

## FACT SHEET

# Background to Statistics

---

### Introduction

'Statistics' include a broad range of methods for manipulating, presenting and interpreting data. Professional scientists of all kinds need to be proficient in the use of statistics: they must be able to recognise and communicate what their data show.

Advanced level coursework therefore requires you to demonstrate the ability to use the statistical analysis of data effectively. You must make calculations and draw graphs or charts to show and identify patterns and trends, including central tendencies and spread. You need to carry out tests to confirm relationships. Applied science presents many opportunities for the presentation and analysis of data. The ability to use *graphs* and **descriptive statistics** is of particular significance in AQA Unit 7 *Planning and carrying out a scientific investigation*, AQA Unit 16 *Ecology, conservation and recycling*, OCR Unit 8 *Investigating the scientist's work* and OCR Unit 9 *Sampling, testing and processing*.

OCR Unit 14 *Ecology and managing the environment* additionally requires the use of **inferential statistics**. You can also use these elsewhere, to make data analysis and the formation of valid conclusions more effective. Inferential statistics help you to form conclusions about *populations* of data from *samples* – you don't need to obtain every single possible result to be able to form a valid conclusion. You can use your analysis to **support or reject a hypothesis** (testable statement) and establish how confident you can be in forming your conclusions. If you use statistics well, you can collect the minimum amount of data and reduce the amount of work that you have to do!

Statistical techniques include:

#### *graphs and descriptive statistics*

- scatter graphs
- bar charts
- histograms
- error bars
- mean, mode and median
- variance and standard deviation
- standard error of the mean
- confidence intervals

#### *inferential statistics*

- association statistics
  - correlation coefficients
  - $\chi^2$  test (chi-squared) for association
- comparative statistics
  - t-tests for paired or unpaired data
  - Mann-Whitney U-test
  - analysis of variance (ANOVA)
- frequency statistics
- $\chi^2$  test (chi-squared)

## Graphs and descriptive statistics

---

### Average

'Average' is not a precise mathematical term. It can be:

mean ( $\bar{x}$ ) = sum of values divided by number of values

mode = most frequent value

median = middle value for results listed in order

The *mean* is the arithmetic average, found by adding up all your values and dividing by the total number of separate values. It is the most often used 'average' and can be used in further statistical analysis. However, it can be distorted by a single extreme value.

The *mode* is the most frequently occurring value or event. It is not useful for further statistical analysis but, unlike the mean, it is not affected by extreme values.

The *median* is the central value (or average of the two central values in an even number of values) when all the values are placed in order of magnitude (ranked). It is not useful for further statistical analysis but, unlike the mean, it is not affected by extreme values.

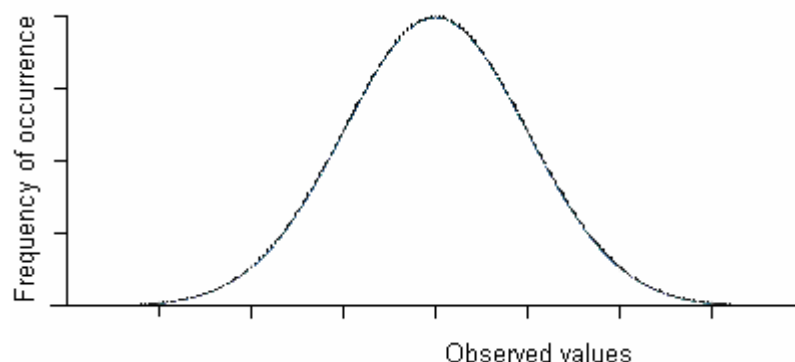
Choose by deciding what you are going to use the statistic for: does the value give you the information you need?

### Sampling

A population contains *all* the possible results, so usually a *representative sample* is obtained, i.e. representative of the population as a whole. The aim is to use the smallest sample that will reliably give the same statistics as the population.

### Variance, standard deviation and normal distribution

The mean, mode or median may be the same for sets of data that spread out in different ways. Finding the **range**, the difference between the highest and lowest values, is a useful measurement for comparing *spread*. However, this does not indicate the *pattern of variation* in the values. Many measurements, especially biological measurements, such as adult human heights, show a **normal distribution**. They cluster around the mean with fewer and fewer values as you move further away on either side, giving a characteristic *bell curve* :



The spread of the values is symmetrical about the mean, so the mean, mode and median all coincide at the same value.

**Standard deviation** is a measure of the spread of the values. For a sample it is usually given the symbol **s**. Sometimes the population symbol  $\sigma$  (Greek letter sigma) is used. Standard deviation is the square root of the **variance**. The variance is the average of the squared differences from the mean of each of the values.

For example, if you have two values, 3 and 9:

$$\text{mean, } \bar{x} = \frac{\sum x}{n} = (3 + 9) \div 2 = \mathbf{6}$$

( $\sum x$  = the sum of all the different values,  $n$  is the number of values)

differences (deviations) from the mean,  $\bar{x} - x = 3$  and  $-3$

deviations squared,  $(\bar{x} - x)^2 = 9$  and  $9$

so average deviation or **variance** =  $\Sigma(\bar{x} - x)^2 / n = 18 \div 2 = \mathbf{9}$

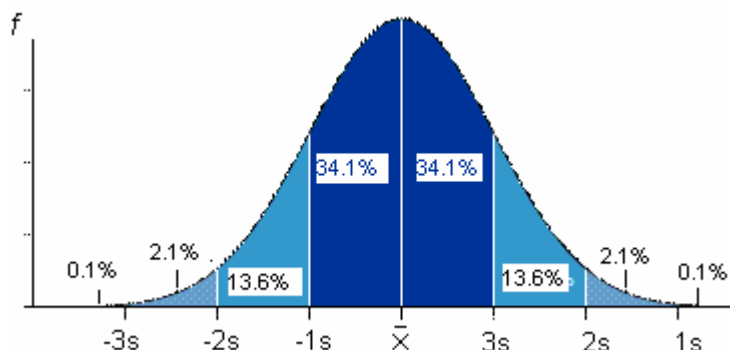
and **standard deviation** (square root of the variance),  $s = \sqrt{[\Sigma(\bar{x} - x)^2 / n]} = \sqrt{9} = \mathbf{3}$ .

If you have a *statistical calculator*, you may be able to use a standard deviation key to obtain its value. There are usually two keys labelled  $\sigma_n$  and  $\sigma_{n-1}$  ( $\sigma$  is used, even though the keys are only for sample calculations). The  $n-1$  refers to finding an 'unbiased' value for the sample standard deviation: calculations based on small samples (<25 values or so) tend to slightly underestimate the true value of the standard deviation.

Should standard deviation be calculated using  $n-1$ ? You can use  $\sqrt{[\Sigma(\bar{x} - x)^2 / n-1]}$  if you have fewer than about 25-30 values in your sample, but this is not essential. Some statisticians think that if you need to do this, it is because your data are not good enough in the first place!

Variance is measured in units squared, but standard deviation restores the measurement to the *same units* as the original values, making it much more useful. It is the root mean square (RMS) of the deviations of the values from their mean.

The standard deviation is the most commonly used measure of how widely spread values are. If many data points are close to the mean, then the standard deviation is small.



Modified from an original file created by Petter Strandmark, based on an original graph by Jeremy Kemp  
[http://en.wikipedia.org/wiki/Image:Standard\\_deviation\\_diagram.svg](http://en.wikipedia.org/wiki/Image:Standard_deviation_diagram.svg)  
<http://creativecommons.org/licenses/by/2.5/>

If many data points are far from the mean, then the standard deviation is large. If all the data values are equal, then the standard deviation is zero.

The *68-95-99.7* or *empirical rule* for a normal distribution:

- 68% or roughly 2/3 of the observations fall within 1 standard deviation of the mean
- 95% of the observations fall within 2 standard deviations of the mean
- 99.7% of the observations fall within 3 standard deviations of the mean

## Inferential statistics

---

### Statistical tests, probability and significance

If you measure limpet shells at two different rocky shores, you may find that your mean values are different. Statistical tests allow you to decide if the difference is big enough to be *significant*.

This is based on the idea that results *can be influenced by chance as well as caused by an effect*. Even if you measured two samples of shells in the *same* area, you would expect a difference between the means of your samples, *due to chance alone*.

So you need to decide if the observed difference is a *real* difference. You must test the hypothesis that there is a *significant* difference in shell diameters between the limpets, not due to chance alone.

For statistical tests, a **null hypothesis** must often be used (usually given the symbol  $H_0$ ). This is a hypothesis that you try to *reject* so that your original hypothesis (the **alternative hypothesis**,  $H_A$  or  $H_1$ ) is supported.

In this case, you could test the null hypothesis: 'There is *no significant difference* between the diameters of the limpet shells found in the two places, any apparent difference is due solely to chance'.

If you can reject this hypothesis, the converse alternative is supported: 'There **is** a *significant difference* between the diameters of limpet shells in the two places'.

To make the decision, you use statistical tests to find the **probability** ( $p$ ) of getting a particular set of results if a hypothesis is true.

$p = 1$  represents certainty,  $p = 0.5$  is 1 in 2 or 50% and  $p = 0.1$  is 1 in 10 or 10%.

Probability can be worked out by considering how many times an event can occur and dividing by the total number of all possible events. Hence, the probability of obtaining a 'head' when tossing a coin is  $\frac{1}{2}$ , i.e.  $p = 0.5$ . For being dealt an ace at random from a pack of cards,  $p = \frac{4}{52} = 0.077$ .

**Significance** in biological investigations is usually set at  **$p \leq 0.05$**  (probability equal to or less than 0.05).

Looking at the data that you would expect if the hypothesis is true, at  $p = 0.05$  you would expect to get that type of data 1 in 20, or 5% of the time. This probability is chosen because more than 1 in 20 is thought to be reasonably likely or better. At a probability of 1 in 20 or less, events become more and more unlikely.

- If  $p \leq 0.05$  the hypothesis can be rejected with confidence. The data are unlikely and don't support the hypothesis.
- If  $p > 0.05$  the hypothesis cannot be rejected with confidence. The data are as expected; it is reasonably likely that the hypothesis is true.

### Confidence levels

You can use the  $p \leq 0.05$  cut-off level or, if a probability value can be found, the smaller the probability becomes the greater your confidence:

- $p \leq 0.05$  reject with confidence
- $p \leq 0.01$  reject with a high level of confidence
- $p \leq 0.001$  reject with a very high level of confidence.

### Is $p \leq 0.01$ better than $p \leq 0.05$ ?

When you reject a hypothesis at  $p = 0.05$ , on average you will be wrong once every 20 times. *Why not set a stricter significance level in the first place, say  $p \leq 0.01$ ?* The trouble with biological data is that it can vary quite a bit and give inconsistent results. If you make significance too difficult, it becomes more likely that you will fail to support hypotheses for effects that are really there.

$p \leq 0.05$  has been found by experience to be a good 'rule of thumb' that seems to work well for biology investigations where results may vary quite a lot. In physics or chemistry investigations data may be much more precise. With more highly controlled variables and accurate measuring instruments,  $p \leq 0.01$ ,  $p \leq 0.001$  or even more exacting confidence levels are more appropriate. It is important to choose your minimum confidence level before you have analysed your data! If your data show strong support for a hypothesis, you can report that you have exceeded the confidence level *after* the choice has been made and justified.

### Parametric and non-parametric tests

You should choose a 'parametric' test for your data if you can show (or reasonably assume) that your data are measured using a continuous interval scale and are more or less normally distributed. The mean, mode and median should coincide or a histogram should give a 'bell curve'. These tests include *Pearson's correlation coefficient* and *t-tests*. Parametric tests are more powerful than non-parametric tests because they require more precise data and make use of the extra assumptions that can be associated with the normal distribution of data.

You should choose non-parametric tests if data are not reasonably normally distributed or not measured on a continuous interval scale (divisible into smaller equal sized units). These include Spearman's correlation coefficient,  $\chi^2$  (chi-squared) tests and the Mann-Whitney U-test. These tests will work for data where ranks or arbitrary scores have been awarded rather than using accurate measurements using small scale divisions.