

Handout: A beginner's guide to understanding statistical terminology in research articles

The purpose of this 'beginner's guide' is to enable people who are not familiar with statistical concepts to have the confidence to read counselling and psychotherapy research papers that make use of quantitative analysis. This guide attempts to explain some basic concepts. Information about more advanced concepts can be explored by following-up the books and internet resources listed at the end.

When thinking about the use of numbers and statistics in research, it is helpful to keep these ideas in mind:

1. Numbers and statistical tests are a really good way of summarising information that has been collected from groups of people, and looking for patterns in that information. Essentially, that is what it is all about. A typical scenario is that a sample of people complete a questionnaire. By using stats, it is possible to look at questions such as:
 - a. What is the average score?
 - b. What is the range of scores (ie highest to lowest)?
 - c. How spread out are the scores (ie are the scores spread out across the whole scale, or have all the participants obtained fairly similar scores?)
 - d. Do different groups (eg men vs women; CBT clients vs psychodynamic clients) have different scores?
 - e. If you are measuring two different things (eg therapeutic alliance and amount of client change after therapy), to what extent do the scores on them go up and down together (correlate)?
2. Quite often, researchers make use of advanced and sophisticated forms of statistical analysis. This is because they are working in universities and have expert statisticians to help them. It is not unusual for the researcher him/herself to not fully understand what the stats are about. Much of the time, the majority of readers are confused too – even people who have had some stats training. As you might imagine, this is a somewhat risky situation.
3. In any well-written paper, the author should provide a narrative summary of the findings, alongside the stats. It should always be possible to 'skip' the stats, and pick up the thread of the study and its results by reading the non-statistical sections of the report.
4. In an article published in a reputable journal, it is unlikely (but not impossible) that the stats have been calculated wrongly. This is because the figures are checked carefully before the paper is published. However, authors of papers may sometimes neglect to discuss statistical findings that do not fit their theory.
5. Statistics is a language, which includes some new ideas and words. It takes a bit of time to get to grips with it. As with any other language-learning, regular practice is important. Also – there is always more to be learned.

Statistical methods used in therapy research can be divided into **descriptive** and **inferential** statistics. Descriptive statistics provide a description of the people in the research sample, in terms of the information collected on these people by the researcher. Inferential statistics, by contrast, enable the researcher to test hypotheses, and assess the confidence with which statements made about the sample being investigated will also be likely to hold true for the population as a whole.

Descriptive statistics

In almost all research that uses measurement, the average, or *mean*, score is reported. This is simply calculated by adding up everyone's score and dividing it by the number of people. The mean score for a sample is usually symbolised as 'm', or occasionally as an X with a bar over the top of it.

It is also important to know about the *variability* of scores – the amount of spread in the distribution of scores. The *standard deviation* statistic (abbreviated as s, s.d., sd or SD) is usually given in research reports as a measure of variability. In fact, the SD figure is the average variability of scores. That's where the term 'standard deviation' comes from: it is a kind of average or typical (hence 'standard') difference (hence 'deviation') between each person's score and the mean score. It tells us how far off the average person is from the average, which also tells us how far we should trust the average as a way of representing the people in the sample: The bigger the SD, the farther people tend to be from average. Interesting facts about SDs:

- About two-thirds of people in a sample will have scores that are within one SD above or below the mean.
- About 95 per cent of people will have scores that are within 2 SDs above or below the mean.
- On a standard five-point rating scale, such as that used by the CORE Outcome Measure, many items will have SDs of about one scale point, so it's a good idea to pay attention to items with SDs substantially above (1.5) or below (.5).

The other bit of jargon that is found is 'n' or 'N' – the number of people in a sample.

Inferential statistics

The purpose of inferential statistics is to enable the researcher to **test hypotheses**. For example, a researcher may hypothesise that clients will report higher levels of satisfaction in sessions where the counsellor displays a higher proportion of accurate empathy responses. Or the hypothesis may be that group therapy is more effective than individual counselling for students with loneliness problems.

The way that inferential statistics operate is not to test a research hypothesis directly, but to calculate the confidence with which the so-called "Null Hypothesis" (most commonly that there is no difference) can be rejected. The Null Hypothesis is a statement that the sets of scores that have been obtained for different groups are **equal** (eg that the high- and low-empathy groups are no different in satisfaction). If the scores are unequal, then this is evidence to support the research hypothesis.

The Null Hypothesis is a precise and exact statement: there is *no* difference between the groups. In reality, there will always be *some* difference in the scores on a test or measure that two sets of people have completed. What needs to be known is whether this difference that has been found is due to random or chance factors or whether it is such a substantial difference that it would be unlikely to occur merely through random fluctuations in the sample of people included in each group. The statistical formulas used in inferential statistics essentially give the researcher a way to ascertain the level of *probability* (p) that a particular result could have been obtained by chance.

The concept of probability refers to the likelihood of occurrence of an event. A high probability would mean that the event was almost certain to happen. A low probability indicates that the event is very unlikely indeed. The convention in mathematics is to report probability on a scale of 0 to one, with a probability of 0 being equivalent to 'never' and a probability of one meaning that the event was certain. A probability of 0.5 would reflect a 50 per cent, or one time in two chance of the event occurring. A probability of 0.01 would suggest a 1 per cent or one time in a hundred chance.

In the type of research being considered here, the question that is being asked is: what is the probability that the differences between the two groups are due to chance? If the statistical calculations carried out to answer this question suggest that there is a high probability that the differences are indeed due merely to chance, then the Null Hypothesis is supported (or, strictly speaking, cannot be rejected) and the research hypothesis is refuted. On the other hand, if the differences are large, the calculation will tell the researcher that there is a low probability that this situation is a chance occurrence. In other words, the lower the probability estimate obtained, the more confidence there can be that the Null Hypothesis can be rejected and the research hypothesis supported.

But, for a probability estimate, how low is low enough? Again, this is another situation where there exist generally accepted conventions. Usually, a probability of less than 0.05 (or one time in 20), would be considered as the threshold for statistical acceptability. If a result is likely to happen more than one time in 20 simply because of random error or chance factors, it is not a very robust or meaningful finding. In research studies, therefore, statistical significance values are usually reported in a form such as ' $p < 0.05$ ', which would indicate a probability of less than one in 20 that a result has been obtained by chance. Other probability threshold levels commonly employed are 0.01 (one in 100) and 0.001 (one in 1000). Some researchers prefer to report the exact probability value obtained, rather than a 'less than' figure.

A number of statistical techniques can be used to decide whether to reject the Null Hypothesis (the reasons why different techniques are used in different situations is too complex to explain here). The main hypothesis-testing techniques used in therapy research are:

- chi-square (X squared)
- Mann Whitney U test
- T-test
- Analysis of Variance (ANOVA) and Multiple Analysis of Variance (MANOVA).

The results of t-tests and ANOVA/MANOVA calculations are sometime reported as F ratios (or as $F = \dots$) or as $t = \dots$. This is sometimes also accompanied by a figure for 'df' (degrees of freedom). Just ignore these – they are only of interest to readers who want to know about the statistical detail.

Another statistical test that is widely used in counselling research is the *correlation coefficient* (reported as ' r '). This is a statistic that shows the strength of relationship between two variables, that is, the extent to which they go up and down together. The value of a correlation coefficient (r) can range from +1.00 to -1.00. A positive (plus sign) correlation means that there is a positive linear relationship between two variables (ie higher scores on one variable are associated with higher scores on the other). A negative correlation means that high scores on one variable are associated with low scores on the other: a negative linear relationship. The nearer r is to plus or minus 1.00, the stronger the relationship is between the two variables. (Note: a linear relation is one that you could make a graph of and represent by drawing a straight line through it, which happens when one variable goes up to the same degree or at the same rate as the other variable.)

This discussion of some of the statistical techniques employed in counselling and psychotherapy research is intended to do no more than supply an initial basis for understanding the role of statistics in studies of therapy. There are many more statistical tests than could possibly be covered here – references to further reading are provided below.

Advice: the main things to look for are the mean (m) and standard deviation (SD) scores that are reported, and whether any analyses that have been carried out have produced statistically significant results ($p < .05$).

Further reading

There are more extended introductory stats sections in:

Barker, C., Pistrang, N. and Elliott, R. (2002) *Research methods in clinical and counselling psychology*. 2nd edn. Chichester: Wiley.

Cooper, M. (2008) *Essential research findings in counselling and psychotherapy: the facts are friendly*. London: Sage.

McLeod, J. (2003) *Doing counselling research*. 2nd edn. London: Sage.

More detailed explanations of statistical concepts and procedures can be found in a wide range of stats textbooks that are available. For example:

Argyrous, G. (2000) *Statistics for social and health research*. London: Sage.

Coolidge, F.L. (2000) *Statistics: a gentle introduction*. Thousand Oaks, CA: Sage.

Field, A. (2005). *Discovering statistics using SPSS*. 2nd edn. London: Sage.

Greene, J. and d'Oliveira, M. (1999) *Learning to use statistical tests in psychology*. 2nd edn. Buckingham: Open University Press.

Hinton, P. (1995) *Statistics explained: a guide for social science students*. London: Routledge.

Kerr, A.W., Hall, H.K. and Kozub, S.A. (2002) *Doing statistics with SPSS*. London: Sage.

Salkind, N.J. (2000) *Statistics for people who (think they) hate statistics*. Thousand Oaks, CA: Sage.

Wright, D.B. (2002) *First steps in statistics*. Thousand Oaks, CA: Sage.

If you are reading an article, and come across statistic jargon that is confusing, there are several websites that are helpful (in addition to a simple Google search):

Statistics textbook on-line:

<http://onlinestatbook.com/rvls.html>

Glossary of statistical terms.

<http://www.stats.gla.ac.uk/steps/glossary/index.html>

List of statistics glossaries

<http://www.glossarist.com/glossaries/science/physical-sciences/statistics.asp>

American Psychological Association: glossary of psychological terms

<http://www.psychologymatters.org/glossary.html>