Statistical Tests (Inferential Statistics) & p values

You can use statistical tests to look for **differences** or **associations** in your data or to make **predictions**.

In most cases you will have data from a **sample** of the **population** that you are interested in and want to use your sample to make predictions about the population.

For instance, you might have found that the mean of a particular measurement in your data is higher in one group than in another. A statistical test is used to see if this difference is just a chance outcome of your sample or whether the difference is **significant**. One way to think about this is that you are using the test to see how likely it is that the difference would be seen in the population as a whole.

When you run a statistical test using software the program will give you a **p value (sig value).** It is important to understand what this is telling you. All statistical tests start with the assumption that there is no difference, or no association, in the data – this assumption is called the **null hypothesis.** The test then uses calculations (usually done behind the scenes by the software) to work out how likely it is that you would see your data if the null hypothesis were true. This probability is given as a p or sig value and will be a number between 0 and 1. The closer to zero the p value is the less likely it is that you would have seen the data in your sample if there were no difference or association.

For instance, if the p value is 0.01 this means there is just a 1% chance that you would have found the data in your sample if there were no difference or association.

A p value of 0.6 would mean that is very likely that you would have seen the difference or association in your data even if this was not present in the population you are interested in.

The p value does not tell you the probability that the null hypothesis is true.

When you run a statistical test you need to decide (and state) *in advance* what cut-off you will use to decide whether your result is significant. This value is often called alpha.
Often alpha is chosen to be 0.05 (so if $p<0.05$ we say there is a significant difference or association) but it is important to realise that this is an arbitrary choice.

The choice for alpha might depend on the consequences of that choice. For instance, if a health service is trying to decide whether a new, more expensive drug will improve patient outcomes they may choose a cut-off of 0.01. They will want to have stronger evidence that the drug is effective before investing in it.

One way to think about p values is in terms of a court case. The starting position (the null hypothesis) is that the defendant is innocent. If there is enough evidence (a low enough p value), the case is proved beyond reasonable doubt and the defendant is found guilty (the null hypothesis is rejected).

Note that we are always dealing with chance. There is a chance that we have rejected the null hypothesis (found the defendant guilty) when it is in fact true (the defendant is innocent). Also retaining the null hypothesis does not mean than we think that it is true (the defendant is innocent), just that there is not enough evidence to reject it (find the defendant guilty).

This leaflet gives more information:
https://www.statstutor.ac.uk/resources/uploaded/statisticalhypothesistesting3.pdf

This video provides an informal introduction to p values:
https://www.youtube.com/watch?v=eyknGvncKLw