



Share this article:

Back to Basics: Survey Tabulation (October 2011)

The purpose of this article is to round out some of the topics introduced in a previous StarTips issue, [Anatomy of a Crosstab](#), and expand on some of the terms briefly described there. Together, these articles are intended to provide a survey researcher with enough knowledge to create a tab plan and discuss data analysis needs with a supplier. There are many options available when analyzing survey results. A better understanding of these options will enable you to get the most from your data.

Tabulation of data in its simplest form involves the counting of responses - the researcher wants to know how many (or what percentage of) respondents selected each possible response choice. Generally, survey questions fall into closed-ended and open-ended categories:

- ★ **Closed-ended questions** are already set up for ease of tabulation. Response values are prelisted in the survey for the respondent to choose from, allowing for simple tabulation of the results without the need for complex data manipulation for analysis.
- ★ **Open-ended questions** are generally not in a "ready-to-tabulate" form. In order to quantitatively analyze open-ended comments, the responses must first be "coded" to allow for tabulation. Coding is usually a manual process of grouping individual responses into categories. Once coded into standardized numeric values, results can readily be tabulated in the same way as closed-ended data.
- ★ A third type, **short answer** or **numeric value questions** may require some relatively straightforward preparation before the data is ready for tabulation. Single words or phrases can be consolidated to allow for spelling and/or other minor variations. Numeric responses may be collapsed into ranges rather than discrete values to facilitate analysis.

A **tab plan** is a set of tabulation specifications, aka "tab specs," in which a research analyst would outline all of the required tables, statistics and other special requests needed for their analysis. The tab plan would act as a guide to be followed as the data is converted into meaningful results. A tab plan may describe specific tables with detailed options for analysis of each survey question, or may simply provide a general overview of the types of analyses desired.

Marginals, also known as frequencies or straight tabs show the distribution of raw counts and/or percentages of each response to the variables (questions) in your survey dataset. They are generally provided as a preliminary look at the data or to provide counts for checking subsequent analyses.

A **crosstab** is a table displaying survey results, usually for a single question. It resembles a spreadsheet with multiple rows and columns. Most commonly, the rows are the answer categories for a given question and the columns are the key variables for analysis. Survey data can be analyzed in a variety of ways, but basic crosstabs remain a powerful tool in exploring your data. Crosstabs are sometimes referred to as stub and banner tables, banner books, or simply, "tables." What follows below is a guide to the various components of a crosstab.

- ★ The **stub** contains the rows of a table - typically a list of response choices to a given question, plus any summary rows, nets or additional statistics to be shown within the table.
- ★ The **banner** is the set of table columns - the key subgroups for data analysis. The banner is shown as the header for each table, therefore the same banner is often run against each survey question. Banners usually contain 15-25 columns known as banner points. Demographic breakdowns or other classification questions are good candidates for a banner. It may be helpful to look at the marginals/frequencies to determine if the subgroups are large enough to warrant the proposed banner points. Often, individual survey questions such as overall satisfaction are also included as banner points (i.e., "high" vs "low" subgroups).

- ★ **Summary rows:** Each table typically begins with summary rows which show the total sample ("Total Respondents"), the number who have answered a given question ("Total Answering") and the count of respondents with missing data ("No Answers"). "Total Mentions" (sometimes referred to as "Sigma") is the sum total of the responses to a question and can also be included. Total Mentions may exceed 100% where multiple responses are allowed for a given question.
- ★ One of these summary rows is generally used as the **percentage base** throughout the crosstabs:
 - ★ Tables based on "**Total Respondents**" have the same base size throughout the crosstabs, with the exception of any tables that are involved in a skip pattern. For example, if the overall sample size is 500, most tables would be based on all 500 respondent's data. Frequencies and percentages will be shown for any missing values (the "No Answers" row).
 - ★ Tables based on "**Total Answering**" have a different base size for every table according to how many respondents answered each individual question. If a question had 400 out of 500 respond, the percentages would be calculated based on those 400. Typically, the number of "No Answers" will not have a percentage shown but the frequency count may be shown.
 - ★ Tables may also be percentaged on "**Total Mentions**" - typically when the researcher wants to look at results as a "share" of responses for a question in which multiple mentions were accepted for each respondent. For example, a question about restaurants visited could be percentaged on respondents, showing the incidence of visitors to a particular restaurant (i.e., of 100 respondents, what percentage visited Restaurant A), or the number of mentions (i.e., if the same 100 respondents visited 250 restaurants, what percentage of the visits were at Restaurant A).
- ★ **Nets:** Groups of similar response categories can be combined in the stub as a single row. For example, an open-ended question asking for likes and dislikes could have a net for the 'like' responses and another for the 'dislike' responses. Nets can also be used in rating scale questions to show the top-2 or bottom-2 responses. If the question has a large scale, such as 1-10, it might be helpful to show the top-3 and bottom-3 nets as well. For a question with multiple responses, the net is an unduplicated count of the items contained in the net - in other words, each respondent is counted only once no matter how many categories within the net they may fall into.
- ★ **Ranking:** Open-ended questions and non-rating scale questions are often displayed in descending order by their frequency counts, showing the answers that received the most responses at the top of the table. Ranking provides an easy way for the researcher to view the most frequently mentioned responses without sifting through a long list of items.
- ★ A table may be **based** (or filtered) on a subset of the sample. A common example of this is when skip patterns are present. Since not all respondents may be asked all questions, tables would typically exclude those who skip a question. Tables may also be based to restrict an entire analysis to various major subgroups of a sample such as region, gender, etc.
- ★ **Summary tables:** Generally, separate crosstabs are provided for each variable (item) in a dataset. Repetitive series of questions (i.e., rating scales) are sometimes shown as summary tables in addition to, or instead of, individual tables. For example, a single "top-box" table would be set up to show an entire series of 20 attributes on which a product is rated highly (i.e., "excellent" on a 5-point scale). Summary tables may also include top-2 box, bottom box or tables of means to simplify the reporting of a series of questions.

With the availability of spreadsheets and other desktop tools, survey tabulation is no longer the advanced task it was once considered. However, to get the most out of your survey data, it is critical to at least have a basic understanding of the tabulation options available. We hope this article served this purpose.

Upcoming issues of StarTips will cover additional details on other aspects of data tabulation such as descriptive statistics, weighting, and other more technical options. Feel free to let us know if there are any other areas you would like us to cover!