DataStar

StarTips

...a resource for survey researchers

Share this article:

Survey Tabulation Basics: Statistics (January 2012)

This article provides a summary of the basic descriptive statistics typically shown in crosstabs and other survey data analyses. Included are the basic measures with which all researchers should be comfortable.

Descriptive Statistics are those measures which help "describe" a distribution of survey responses. Where applicable, the mean (average), median, standard deviation and standard error are often included on tables for analysis purposes. For example, it might be helpful to show the mean of a rating scale question and other numeric fields (i.e., age or income values). These measures summarize the key results in a few succinct numbers.

A simple **mean** of a distribution is the arithmetic average - the sum of all responses divided by the number of responses. The mean of response values 1,2,2,2,3,4,5 is 2.7(19 divided by 7). A weighted mean takes this one step further by assigning weights to each response value. This is typically used when response values represent ranges such as age ranges - in this case, the mean is typically calculated based on the midpoints of the ranges (i.e., if the response value "1" is used to represent the age range 18-24, the range midpoint - "21" would be used to calculate a weighted mean which reflects age values. Weighted means may also be used to alter the coded values in a rating scale, for example, reversing a rating scale so the highest (or lowest) value indicates the value of greater importance.

The **median** of a distribution is the "middle" value when all values are listed in order from lowest to highest. In the example above, the median value is "2" (the 4th value in the 7-value list). Medians are often used where the presence of outliers (extreme responses) would skew the mean. For example, a distribution of income ranges of $18k, $24k, $35k, $42k, $46k, $65k, $72k, $125k, $4.5M would have a mean of $547k but a median value of $46k, a statistic that better describes the income level of the sample group.

Standard Deviation (SD) and **Standard Error** (SE) are perhaps the two least understood statistics shown in data tables. Both provide additional insight regarding the mean of a distribution. The standard deviation describes how far, on average, the individual values fall from the mean. A small SD would indicate that most values are clustered close to the mean value, which a large SD would describe a distribution where the values vary widely from the mean. For example, the distribution 11,11,12,12,12,13,13 has the same mean (12) of this distribution: 1, 2, 4, 6, 15, 21, 35 but very different SDs. If distribution values were represented on a frequency curve, a small SD would be indicated by a narrow, tall shape, while a large SD would be depicted by a short, wide shape The SE, on the other hand, is an indication of the reliability of the mean. A small SE is an indication that the sample mean is an accurate reflection of the actual population mean, usually based on a combination of a large sample size and low SD.

Significance Testing was covered in detail in a previous edition of StarTips (What Every Researcher Should Know About Statistical Significance). Data tables are often provided with results of statistical tests which show the reader at a glance which results are "statistically significant" when making comparisons across groups. DataStar also provides a complimentary tool, StarStat (and an iPhone app) which calculates z-tests, t-tests, and sample precision estimates.