PROBABILITY AND SIGNIFICANCE, INCLUDING THE INTERPRETATION OF SIGNIFICANCE AND TYPE 1/TYPE 2 ERRORS

Probability and significance

Probability refers to how likely something is to happen, and varies from 0 (something definitely **won't** happen) to 1 (something definitely **will** happen). Science deals with probabilities, and makes judgements about how confident we can be that the results we have obtained in an investigation can be considered to be 'significant' rather than the result of chance factors operating.

Suppose that we conduct an experiment in which there are two conditions. In one condition, participants hear a list of words and then have to immediately recall them. In the other condition, participants hear the same list of words, but have to count backwards in threes for ten seconds. The results are as follows:

Immediate recall: Mean number of words recalled = 3.9
Delayed recall: Mean number of words recalled = 2.8

What we want to know is whether delaying recall by 10 seconds has had a **real effect** on how much has been recalled, or whether the difference in the amount recalled in the two conditions is just the result of **chance factors operating**. When we devise an investigation, we always predict that something *will* happen.

The hypothesis that the researcher proposes is called the **experimental hypothesis**. The experimental hypothesis *always* predicts that whatever happens will be a **real** effect. In the case of an experiment, the researcher's hypothesis is that the difference between two conditions will be a result of a change in the independent variable.

However, there is another kind of hypothesis, which is called the **null hypothesis**. This hypothesis predicts that nothing will happen in our investigation, and that anything which does happen is simply the result of **chance factors operating**. So, if the experimental hypothesis is that 'X *will* affect Y' or 'X *will* be correlated with Y', the null hypothesis is that 'X *will **not*** affect Y' or 'X *will **not*** be correlated with Y'.

We use **statistical tests** in psychology to them us howconfident we can be that a difference between conditions in an experiment *is* real, or that a correlation between two variables *is* real.

If the outcome of a statistical test tells us that we can be **at least 95% confident** that a difference or correlation is real, then that is good enough, and we conclude that the difference or correlation is a **statistically significant** one. When this happens, we **accept the experimental hypothesis** and **reject the null hypothesis**.

If we can be at least **95% confident**, then the difference or correlation is said to be **significant at the 5% level**.

On some occasions, the outcome of the statistical test tells us that we can be even more than 95% confident:

99% confident = significant at the **1% level**
99.9% confident = significant at the **0.1% level**
99.99% confident = significant at the **0.001% level**

When we can be 95% confident, it is convention to write it like this: $p < 0.05$. What this means is that the probability (**p**) of finding a difference or correlation as large as this if chance factors were operating is less than (<) 5% (**0.05**). 99% confidence is written as $p<0.01$, and 99.9% confidence is written as $p<0.001$.

However, if the outcome of the statistical test tells us that we can't be at least 95% confident, then we have to conclude at the difference or correlation is not big enough to be treated as real, and it is **statistically non-significant**. When this happens, we have to **accept the null hypothesis** and **reject the experimental hypothesis**.

Type 1 and Type 2 errors

If the outcome of a statistical test tells us that the difference or correlation **is** significant, and it *really is* significant, then we would correctly accept the experimental hypothesis and reject the null hypothesis. Equally, if the outcome of the statistical test tells us that the difference or correlation **is not** significant, and it *really is not* significant, then we would correctly reject the experimental hypothesis and accept the null hypothesis.

However, there may be occasions on which the outcome of the statistical test we have done tells us that the difference or correlation **is not** significant when in fact it actually **is**. This would lead us to accept the null hypothesis when in fact it was actually false. In other words, we would conclude that there wasn't a difference or correlation when in fact there was. This is called a **Type 2 error**.

Type 2 errors occur for a very simple reason: Some statistical tests are really simple to use, because they only require us to do basic things with the data we have collected. A good example is the **Sign test**. Other tests, such as the **Wilcoxon test,** require us to do more complicated things with the data we have collected. The problem with simple tests is that because they don't involve us doing much with the data they can sometimes fail to detect a difference when there actually is a difference. A **Type 2 error**, therefore, occurs when we use a '**weak**' test (like the **Sign test**) when we *could* have used a '**stronger**' test like the **Wilcoxon test**.

There may also be occasions on which the outcome of the statistical test we have done tells us that the difference or correlation **is** significant when in fact it actually **is not**. This would lead us to reject the null hypothesis when in fact it was actually true. In other words, we would conclude that there was a difference or correlation when in fact there wasn't. This is called a **Type 1 error**.

This type of error is not the fault of the statistical test that has been used. Instead, it is the researcher's fault. Remember that it is 'conventional' to use the **5% significance level**, that is, to conclude that a difference (or correlation) is real if we can be 95% confident that it is real. However, there is no law which says the 5% significance level has to be used.

What some researchers do is to say that they are prepared to accept that a difference or correlation is real if they can be 90% confident that it is real (the **10% significance level**). If a test tells them that they can be 90% confident, but not 95% confident, then they are wrongly accepting the experimental hypothesis (because the rest of us are using the 95% confidence level). A **Type 1 error** is therefore most likely when a researcher doesn't use the conventional 95% confidence level.