the pulse

A SURVEY RESEARCH PUBLICATION OF NCDOT'S PRODUCTIVITY SERVICES SECTION FEBRUARY 2007

Probability or Non-probability?

The Choice in Sampling Approach Makes a Difference

There are good surveys and there are bad surveys. Many factors contribute to the quality of a particular study, but one of the most important considerations is the type of sample used. In short, there are two classifications of samples – probability and non-probability. The decision to use one type over the other says a lot about just how accurate a survey will be.

Probability samples are considered the gold standard of sampling and are backed by a well-formed body of statistical theory. Conversely, non-probability samples, while used in some survey applications, have limitations that restrict the ability to make inferences about a broader population. In most survey endeavors, gathering data from a smaller group to derive estimates about how a larger group feels is precisely the point. So, let's take a closer look at each of these two approaches to survey sampling and see if a good rationale can be made for using one type instead of the other.

Survey *sampling* naturally implies that the entire population of interest will not be queried – only a part of it. With a non-probability sampling approach, the selection process of members of the population chosen to participate in the survey is conducted in a non-random way. In other words, the selection of the sample follows no formalized protocol for ensuring that chosen individuals are representative of the population. Membership is ultimately left to human judgement.

The implication of non-probability samples is that conclusions cannot be drawn about the study population from the results derived from the sample. The inability to determine the representativeness of the sample offers no basis for stating that the attitudes and opinions captured through the sample are reflective of the full population.

As an example of non-probability samples, consider a selection procedure known as judgement sampling. Here, interviewers are instructed to use their own discretion when selecting respondents they judge to be "typical" or "representative" of the population being surveyed. There is an obvious degree of human bias that determines who is chosen.

Probability sampling, on the other hand, relies on a randomized selection mechanism to choose the sample, such as numbers randomly generated by computer, using a random numbers table, or simply pulling numbered slips of paper from a hat. This approach eliminates the element of human involvement in the specific selection of one person over another. With probability sampling, each member of the study population has a quantifiable and non-zero chance of being selected to participate in the survey. Measurements can subsequently be taken to determine the precision of the survey's findings.

Suppose, for instance, that a national organization wants to conduct a survey among its 25,000 members. It could survey the first 500 members who call in to the help desk seeking assistance, but the question is raised as to whether these people are "typical" members. A better approach would be to generate a unique identification number for each member and then randomly select 500 numbers that would comprise the survey sample. This probability approach gives all members an equal chance of being selected.

To further illustrate the difference between the two sampling methods, stopping visitors in a city park and asking them to take a brief survey is considered a non-probability sampling approach while surveying every 10th visitor who enters the park gate is a form of probability sampling. Calling the first 500 listings that appear in the telephone book is a type (and poor means) of non-probability sampling. Dialing telephone numbers that have been randomly generated by computer is a method of probability sampling.

Most surveys involve collecting data from just a small portion of the population of interest with the intention of generalizing the findings to the broader population. A probability sample meets the requirements necessary to do this whereas a non-probability sample does not. Probability and non-probability sampling are two distinct approaches and should not be confused as providing the same quality of survey data.

Survey Tip

When sampling from a list that has a finite population, such as a roster of employees or all subscribers to an Internet service provider for example, a *systematic circular* sampling technique of choosing every "n^{th}" name is a sound and cost efficient method of probability sampling.

The first step of this technique is to determine the number of names on the list. Then a decision is made in terms of the appropriate size of the sample. By dividing the desired sample size into the total number of names on the list, the value of "n" is determined.

Did you know...

With the Super Bowl once again upon us, it is interesting to note that professional football continues to be the most popular sport among U.S. adults. In a survey conducted by Harris Interactive among 1,219 respondents who follow sports, twenty-nine percent cite pro football as their favorite, followed by major league baseball (14%) and college football (13%). Auto racing - including NASCAR (9%) and pro basketball (7%) round out the top five.

Pro football is especially popular among respondents living in the East, those with household incomes of $35,000 to $44,999 a year, Generation X (30 to 41 age group) and African-Americans. Groups not as likely to prefer pro football are those earning less than $15,000 a year and people living in the western U.S.

Baseball is most popular among respondents who consider themselves to be liberal in their politics, while auto racing is most popular with those who have household incomes of $25,000 to $34,999 and people with a high school education or less.

Source: Harris Interactive

Comments, suggestions and questions related to survey research should be directed to Doug Cox - NCDOT Market Research Manager at (919)733-2083.