ACET inc

Evaluation Take Aways

Free Evaluation Technical Assistance from the Staff of ACET, Inc.

Statistical Significance

Statistical significance refers to whether an observation (e.g., a difference in test scores between two different groups) represents a pattern rather than the results of random chance. Let's look at a hypothetical study comparing 10 students, five of whom are classified as English Language Learners (ELL) and five for which English is their primary language. These 10 students are all given the same reading test and we would like to know whether or not one's ELL status has an effect on reading scores. The simplest way to assess this would be to compare the mean (average) score for each group. Maybe our results look something like this:

	English Language Learners	English
Average Score	83	100

We observe a mean score of 83 for ELL students and 100 for English-speaking students, and we might conclude that one's ELL status does have an influence on test scores because we observe a difference of 17 points between the two groups. But how do we know the difference in test scores isn't due to random events? Here is where we can use a significance test. Examining just the means can mask differences in individual scores, but a significance test will compare individual scores and allow you to dig deeper into the results.

First, let's look at an example that shows results that are *not* statistically significant. As can be seen in the table below, four of the ELL students have scores that are identical to scores for the English-speaking students and one ELL student who scored substantially lower than any other student in either group. That one student's score had a substantial impact on the mean for the entire group and created the 17-point difference between the ELL and English-speaking students' mean scores. In this example, one single student's score does not represent a pattern and a significance test would indicate that the observed difference in group means is not statistically significant.

ELL Students' Scores	English Speaking Students' Scores
5	90
95	95
100	100
105	105
110	110
Mean = 83	*Mean = 100*

Below is an example a statistically significant result. In this example notice that all ELL students have scores that are lower than the scores of the English speaking students. In addition, the highest score observed among the ELL students is 85 while the lowest score for an English-speaking student is 90. In this example there is a clear pattern of ELL students scoring lower than the English-speaking students. It is this non-random pattern of scores that results in a statistically significant finding. In this example we would use an independent samples *t*-test to test for statistical significance.

ELL Students' Scores	English Speaking Students' Scores
80	90
80	95
85	100
85	105
85	110
Mean = 83	*Mean = 100*

Notice that, in all three tables, the means for the two groups are identical; in the first table the differences between the group means could be due to chance (but we don't know for sure), in the second table the differences in means are *not* statistically significant, and in the third table the differences between the means *are* statistically

significant. Why? It goes back to the definition of statistical significance. Tests of statistical significance are looking for consistent, non-random patterns of scores. In the first two tables, there were no consistent patterns to establish that the two groups are different. But, in the third table, a consistent pattern clearly emerges that discriminates between the two groups.

Typically statistical significance is expressed with a number, usually expressed as "p <=.xx," where xx would be replaced by a two-digit number. That two-digit number represents the probability or likelihood that the observed differences were due to chance. In general, the smaller that two-digit number the less likely that the observed differences are due to chance. In evaluation and most research the 'gold standard' of statistical significance is 5%, or p <= .05. This would indicate that there is a 5% chance (or less) that the observed differences were due to chance or random events.

It is important to note that statistical significance only tests whether or not observed differences are due to chance; tests of statistical significance do *not* indicate whether or not a difference is important or meaningful. Decisions about the importance or meaning of results are up to interpretation by program staff, evaluators, or researchers. It is possible to find differences between groups are statistically significant but have little importance and it is possible that differences between groups are not statistically significant but still have great importance or meaning.

Resources

Fleming, S. (2011, May 26). Steve Fleming on explaining statistical significance [Web log post]. Retrieved from http:// http://aea365.org/blog/?p=3528

Gonick, L. & Smith, W. (1993). *The cartoon guide to statistics*. HarperPerennial: New York.

Salkind, N. J. (2010). *Statistics for people who (think they) hate statistics* (4th ed.). Thousand Oaks, CA: SAGE.

For additional information about this or other ACET, Inc. resources, or for evaluation assistance, please contact:

ACET, Inc.
info@acetinc.com
952.922.1811

Suggested Citation:
ACET, Inc. (2013). *Statistical significance.* Minneapolis, MN.