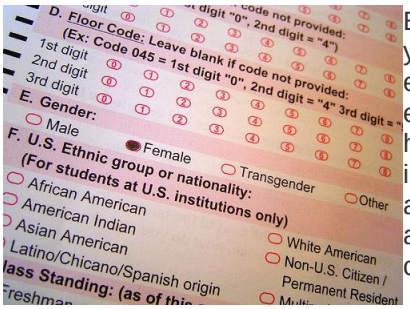
Analyze This! Doing Quantitative Analysis

You have piles of surveys on your desk, and your computer practically spills over with data. Now what? To make sense of all of those responses and figure out what they can tell you will require quantitative analysis. Quantitative refers to measurable numbers and percentages, including relationships between variables and statistical significance.

Let's look at why the numbers are important. An example of a quantitative evaluation activity often used in process evaluation is a public opinion survey. The results will be used to help convince policy makers, so the main purpose is to gather answers to the question "Are you in favor of XYZ policy?" Numbers matter in this case because the more people voice that they are in favor of a policy, the stronger the campaign becomes.



But chances are you will also ask additional questions that can inform your campaign -- questions that measure awareness of the issue or explore various options for the way the policy could be framed, for example. What's key is that you can quantify the answers and show how many people answered one way or another. If you are also interested in showing how people with different characteristics answered the questions, you can add some demographic questions and analyze results by grouping responses according to demographic characteristics.

In outcome evaluation, quantitative data can show how much change occurred because of an intervention. Is there less cigarette litter now than before the non-smoking policy? Are fewer participants smoking than before they took a cessation class? Are more apartments designated as non-smoking than before the housing campaign? And did the difference occur across the board or only for segments of the population or the group of interest?

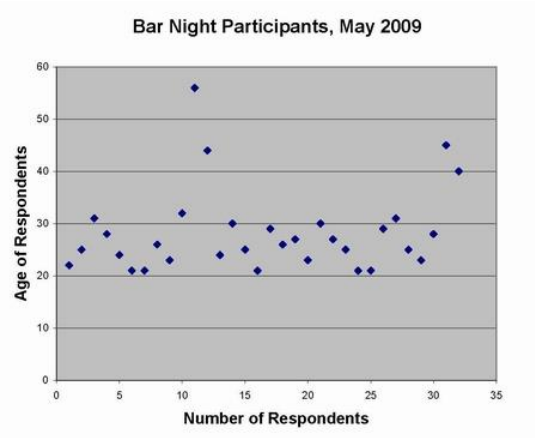
Although statistical analysis of quantitative data can get rather complicated, most Tobacco Control evaluation activities require only the most basic statistical concepts, which are percentages, means, and median. These calculations can be fairly easily arrived at, especially with the help of software like Excel or an online program like SurveyMonkey which does the analysis for you.

Percentages

Once data are entered, you can start calculating. If calculating percentages, make sure you look at how many people answered each of the questions. For instance, "30% of respondents said 'yes' when asked if they smoked." Note that the percentage is of respondents rather than the total number of people asked. If your sample size was 120 but only 100 of those approached actually answered, then the percentages are calculated using the respondents only (n=100).

Means and Medians

Let's say you did a survey of bar night participants where the average age was important to know. Using the data from the chart below, the mean (or average) age of respondents was 28 years. The mean is easily calculated by adding all the ages and dividing by the number of respondents. But looking at this scatter plot graph which shows us each of the data points, we can easily see that a few respondents were much older than the rest. This shifted the average age higher than the age of most participants. The actual midpoint, or median, of the ages of all participants is 26 years. This difference between the 26 years of the median and the 28 years average shows us



what effect a few outliers, or atypical cases, can have on the mean. For this reason, it is a good idea to report both the mean and median and/or standard deviation (a measure of the spread of data points from the mean) because a sophisticated critic might challenge results that show only means.

Relationships between Variables

If you are interested in showing relationships between variables (the things that were measured) or showing statistical significance (which can be useful if only a small percentage of difference is expected between categories or in change over time), a statistical program like SPSS is best suited for calculation.

Let's look at an example: Casino employees and patrons answered the same questions about smoking preference in the casino. When analyzing results, you want to see if employees answered a particular question differently from patrons. A statistical program lets you compare the variables and will tell you whether or not the difference is "statistically significant," meaning that it is not only accidentally different, but that the difference is big enough given the sample size that it can be considered to have something to do with the variable (e.g., the status of the person answering -- employee or patron) and not by chance.

Statistical significance is especially relevant for comparisons over time or between groups; for instance you may want to show whether or not and to what extent a difference occurred after your intervention. When the difference is very obvious, it is not necessary to figure out statistical significance. If before a smoke-free parks policy was passed 200 pieces of cigarette litter were collected compared to only 10 pieces after its passage, the difference is obvious. But what if smoking incidences among students at a school declined by 1.4 percent? Is that statistically significant or just a matter of chance? Would another round of the intervention yield the same results? A statistical program can easily tell us whether mere chance was at work or if we are actually on to something.

Be aware, though, that statistical significance only shows that there **is** a difference. It cannot tell us whether or not the difference occurred **because** of the intervention.

Using Results to Guide Program Efforts

So what if a person's status (employee or patron) predicts their smoking preference? What does this mean for the intervention? Let's say most patrons want to smoke but the employees don't. The purpose of pursuing a smoke-free casino is to protect employees and patrons from secondhand smoke, but casinos are concerned about losing patrons whose desire to smoke is not accommodated. One recommendation could be to launch an education campaign directed towards patrons who may become more supportive if they understand that a smoke-free casino policy is an attempt to protect a group of people rather than to take freedoms away from smokers.

If even the patrons remain non-supportive, the campaign could take other tacks like collecting information about the experience of casinos that have gone smoke-free, exploring the financial benefit of reduced employee sick time or cleaning costs, or touting the likely economic advantage of marketing to patrons who would be attracted to a smoke-free venue.

So whatever your quantitative data reveals -- whether support or opposition for your campaign -- it can be useful to your program in deciding what needs to be done next.

For more ideas on how to analyze quantitative data with descriptive statistics, check out Tips & Tools #10 on the TCEC website.

Photo by Robyn Lee