How Many Surveys Do I Need?

A Simple Plan for Determining Sample Size

Prepared by
Market Directions
Analytical Group
Summer 2001

Contents

THE IMPORTANCE OF RANDOM SAMPLING

It is very important to develop a sampling plan that permits legitimate generalization from the survey results to the population of interest. The key is the use of statistically derived random sampling procedures as described below. These ensure that survey results can be defended as statistically representative of the population. Surveys that do not follow these procedures can produce results that lead to misguided market research, strategic, or policy decisions. Any so-called "survey" in which no attempt is made to randomly select respondents, such as call-in readers' or viewers' "polls", is likely to produce results that in no way reflect overall public opinion—even if many thousands of individuals participate.

FACTORS DETERMINING SAMPLE SIZE

Sample size depends on:

- how much sampling error can be tolerated;
- population size;
- how varied the population is with respect to the characteristics of interest; and
- the smallest subgroups within the sample for which data are needed.

Factors Indicating a Large Sample

- The decisions to be based on the survey have very serious consequences.
- Reliance on the data for critical decision making.
- There is likely to be a high level of variance among the participants in the population to be sampled
- The sample is to be divided in small subsamples during analysis and interpretation of results.
- Costs and timing are not an issue.
- Time and resources are readily available

Factors Indicating a Small Sample Size

- There are few major decisions or commitment based on the survey data.
- Rough estimates are only needed concerning the parameters of the population.
- The population is very homogenous, with little variance.
- The analysis and interpretation will be based on the entire sample, not subsamples.
- Budget constraints and/or time requirements limit the volume of data that can be collected.

STEPS IN SAMPLING

Identify Target Population

The target population should be identified as precise as possible and in a way that makes sense in terms of the purpose of the study. It is important to understand who is eligible and who is not. Sometimes, the entire population will be sufficiently small, and data is gathered on every member of the population. This type of study is called a census study.

Develop a List of Target Population

Develop the list frame, that is the list from which the sample will be drawn. The list should be in the form of a database in a spreadsheet program such as Excel® or a database program such as Access®. This will allow random number assignments in choosing the participants from the population.

Select the Sample

Sampling methods range from simple to complex. The sample reflects the characteristics of the population from which it is drawn. Sampling methods are classified as either probability or nonprobability.

Sample Selection Methods

In probability samples, each member of the population has an equal probability (chance) of being selected. Probability methods include random sampling, systematic sampling, and stratified sampling.

In nonprobability sampling members are selected from the population in some nonrandom manner or based on subjective judgment. Nonprobability sampling includes: convenience sampling; judgment sampling; quota sampling; and snowball sampling.

The advantage of probability sampling is that sampling error can be calculated. Sampling error is the degree to which a sample might differ from the population. When inferring to the population, results are reported plus or minus the sampling error. In nonprobability sampling, the degree to which the sample differs from the population remains unknown.

Probability Sampling Methods

Random Sampling

Random sampling is the purest form of probability sampling. Each member of the population has an equal and known chance of being selected.

Systematic Sampling

Systematic sampling is often used instead of random sampling. It is also called an Nth name selection technique. After the required sample size has been calculated, every Nth record is selected from a list of population members.

Stratified Sampling

Stratified sampling reduces sampling error. A stratum is a subset of the population that shares at least one common characteristic. Stratified sampling is often used when one or more of the stratums in the population have a low incidence relative to the other stratums.

Nonprobability Sampling

Convenience and Judgment Sampling
Convenience sampling is used in exploratory research where the researcher is interested in getting an inexpensive approximation of the truth. The sample participants are selected because they are convenient. Judgment sampling is an extension of convenience sampling.

Quota Sampling
Quota sampling is the non-probability equivalent of stratified sampling. Like stratified sampling, the researcher first identifies the stratums and their proportions as they are represented in the population. Then convenience or judgment sampling is used to select the required number of subjects from each stratum.

Snowball Sampling
Snowball sampling is a special non-probability method used when the desired sample characteristic is rare. It may be extremely difficult or cost prohibitive to locate respondents in these situations. Snowball sampling relies on referrals from initial subjects to generate additional subjects.

SELECTING THE SAMPLE.

What is the optimum percentage of a population to accurately predict from this sample the characteristics of the population? Five percent of a population of 2,000 is not enough however, for a population of 8,000, it is correct. For a population of 40,000, a five-percent sample is five times larger than necessary. The key is to sample just enough people to assure confidence in the results, but no more. (Why waste money surveying more people than you need?)

Above a certain sample size, the margin of error decreases only slightly, regardless of the size of the population. You can achieve the "95-percent sure that the margin of error is less than 5 percent" standard with a sample as small as 385 respondents from a population of 8,000. Doubling the size of the sample (to 770 respondents) only reduces the margin of error by 1.4 percentage points. Decreasing the margin of error, and increasing the level of confidence both require drawing a larger sample.

The table below gives the sample size necessary to estimate population characteristics given various levels of sampling error, population size and variation. **The sample sizes indicated are based on a 95 percent confidence level.**

	+/- 3% Sampling error		+/- 5% sampling error		+/- 10% sampling error	
Population Size	**50/50 split**	**80/20 split**	**50/50 split**	**80/20 split**	**50/50 split**	**80/20 split**
100	92	87	80	71	49	38
250	203	183	152	124	70	49
750	441	358	254	185	85	57
1,000	516	406	278	198	88	58
5,000	880	601	357	234	94	61
10,000	964	639	370	240	95	61
25,000	1,023	665	378	234	96	61
100,000	1,056	678	383	245	96	61
1,000,000	1,066	682	384	246	96	61
1000,000,000	1,067	683	384	246	96	61

How to read this table: For a population with 1,200 members whom are expected to be about evenly split on the characteristics in which they are being measured, a sample of approximately 516 is needed to make estimates about the population with a sampling error of no more than +/- 3 percent, at the 95 percent confidence level. A "50/50" split means the population is relatively varied. An "80/20" split means the population is less varied; most people have a certain characteristic, a few do not. Unless the split is known with certainty it is best to be conservative and use 50/50. Numbers in the table refer to complete, usable questionnaires. – ***Table was developed using statistically valid methodologies by Priscilla Salant and Don A. Dillman, authors "How to Conduct a Survey"***

RESPONSE RATES

A survey's response rate is the proportion of persons included in the sample who actually complete the questionnaire or interview. A very low response rate leads one to suspect that the respondents are somehow quite different from the many who chose not to respond simply because they did respond. In general, the lower the response rate, the greater the potential that results will be misleading. Thus, surveys should focus on achieving high response rates rather than simply on obtaining large response numbers.

If the sample size for each group of interest is more than about 200, the accuracy of the survey results is often more seriously affected by low response rates than by small sample sizes. Thus, funds should be invested to enhance the response rate rather than to increase the sample size beyond the minimum number necessary to achieve the desired level of sampling accuracy.

Using the table above to determine the sample size for a population of 1,200 (table indicates a completed sample size of 516) and a 70% response rate (average) a starting sample of: 516 divided by .7 = 738 is needed.