

Maths/stats support 9 What test should I use?

Getting started with statistics

When undertaking any investigation, much mental anguish will be saved if you can go straight into using an appropriate statistical test because you have planned what you are going to do and collected the right sort of data. There are several types of statistics including:

- **Descriptive statistics** such as the mean or the median; they tell you something about a set of data (i.e. the average and middle values in this case).
- **Hypothesis testing statistics** such as a t-test or a χ^2 test; these tests compare one set of data with another set of data.

Maths/stats support sheet 7 Averages (M0.07S) and Maths/stats support sheet 10 Standard deviation (M0.10S) describe some descriptive statistics. This sheet and all the following Maths/stats support sheets, M0.11–M0.14, are about hypothesis-testing statistics. There are several different statistical tests you can use depending on what you are investigating, but they all test a hypothesis and include the following steps.

Invent a null hypothesis

In normal English a hypothesis is a statement that you think explains your observations. For instance, you are walking down the street and notice three giraffes, a stoat and an elephant fighting outside a pub. What could possibly explain this bizarre phenomenon? You could invent several hypotheses:

Hypothesis 1: Someone has spiked your mineral water with hallucinogenic drugs.

Hypothesis 2: They're filming the 14th series of *Rolf's Animal Hospital* and it's all gone horribly wrong.

Hypothesis 3: The Chief Zookeeper has left the main gate unlocked again.

These are all possible explanations for your observations.

In hypothesis-testing statistics you start with a special sort of hypothesis that you set up purely for the purpose of the test. It is called a **null hypothesis** (symbol: H_0).

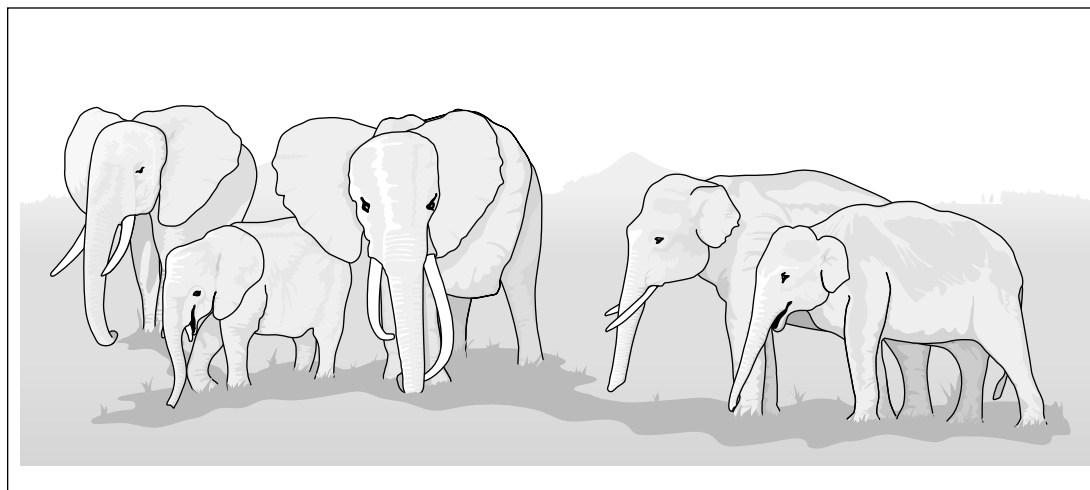
All of the tests compare one set of data with another set of data. A null hypothesis is a statement that says that there is no significant difference between the two sets of data.

This statement 'of no significant difference' is always made even if you think there really is a significant difference.

Trunk lengths: an example

Let us imagine you are comparing the mean lengths of trunks of African elephants and Indian elephants. It looks pretty clear to you from Figure 1 on the next page that the African ones are longer by far. However, you would state your null hypothesis as follows:

There is no significant difference in mean trunk length between African and Indian elephants.



▲ **Figure 1** How are these elephants different?

Always stating there is no difference may seem peculiar until you get used to it. The point is that in the end you get to either *accept* the null hypothesis if there is no significant difference, or *reject* it if there is a significant difference. Even more thrillingly, you get to pick how certain you want to be that you are correct in accepting or rejecting the null hypothesis; this is the real value of such tests.

Null hypotheses are sometimes called hypotheses of no difference because you always state them this way. Null means nothingness, lacking distinction, characterless, i.e. no difference.

Calculate the value of the test statistic

The value of the test statistic is the number you calculate from your data. It will have a name that depends on the test you are using (e.g. 't' or 'r_s'). It is called the **calculated value** of the test statistic. The procedures for working these numbers out are to be found on the Maths/stats support sheet for the individual tests (10 Standard deviation (M0.10S), 11 t-test (M0.11S), 12 Spearman's rank correlation (M0.12S), 13 χ^2 (Chi-squared) test (M0.13S), 14 Mann-Whitney U test (M0.14S)).

Work out the critical value of the test statistic

Next, you obtain a table of **critical values** for the appropriate test (supplied by statisticians who have worked out all combinations of circumstances and sets of data). You extract from the table the critical value that applies to your combination of circumstances, which depends on the size of your sets of data and the degree of precision you require in your acceptance or rejection of your null hypothesis.

At this point you need to have chosen how certain you want to be that you are correct in your acceptance or rejection of H_0 . Many tables of critical values have significance levels (given as percentages or probabilities along the top). You can pick any of them but the smaller the percentage significance level you pick, the more certain you will be that you are correct in rejecting H_0 .

For example: If you chose to test at the 1% significance level ($p < 0.01$), on 99% of the occasions that you reject H_0 you are likely to be correct. About 1% of the time you will conclude that a significant difference exists when it actually doesn't. It's rather like saying you're 99% certain if you find a significant difference that you are correct.

In biology, because of the inherent variability of biological systems, it is generally accepted that a significance level of 5% is absolutely fine. This means that if you did the same investigation lots of times you'd expect the same result when you perform the statistical test 95% of the time and a different result due to chance 5% of the time. It's like saying you're 95% certain that if you reject H_0 you are correct.

Compare the calculated value with the critical value

You now compare the calculated value with the critical value; depending on which one is the bigger, you either accept or reject your opening null hypothesis. You do this in the happy knowledge that you would expect this result 95% (or whatever significance level you pick) of the time if you repeated the investigation forever.

Deciding which statistical test to use

Beware, there is nothing worse than doing your experimental work or fieldwork only to find that the results you have collected are useless for the statistical test you need to use to test your hypothesis. So, once you have come up with your idea for an investigation (often the hardest part), done your background research and formulated your null hypothesis, you need to decide what sort of data you will collect. You need to know what sort of data you will have before you can select the correct test. There are several types of data.

Interval data

This is best thought of as 'real' measurements of things. Length is a good example of interval data. If you have two pieces of data and you can say how much bigger or smaller one is than the other, then you have interval data. For example, an elephant's trunk of 2 metres length is 1 metre longer than an elephant's trunk of 1 metre length.

Ordinal data

Let us suppose that you were surveying elephant's trunks but were too nervous to get close enough to measure them directly. You could devise a scale thus:

- 1 = embarrassingly small
- 2 = short
- 3 = medium
- 4 = long
- 5 = very long
- 6 = probably works for Fire Brigade.

You could assess trunk lengths at a safe distance and you could say that a trunk rated at level 3 was shorter than a 4 (or whatever) but you could not say by how much. This is an example of *ordinal* data, where the numerical value gives the relative position in a series but does not tell you about the size of the interval between the measurements.

Categorical data

Some kinds of data comprise counts of things in categories; these are categorical data. For example, during your elephant trunk studies you notice that some animals have flared nostrils (possibly as a result of over-zealous water squirting) and some have compact nostrils. Each individual belongs in one of these two categories and you could count the frequency of occurrence of each type; e.g. frequency of flared nostril individuals = 20, frequency of compact nostril individuals = 22.

The tests

t-test

A t-test will tell you if the means of two sets of normally distributed (symmetrical about the mean), unmatched, (see Spearman's rank correlation below for what this means) continuous data with interval level measurements are significantly different to one another. (If you have a big sample, 25+, you can use it for ordinal data as well.)

For any t-test you do the null hypothesis will be:

There is no significant difference between the means of the two sets of data.

Examples of where you might use a t-test:

- comparing mean heights of limpets on two different seashores
- comparing mean masses of plants grown with and without fertiliser
- comparing mean tree heights on North- and South-facing slopes
- comparing mean vegetation heights on trampled and untrampled areas
- comparing mean vitamin C contents of pasteurised and unpasteurised orange juice.

Spearman's rank correlation

A Spearman's rank correlation test will tell you whether two variables are correlated, i.e. whether a change in one variable is accompanied by a change in the other variable. It will tell you whether the relationship is a positive correlation (both go up together) or a negative correlation (one goes up as the other goes down) and the strength of any correlation.

For any Spearman's rank correlation you do, the null hypothesis will be:

There is no correlation between the two variables.

You can use this test on interval or ordinal data and the data will always be in matched pairs. This means that one piece of data is associated with one other piece of data only, for example, if you were measuring temperature and water depth, each temperature measurement would belong with only one specific depth measurement (both taken at the same place). If you mixed the matched pairs up the data would be meaningless.

Examples of where you might use Spearman's rank correlation:

- Is there a correlation between temperature and height up a mountain?
- Is there a correlation between mouse density and proximity to a cheese factory?
- Is there a correlation between current speed and mayfly nymph abundance?
- Is there a correlation between cigarette smoking and low intelligence?
- Is there a correlation between species diversity and height on the seashore?

χ^2 test (Chi-squared test)

A χ^2 test does a lot of things but it can be used in a simple way to see if an observed set of data (categorical data, counts of things in categories, i.e. frequencies) differs significantly from what we might expect, given our null hypothesis. For example, it can be used in a genetics experiment to compare the observed data with what might be expected from a cross between two heterozygotes.

For any χ^2 test you do the null hypothesis will be:

There is no significant difference between the observed and the expected frequencies.

Examples of where you might use a χ^2 test:

- Do seashore snails actively select specific microhabitats?
- Does lichen frequency differ between air-polluted and clean sites?
- Are the fruit flies in the lab heterozygous?

Mann-Whitney U test

The Mann-Whitney U test tests for differences between population medians. You can use it on interval or ordinal level data that are not in matched pairs (see Spearman's rank test above). The data don't have to be normally distributed but the distributions have to be the same shape. It can be used on a great many datasets and is one of the most popular statistics used in A Level coursework. It has the great advantage in that it's fairly easy to do as well.

For any Mann-Whitney U test you do the null hypothesis will be:

There is no significant difference between the medians of the two populations.

Examples of where you might use a Mann-Whitney U test:

- To compare median trunk lengths in two populations of elephants. (You could use interval or ordinal data as described in the example on page 3.)
- To compare median body masses of members of 'Chocolate Guzzlers Anonymous' and the 'Lettuce Nibblers Association'.
- To compare the median density of limpets on the middle shore with the median density of limpets on the upper shore.
- To compare the median diversity of lichens in the town centre with the median diversity of lichens in the countryside.

Which test?

Consider the following questions when deciding which test to use.

- 1 Do you want to find a correlation between two variables with data in matched pairs? **Yes**, then use Spearman's rank correlation.
- 2 Do you want to compare your observed data with expected values predicted using a biological theory? **Yes**, then use the chi-squared test.
- 3 Do you want to compare two sets of data with less than 25 measurements in each dataset (unmatched, interval or ordinal data with normal or non-normal distribution)? **Yes**, then use the Mann-Whitney U test.
- 4 Do you want to compare two sets of data with more than 25 measurements in each dataset (unmatched, interval data with normal distribution)? **Yes**, then use a t-test.

Q1 For each of the examples below, think of a null hypothesis and which statistical test might be appropriate, before checking with the answers at the bottom of the page.

- a Investigating the surface area of leaves growing in sun and shade.
- b Investigating the effect of substrate concentration on the activity of the enzyme catalase.
- c Investigating the preference of woodlice for different conditions using choice chambers.
- d Investigating the effect of salt concentration on the germination of seeds.

Answers

- Q1**
- a There is no difference between the surface areas of leaves growing in the sun and shade. Measure surface areas of at least 25 leaves at the two sites. If data normally distributed use a t-test; if not normally distributed use a Mann-Whitney U test.
 - b There is no correlation/relationship between substrate concentration and catalase activity. Measure the enzyme activity using a range of substrate concentrations (try to get 12 or more pairs of measurements). Use Spearman's rank correlation.
 - c There is no significant difference between the observed and the expected frequencies. Collect frequency data on woodlice preference for the different conditions using choice chambers. Use the χ^2 test.
 - d There is no correlation between salt concentration and germination of seeds. Measure the percentage germination at a range of salt concentrations (try to get 12 or more pairs of measurements). Use Spearman's rank correlation.