statsART

statistics | analysis | research | training

Interpreting tests of statistical significance

This guide is intended to help you to interpret the findings of analyses statistical significance.

From samples to populations

In any study, we can only collect data from a small sample of the entire population. For example, if we wanted to look at sex differences in height, we would only be able to measure the height of a small number of men and women – measuring every single person on the planet (the population) would be impossible! We use tests of statistical significance to see whether our finding from our sample can be generalised to the entire population.

Do our findings in the *sample* exist in the entire *population*?
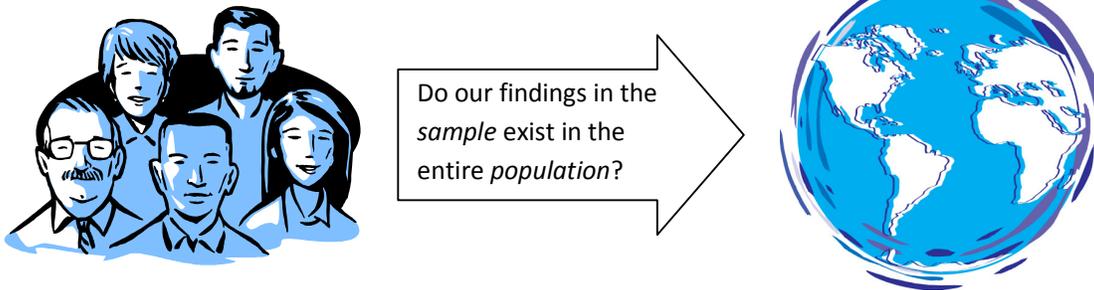
Avoiding making mistakes

When we analyse data from a sample, we want to know whether we would find the same thing if we tested the entire population. This means that there are four possible outcomes from any analysis.

		Population	
		Does the effect exist in the real world?	
		Yes	No
Sample	Yes	Correct analysis	Type I error
Did you find an effect in the study?	No	Type II error	Correct analysis

Two of the outcomes give us the correct result; either finding an effect that really does exist or not finding an effect that does not exist. We are far more worried about possible mistakes that we might make, in particular making a *Type I Error*. This type of error would mean that we found something within our sample that does not exist in the real world, the population.

Think about this within the context of medical research on a treatment. Imagine we found within our sample that the treatment was effective, but that was just a random result that does not exist in the population. An ineffective treatment might be further tested or used on patients when it does nothing at all. This could have medical consequences for the patients and financial consequences for the people conducting the research. Type I errors are very bad.

What does significance tell us?

When looking at the significance of any analysis, we are looking at the probability that we have made a type I error. The bigger the effect in our sample the less likely it is that we have made a type I error. So how willing are we to make this kind of mistake?

The usual criteria, the alpha level, is set at $p = .050$. This means that the probability of us making a type I error is 5%. To say that we have a significant result that can be extrapolated to the population; we want less than a 5% chance of making a mistake.

$p \leq .050$	$p > .050$
Less than 5% chance of Type I error	Greater than 5% chance of Type I error
Significant result	Not significant result

The smaller the p value resulting from a statistical test, the less likely it is that you made an error.

Things to think about when interpreting a statistically significant result

1. Is an alpha level of .050 suitable for your analysis? Sometimes you may want a stricter level, for example an alpha level of .010 for medical research – you want less than a 1% chance of making a Type I error. Alternatively, if you are conducting some early exploratory research, you might be willing to accept a higher alpha level, say .100 (10% chance of a type I error).
2. The significance of a result is influenced by the sample size. The bigger the sample the easier it is to extrapolate your findings. This means that an effect found in a study that included 1,000 participants will be far more significant than exactly the same effect found in a sample 10.
3. Most tests of statistical significance look for consistency in the effects across all participants. This means that a small effect that occurs in the same way for every person in the sample is likely to be significant. An inconsistent result, even if it produces some big overall effects, is far less likely to be statistically significant.

What next?

If you have any further questions about statistics, inferential statistics and significance, please visit our website (www.statsART.com) or email us (info@statsART.com).