail. This is described as saying that the ***probability*** of getting the head is ½, 0.5 or 50%. The probability of getting a tail is the same. The sum of these two probabilities (½ + ½) is equal to 1. A probability of 1 means that something is *certain* to happen, i.e. the coin will either land as a head or as a tail (ignoring the very unlikely outcome that it lands exactly on its edge).

If a coin is tossed and a head is obtained then the probabilities of getting a head or a tail on the second toss are still the same. The outcome of the second throw is not affected by the outcome of the first throw. Each coin toss is therefore called an ***independent event***. (A dependent event is one whose probability is affected by the outcome of another event.) Many people do not find this idea at all intuitive and believe that if a coin is tossed five times and it lands as heads every time then the probability of getting a tail the next time will be much higher. However, since these are independent events, the probability of a head is still ½ (assuming it is a 'fair' coin).

If you toss a coin a very large number of times (say a billion times) then you would expect the proportions of heads and tails to be very close to 50% of each. However, if you only toss a coin 10 times (a small sample size), you might get 5 of each, but it is also quite likely you will get a different proportion, such as 4 heads and 6 tails. The larger the sample of tosses, the more likely you are to get closer to half being heads and half being tails.

If a coin is tossed twice, there are a total of *four* possible outcomes: HH, HT, TH and TT. These are shown in Figure 6.4 (this is known as a *sample space*).

Figure 6.4 Outcomes of tossing a coin twice

		2nd toss	
		H	**T**
1st toss	**H**	HH	HT
	T	TH	TT

A pair of coin tosses, consisting of two separate events, is an example of a ***combined event***. The probability of a combined event can be calculated by multiplying together each of the probabilities of the separate events but *only* if these are *independent events*. So, for a pair of coin tosses, the probability of getting a head on the first toss is ½ and the probability of getting a head on the second toss is also ½. The probability of throwing two heads is therefore ½ × ½ = ¼. The probabilities for each of the four outcomes are shown in Figure 6.5a. All of them are ¼. This means that in a very large number of coin tosses, you would expect there to be 25% of each of the four combinations, as shown in Figure 6.5b.

Figure 6.5 Probabilities and percentage frequencies

(a) Probabilities for each combination

	H	**T**
H	¼	¼
T	¼	¼

(b) Outcome for very many pairs of coin tosses

	H	**T**
H	25%	25%
T	25%	25%

(c) A possible outcome for 100 pairs of coins tosses

	H	**T**
H	30%	19%
T	27%	24%

For a smaller number of pairs of coin tosses, you would not expect to get 25% of each combination. Figure 6.5c shows a possible outcome for 100 pairs of coin tosses. As the number of pairs of tosses increases, the more likely it is that the proportions will approach 25% of each.

These ideas about the outcomes of coin tossing are a direct analogy for the way that the outcomes of genetic crosses are predicted (and as represented using a Punnett square). The outcome of a single coin toss corresponds to a particular allele, while the pair of coin tosses corresponds to the pair of alleles inherited from each parent.

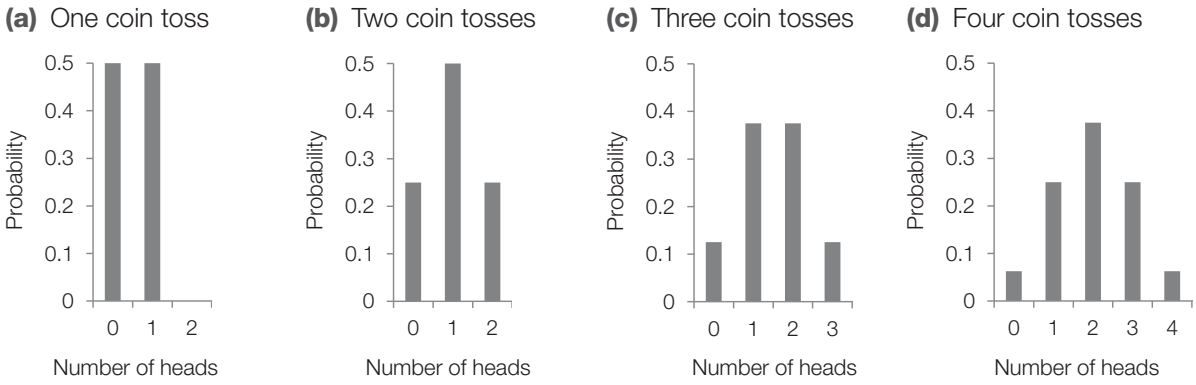
So far, we have looked at the outcomes of a single coin toss and of a pair of coin tosses. This can be continued, looking at the probabilities of each of the outcomes for three, four or more sets of coin tosses. The mathematical calculations get rather more difficult but the principle is just the same. Figure 6.6 shows the probabilities of possible outcomes for various numbers of coin tosses, expressed in terms of 'number of heads'.

For example, Figure 6.6a shows that, for a single coin toss, there are two outcomes: 0 heads (i.e. a tail) or 1 head. The probability of each is 0.5.

Figure 6.6b shows that, for two coin tosses, there are three outcomes: 0 heads (i.e. TT), 1 head (i.e. HT, TH) and 2 heads (HH). These probabilities are 0.25, 0.5 and 0.25, respectively.

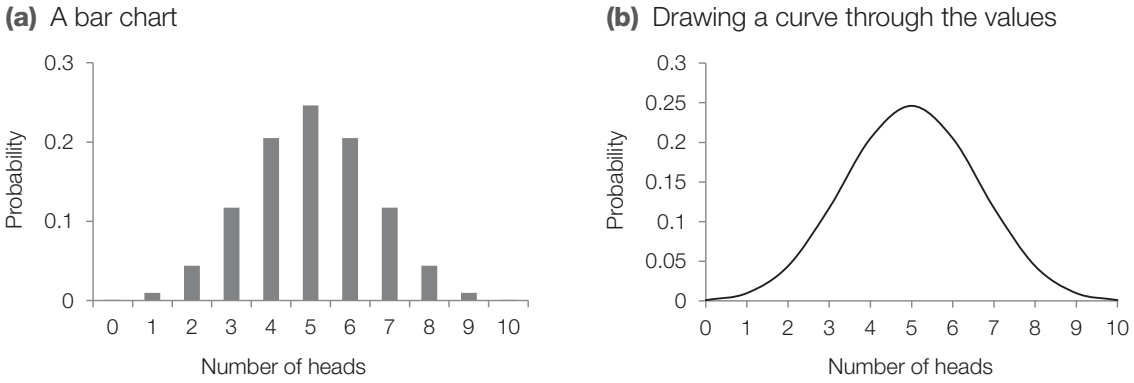
Figure 6.6c shows the outcomes for three coin tosses and Figure 6.6d shows the outcomes for four coin tosses. For sets of four coin tosses, the most likely outcome is to get 2 heads (and thus 2 tails). Much less likely, though still quite possible, is to get 0 heads or 4 heads.

Figure 6.6 Probabilities of outcomes from various numbers of coin tosses



We can carry on looking at larger and larger numbers of coin tosses. Figure 6.7 shows the probabilities of all the outcomes for sets of 10 coin tosses. It shows that 5 heads (and thus 5 tails) is the most likely but several other combinations can occur quite often; for example, getting 3 heads has a probability of over 0.1. The probability of getting 9 or 10 heads in 10 throws, however, is very low.

Figure 6.7 Probabilities of outcomes from 10 coin tosses



With this larger number of values, it is interesting to see the shape of the graph obtained by drawing a curve through these data points (Figure 6.7b). This shows the classic 'bell-shaped' curve – the sort of distribution that you get, for example, by measuring the heights of people in a population. Most people are of middling height with smaller numbers of very short or very tall people.

The shapes are similar for the same underlying reason. The heights of people, like the outcomes of coin tossing, are determined by a combination of many random events. The same is true for the variability due to measurement uncertainty (e.g. repeated measurements of the time for a marble to run down a slope) and for the natural variability in the characteristics of any population (e.g. the times for different people to run a 100 m race).

6.10 Estimating risks

In everyday language, the words 'hazard' and 'risk' are used more or less interchangeably with the word 'danger'. However, they are also technical terms with precise meanings. A *hazard* is something that is potentially harmful to people, property or the environment. A ***risk*** relates to the *probability* of harm occurring when exposed to a hazard. For example, a river near a house is a *hazard* because it can potentially cause flooding. The risk of the land around the house being flooded by the river might be assessed as a 1% annual probability.

There are essentially two ways in which risks can be assessed. Thinking about coin tosses is a simple way to illustrate this. If we want to know what the 'risk' is of getting two heads when we toss two coins, there are two ways of working this out.

One method is to collect a lot of data, by noting the outcomes of many pairs of coin tosses. The more coin tosses we do, the more likely the estimate is to be accurate. With a very large number of pairs of coin tosses, the 'risk' of getting two heads is found to be ¼.

The other method is to work out the probability from what we know about the behaviour of the coins. For each coin toss, the probability of getting a head is ½. Thus the 'risk' of getting two heads is $½ \times ½ = ¼$.

In the real word, both of these methods are used to calculate risks. When a lot of data is available, the first method can be used. The risk of getting lung cancer from smoking, or the risk of certain injuries in road accidents, can be found by analysing statistics that are routinely collected.

Other risks relate to events that happen rarely or have never happened. Calculating the risks of damage to a nuclear power station due to an earthquake, or the risks to human health on a mission to Mars, cannot be calculated by analysing large data sets of previous cases. Instead, they are calculated by combining the estimated probabilities of the events that lead to the outcomes being assessed.

Estimating risks in the real world is of course much harder than working out probabilities in coin tossing. For example, despite having a great deal of data about people's health and behaviour, it took a long time and much analysis before the link between smoking and lung cancer was established. In the case of nuclear power stations, different people may make different estimates of risk because they make different assumptions about the contributions to the risk.

6.11 Interpreting reports about risk

Everything that we do entails ***risk***. Running, walking or any kind of physical activity involve the risk of injury, while lack of exercise is a risk to health. In order to make decisions, one needs to weigh up the sizes of different risks. There are frequent media reports concerning risk, such as the impact of new drugs or the effects of diet on health. However, people may have difficulties in understanding the figures, and the studies are often reported in a way that makes them hard to interpret. This can have a serious effect on the ways that risks are perceived.

For example, here are some headlines about cancer risks:

- 'CT scans in childhood can triple the chance of developing brain cancer'
- 'One drink a day increases breast cancer risk by 5%'.

The figures themselves may not be easy for many people to compare because they are expressed in different ways. A 'tripling' of a risk can re-expressed as a '200% increase' (conceptually not a simple calculation) and, put this way, perhaps it sounds more alarming (after all, '200' is a bigger number than '3'). When both figures are expressed as percentages, they can be compared more easily:

- 'CT scans in childhood can increase the risk of developing brain cancer by 200%'
- 'One drink a day increases breast cancer risk by 5%'.

A casual reading of these figures might suggest that childhood CT scans are a much bigger cancer risk than having one drink a day (40 times bigger).

However, the problem is that the figures in these headlines are referring to *relative risk* and not *absolute risk*. Your risk of being hit by a meteorite is extremely small indeed (an *absolute risk*); if for some reason this risk increases 100 times (a *relative risk*), this would still represent a very small risk of being hit. So, it is hard to interpret a relative risk without knowing the size of the underlying absolute risk.

For the two examples above, the relevant absolute risks are:

- about 1 in 10 000 children aged 0–9 develop brain tumours or leukaemia
- about 11% of women who do not drink develop breast cancers.

Again, some care is needed in comparing these figures for the absolute risk. Some people think that '1 in 10 000' is a bigger risk than '1 in 100' (since 10 000 is a big number), or that '1 in 10' is smaller than '5%'. Re-expressing the second risk allows the two risks to be compared more easily:

- about 1 in 10 000 children aged 0–9 develop brain tumours or leukaemia
- about 1100 in 10 000 women who do not drink develop breast cancers.

The story presented here has been simplified, but the principle should be clear to see. The *absolute risk* in the second example is far, far bigger than the first, even though the *relative risk* is lower. Using all the available data and some intricate calculations, studies have worked out the increase in numbers of cases:

- an extra two cases of cancer for every 10 000 children given CT scans
- an extra 60 cases of breast cancer for every 10 000 women who regularly have a drink.

The 'tripling' of cancer due to CT scans now seems to be less alarming than the headline may have suggested to many people. Although the relative risk is very high, the numbers affected are fairly small because of the low underlying absolute risk.