

Inferential Statistics: A Mini-Digest

Prof. John deRoche

This quick lesson is stripped of all the ifs, ands, buts, and maybes.

Inferential stats for measures of association.

- The purpose of inferential stats tools is to help us judge whether the findings from our sample can be safely generalized to the population that the sample was randomly chosen from.

For instance, the “BigOneThird” data set contains 3934 respondents ‘randomly’ sampled from the whole adult population of Canada (part of the 1991 General Social Survey). Two of the variables in that study were “happiness” and “affect balance”. The gamma association for those two variables, among these individuals, is .58.

But *would you have found a correlation if you had data for the entire Canadian adult population? Or would you instead have found no association – that is, a correlation of zero?*

This is a meaningful question because of sampling error. When we take a random sample, there is no guarantee that it will perfectly represent (be a perfect replica of) the population, because it’s based on “the laws of chance” (probabilities). In other words, *a correlation in the sample might be just the result of chance* – just a fluke because we had the bad luck to get an unrepresentative sample.

- Here is the way statisticians typically frame that same question. Suppose we had devised a *research hypothesis* saying that happiness and affect balance are directly associated. (“As happiness rises, there is a rise in positive affect.”) The corresponding ***null hypothesis*** says: *The correlation of affect balance with happiness is zero. There is no correlation.*
- Inferential stats tools tell us the probability (chance, likelihood) that the null hypothesis is true for the population, in light of what was found in the random sample. So, in our example, the random sample revealed a gamma of .58, a substantial direct association. What are the chances that the population, nevertheless, has a gamma of ZERO?
- SPSS answers that last question – without your even needing to ask it. This stats program calculates the probability that gamma is 0 in the population. That probability is displayed in the SPSS output in the last column of the gamma table. SPSS labels the column as “***Approx. Sig.***” (approximate significance level).
- Among quantitative researchers, “Approx. Sig.” is most commonly abbreviated as “**p**” – which stands for “***the probability that the null hypothesis is true.***”
Typically, we are attempting to “reject the null hypothesis” – which would mean that our research hypothesis is supported. When working with fairly small samples (where the chances of poor representativeness are greater), we usually set a standard of **p < .05**. This means: “The chance of ‘no association’ has to be smaller than 5 in 100, or 5%, before we can safely reject the null hypothesis. If it is bigger than .05, we have to assume there is really no association within the population.” With larger samples (like BigOneThird), it is common to set a tighter standard, such as **p < .01**.
- In terms of the bottom-line application of all this to your own findings, try this perspective: If you are expecting a correlation between two variables, the good news would be (a) a big correlation value and (b) a small “p” or “Sig” value.
- One more point about the language we use around this. If we successfully reject the null hypothesis (which requires a p or “sig” value under, say, .05), we describe the correlation as “***statistically***

significant". If p is too large, however, and we fail to reject the null hypothesis, we say the correlation is "*statistically non-significant*" or that it "failed the test of statistical significance". The whole process is often described as "doing a test of statistical significance." Thus, the "inferential stats" toolkit can be seen as a set of "tests of statistical significance."

By the way, SPSS shows p values (Sig) to 3 decimal places. A figure of ".000" indicates a value smaller than .0005, rather than actually being zero.

Inferential stats when comparing two or more means (t-test and ANOVA).

The following description is about **ANOVA – analysis of variance** – which we use when the cases are divided into *three or more* groups according to their classification on the independent variable. When there are *only two* groups being compared, we use the **t test** instead (as in Lab 6 for the AN/S-POLS 268 course). The basic logic of these tools is the same, so you can apply the following when you're using a t test as well.

For our example, we are treating "affect balance" as the dependent variable and "happiness" as the independent. We will compare the average "affect balance" scores across the four "happiness" groups, to test our hypothesis that affect tends to become more positive as we go from unhappy to happy groups of people. In fact, the mean "affect" does change in the expected direction. Affect is operationalized on a scale where 0 is extreme negative affect and 20 is extreme positive affect. As we go from the most unhappy of the four groups to the happiest, the mean affect scores are 7.0, 8.6, 11.8, and 14.4. We want to see whether overall *differences among* these averages are *big enough to be "statistically significant"*.

ANOVA – analysis of variance – is an inferential stats tool to accompany this method of checking an association between variables. We are comparing the means of the dependent variable within the subgroups of the independent variable, to see whether the DV is affected by the IV. ANOVA looks at the size differences of the DV means, among all the IV subgroups, and produces a **p** value (or "Sig") that expresses the probability that the differences overall are too small to be statistically significant. *In other words, it tests the null hypothesis that all these means are equal to each other in the population.*

For our set of four "affect balance" means, SPSS yields a **p** value of ".000". From what I have told you, you can now answer this string of questions. Try them.

- What is the probability that the four means are all equal in the Canadian adult population?
- What is the likelihood that the null hypothesis is true?
- Would you reject the null hypothesis, or fail to reject it?
- Are the differences in affect balance scores (across the happiness groups) statistically significant?