schottler consulting
Insight from complexity

What is statistical significance?

Significance testing involves using a range of statistical methods to identify what are termed "statistically significant" differences or trends in data. A significant difference is a difference which is assumed to be real, rather than due to (what is termed) "sampling error". Sampling error simply refers to the fact that when you select people to do surveys, it is often the case that you may not have selected or "sampled" in a way, which represents the real population of people you are looking at. For instance, while a result of 80% for group 1 and 85% for group 2 may look like a real difference, it's quite possible that this 5% difference is just due to sampling error.

To assess whether differences in data are likely to be real, statistical significance testing can be quite useful. By applying a range of statistical methods, we can test whether that 5% difference is statistically significant (or just due to sampling error). This produces a probability figure represented as "p". A probability less than 0.05 is typically used as an indication of statistical significance. This indicates that the theoretical chance of the result occurring by chance is less than 5%.

As such, a result of $p<.001$ (implying that the probability figure was so small and the result is very unlikely as a result of chance) means that the difference is very statistically significant. Conversely, figures greater than a probability or "p" value of 0.05 are generally considered to be non-significant. Note that < implies "less than".

Some examples of results to assist with the interpretation of statistical significance are below:

- t-tests - Group 1 Mean=5.3, Group 2 Mean=5.5 ($p<.05$) – implies that Group 2 was rated significantly higher than Group 1 (ie. The 0.2 difference in mean ratings was "statistically significant")

- Pearson correlations – Between age and attitude - $r=.86$ ($p<.05$) – implies that there was a statistically significant relationship between age and attitude (note that the .86 implies that, as age increased, attitudinal ratings became significantly higher – in contrast, -.86 would have implied the reverse – ie. As age increased, attitudinal ratings decreased)

- chi square - $p>.07$ – this implied that the probability of obtaining the chi square test result was more than 0.05 (0.07) – 0.07 is often described as "tending towards significance" in cases where researchers may suspect a trend, but couldn't prove the trend for sure

Note that there are many statistical tests which can be used to identify trends in data and virtually all can be used to assess whether a result is "statistically significant". Even if you read papers and don't always understand the tests, you can be sure that a result of $p<.05$ is one that's worth looking at.