StarTips
...a resource for survey researchers

Share this article:

What Every Researcher Should Know About Statistical Significance

★ **What is statistical significance?**

Survey researchers use significance testing as an aid in expressing the reliability of survey results. We use phrases such as "significantly different," "margin of error," and "confidence levels" to help describe and make comparisons when analyzing data. The purpose of this article is to foster a better understanding of the underlying principles behind the statistics.

Data tables (crosstabs) will often include tests of significance that your tab supplier has provided. When comparing subgroups of our sample, some results show as being "significantly different." The key question of whether a difference is "significant" can be restated as "is the difference enough to allow for normal sampling error?" So what is this concept of sampling error, and why is there sampling error in my survey results? A sample of observations, such as responses to a questionnaire, will generally not reflect exactly the same population from which it is drawn. As long as variation exists among individuals, sample results will depend on the particular mix of individuals chosen. "Sampling error" (or conversely, "sample precision") refers to the amount of variation likely to exist between a sample result and the actual population.

A stated "confidence level" qualifies a statistical statement by expressing the probability that the observed result cannot be explained by sampling error alone. To say that the observed result is significant at the 95% confidence level means that there is a 95% chance that the difference is real and not just a quirk of the sampling. If we repeated the study 100 times, 95 of the samples drawn would yield similar results.

"Sampling error" and "confidence levels" work hand-in-hand. A larger difference may be significant at a lower confidence level. For example, we might have the ability to state that we are 95% confident that a sample result falls within a certain range of the true population level. However, we can also be 90% confident that our sample result falls within some broader range of the population level.

So when we state that a comparison is significant, we are trying to extend our limited survey results to the larger population.

★ **How large a sample is enough?**

Normally, we deal with this question prior to data collection at the sample design stage of a research project. However, the same statistical analysis can be performed to determine the "margin of error" of a research study. As sample sizes increase, survey results generally prove more reliable; hence, the margin of error becomes smaller. We often see a disclaimer on a research study such as, "results are reliable to within +/- 6 percent at the 95% confidence level."

To determine appropriate sample size, we must consider the maximum sample error we are willing to accept, as well as the confidence level desired. Different research requires different degrees of reliability, depending on the specific objectives and possible consequences of the survey findings.

Often, an "acceptable" margin of error used by survey researchers falls between 4% and 8% at the 95% confidence level. We can calculate the margin of error at different sample sizes to determine what sample size will yield results reliable at the desired level. Another factor in determining sample size is the number of subgroups to be analyzed; a researcher will want to be sure that the smallest subgroup will be large enough to ensure reliable results.

The sample precision analysis can help a researcher make an informed decision regarding sample reliability.

* **What statistical tests do I need to know about?**

  The most common statistics used to calculate significance of survey results are the Z and T statistics, used for proportions and means respectively.

  Z-test of proportions: used to test the difference between proportions (percentages) for two groups. Statistics computed provide the probability that a difference at least as large as noted would have occurred by chance if the two population proportions were in fact equal.

  T-Test of independent means: used to test the difference between means for two groups. Statistics computed provide the probability that a difference at least as large as noted would have occurred by chance if the two population means were in fact equal.

  Sample Precision or margin of error analyses use the T-test to determine a conservative estimate of sampling error for a given sample size. These estimates are helpful in answering questions like: "How large a sample do I need?" or "How reliable are my survey results?"

  Most common tests will yield either a "one-tailed" or "two-tailed" result. A one-tailed test is appropriate when the hypothesis tested implies a directional difference (i.e., "Group A will score higher than Group B"). A two-tailed test, on the other hand, tests the hypothesis that the two groups are "different", regardless of the direction of the difference (i.e., "Group A and Group B perform differently" on a certain measure). Significance test results are most commonly displayed based on 2-tailed probabilities.

* **Calculating Tests of Significance**

  DataStar provides a complimentary tool, StarStat (and an iPhone app) which calculates z-tests, t-tests, and sample precision estimates.

  Having a basic understanding of these tests and the principles involved will help the researcher to interpret significance testing results whether displayed in data tables or calculated using a standalone tool.

Contact DataStar today for more information or a free, no obligation quote on your next survey need.