www.wellcome.ac.uk/bigpicture/numbers

JUST ABOUT AVERAGE

There are different types of average

When we talk about 'an average', what we're really trying to do is get some sense of where the middle is. We can then use that as a way of comparing two groups of data. Unfortunately, there isn't just one type of average – there are several.

To get the mean, add the data together and divide the total by how many data points there are (see equation below). Beware, though – outlying data can often skew the mean to be artificially high or low. Take the following number list: 1, 3, 6, 9, 9, 11, 14. The mean here is 7.57. However, add a much higher number to the end of the list, say 50. The mean is now 12.88. Just one particularly large outlier has almost doubled the mean, and the majority of the numbers are below the mean. Imagine how the mean wealth of biology teachers in a room might change if Bill Gates joined them.

If you place the numbers in ascending order and look for the middle value in the list, you have the median. If there are an even number of values, you take the mean of the middle pair. For the original list, this is 9. Outliers have a much smaller effect on the median than the mean, so adding 50 again does not alter the median.

The mode is the value that occurs most often in a list. For this list, the mode is 9.

WHAT'S YOUR TYPE?

Not all data are the same

Researchers define data in different ways. For example, data are **categorical** if the values can be sorted into non-overlapping categories (e.g. by blood type, species or sex). Every value should belong to only one category, and it should be clear which one it belongs to. Categorical data are also known as 'nominal data', or 'frequencies', as the research looks to find out how frequently data fall into each category.

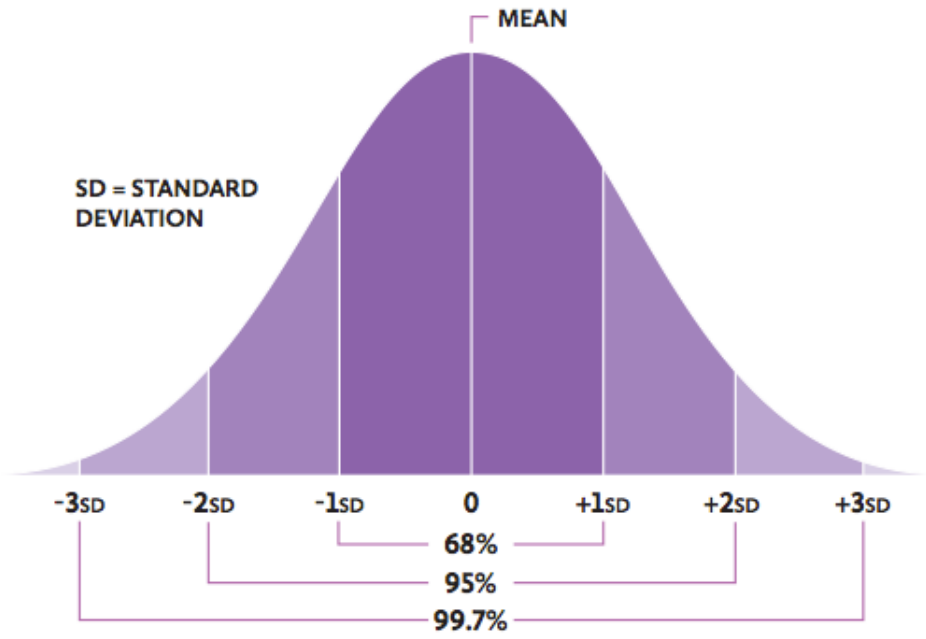
Ordinal data, by contrast, can be ranked or have some sort of rating scale. Ordinal data often come from surveys and questionnaires.

Data can also be defined as discrete or continuous. Data are **discrete** (or discontinuous) if they can take only isolated values. **Continuous** data can take on any value and are limited only by how accurate your measurements are. So, while foot length is continuous, shoe size is discrete because you can't be a size 7.234434 – you have to be a 7 or a 7.5.

WHAT IS NORMAL?

Many things follow a normal distribution

Datasets can be spread out in many different ways. The majority of the data can sit above the mean or below it. In many datasets, however – particularly large ones – the data points seem to settle equally on either side of the mean. Plotted on a graph, the shape of the distribution resembles a bell and so is sometimes called a 'bell curve'. This is also called a normal distribution (see graph below).



Standard deviation is a measure of how spread out the numbers are around the mean. If a dataset follows a normal distribution, approximately 68 per cent of the data will fall between one standard deviation on either side of the mean. Around 95 per cent will fall within two standard deviations on either side. In such circumstances, the mean, median and mode of the data are all equal.

There are many everyday biological examples that follow a roughly normal distribution, including blood pressure, height and foot length. Along with these examples, you could also try looking at stalk height in daisies or the length of holly leaves.