

SELECTING A SAMPLE SIZE

RON SELLERS

GREY MATTER RESEARCH & CONSULTING

Originally published in *The NonProfit Times*, September 15, 2001

How many people must provide their opinions for a survey to be statistically valid – two hundred? Two thousand?

If only the question could be answered with a simple number like that, everyone's life would be much easier. When studies are made public, inevitably there are a few comments such as, "How can they claim the opinions of 600 people represent the entire United States? That's ridiculous."

Actually, it's not ridiculous at all. Unfortunately, proving that to you mathematically would take more room than we have here and prove far more boring than the editors would allow in this publication. A quick analogy is to think of a jar with 1,000 marbles in it. Five hundred are orange and 500 are blue. To find out the *exact number* of each color, you would have to count all 1,000 marbles. In survey parlance, this is known as taking a *census*. Only with a census is there no margin of error – no possible way you'll be off by even one marble.

But few business applications need an exact count. This is where *sampling* comes in. If for your sample you pulled just one marble out of the jar, you'd have a 100% chance of being wrong in your estimate. If you pulled two marbles out, you'd have a 50/50 chance of being right, because there are four color combinations you might pull out (blue/blue, orange/orange, blue/orange, and orange/blue). Two of those numerically predict how many of each color are in the jar.

The more marbles you pull out of the jar, the greater the chances that you'll end up with a close estimate of the actual distribution of the colors. If you pulled 800 of the 1,000 marbles out of the jar, you probably wouldn't end up with exactly 400 of each color, but you would be very close to that distribution. Statistically, you have a greater chance at accuracy by counting 800 of the marbles than by counting 400 of the marbles.

Statisticians – those folks with the pocket protectors we researchers lock up in the back rooms of our companies with high-powered computer programs and lots of Yoo Hoo – have conveniently figured out exactly how great a chance of error we have in each situation. But instead of thinking of a jar of 1,000 marbles, picture your database of 75,000 donors, or maybe the entire U.S. population. And let's get back to our original question of how many people must be included in a survey for it to be valid.

The first question is how many people are in the group the survey is supposed to represent (called the *sample universe*). Statistically, this only becomes an issue if the sample universe numbers below 10,000 individuals. If you want to survey your donor base and it has only 3,000 people, then the sampling figures are different. But if the group of people is 10,000 or more, it doesn't matter whether your survey represents the city of Baton Rouge, with about half a million people, or the entire U.S. population of some 200 million adults – the number of survey responses you need is the same in each situation.

The second question is how much potential margin of error you want in the survey. Most researchers use the 95% confidence interval as their basis. This means that if the same survey were done 100 times, then 95 of those times the answers would fall within the same range. The number of people interviewed determines what that range is.

Let's take an example of a donor base of 30,000 people. This is above our 10,000 threshold, so the same figures would hold true even if the donor base were 3 million people. Assuming the sample is pulled correctly, the questions asked correctly, the response rate is acceptable, etc., then a sample of 600 people is accurate to within ± 4 percentage points. In simple English, this means if you conducted a survey to which 50% of the people answered "yes" to a particular question, 95 out of a hundred times that same question would have between 46% and 54% of the respondents answering "yes."

With this example, different sample sizes provide different levels of accuracy. For instance:

- 200 interviews: ± 6.9 percentage points
- 400 interviews: ± 4.9 percentage points
- 1,000 interviews: ± 3 percentage points
- 2,000 interviews: ± 2.1 percentage points
- 3,000 interviews: ± 1.7 percentage points

As you can see, there's a lot of difference between the level of potential error provided by 200 interviews and the level provided by 400 interviews, but not much difference between 2,000 and 3,000. The law of diminishing returns basically dictates that at a certain point, it's just not worth interviewing any more people, because it won't make a practical difference in the accuracy of your survey. This is why you rarely see national studies with 20,000 or 50,000 interviews – because they wouldn't be that much more accurate than the same study conducted among 2,000 Americans (but they would be dramatically more expensive).

Also, note that researchers usually use these round numbers primarily because they're easier for everyone to understand and deal with. Technically, our survey of 600 people is accurate to within ± 3.96 percentage points. If we wanted exactly ± 4.0 accuracy, we would only need to interview 588 people. However, since this makes no practical difference, researchers usually default to the round numbers for the sake of simplicity.

(A handy online sample size calculator can be found at www.surveysystem.com/sscalc.htm. This is provided by Creative Research Systems, which designs and markets data processing and tabulation software.)

So if you're going to conduct or commission a survey, how many people should be interviewed? Part of this depends on what potential margin of error you are willing to accept. If you want a quick, high-level look at a particular issue, 200 or 300 may be plenty. If you're going to use the data to help build a sophisticated model, you probably want a greater level of accuracy.

Part of the question, however, also depends on the importance of *subsamples*. Within your donor list, there are individual population groups such as men and women, current and lapsed donors, different age groups, various income groups, etc. Each of these subsamples may have important input to the process. For example, let's say a survey of your donors showed strong support for having your organization move in a new direction. On the surface that may sound great. But what might your decision be if you further discovered that your \$10 and \$20 donors supported the new direction, but higher-dollar donors had grave reservations?

If there are subsamples you want to look at within the total population, the overall sample size must be large enough that these subsamples are also large enough to analyze. A survey of 200 people divided evenly into two age groups means 100 respondents in each age group – small, but usable for a high-level look at opinion differences. But try to divide that sample into six key age groups, and your subsample sizes become too small to allow comparison.

All of these numbers and examples assume the sample is actually pulled randomly from the sample universe. Go back to our jar of marbles for a moment. If the 500 orange and 500 blue marbles are randomly mixed together, and you randomly pulled 100 marbles out of the jar, we could predict the accuracy of that sample. But let's say the orange marbles are all on top and the blue ones on the bottom. If you truly pull a random sample from the jar, it's still going to maintain the same level of accuracy. What you have to make sure is that you don't just pull the first 100 marbles from the top of the jar – you'll wrongly assume that the entire jar is filled with orange marbles.

This has an important application when surveying from lists (such as your donor database). Let's say you're doing a donor study by telephone, but the only telephone numbers you have on your file are from people who mail you checks. Those who give monthly from their credit card, or who give through your website, would not be represented in the study. They may be very different from the check writers. If you just sample your check writers because they're the ones you have phone numbers for, it would be akin to pulling 100 orange marbles from the top of our jar. It's imperative that the entire sample universe has an equal chance of being selected for the survey if you want the survey to be an accurate representation of your list.

Finally, your decision will also be impacted by cost. A sample of 1,000 people is more accurate and provides more ability to analyze key subsamples than does a sample of 400 people, but it's also more expensive and time-consuming. Choosing the right sample size often comes down to a balance between what you want to accomplish with the study and what you can afford. If you can only interview 200 people, that's still a study that is usable, but you must realize the limitations placed on accuracy and data use with a sample size that small.

THE NONPROFIT TIMES

“It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.”

SIR ARTHUR CONAN DOYLE (CREATOR OF SHERLOCK HOLMES)



GREY MATTER
RESEARCH & CONSULTING