DETERMINING SURVEY SAMPLE SIZE
A SIMPLE PLAN

Prepared by **Market Directions**

Market Directions

B O S T O N

617-323-1862

800-475-9808

www.marketdirectionsmr.com

info@marketdirectionsmr.com

DETERMING SAMPLE SIZE

A SIMPLE PLAN

TABLE OF CONTENTS

THE IMPORTANCE OF RANDOM SAMPLING

It is very important to develop a sampling plan that permits legitimate generalization from survey results to the population of interest. The key is the use of statistically derived random sampling procedures. These ensure that survey results can be defended as statistically representative of the population. Surveys that do not follow procedures can produce results that lead to misguided strategic, or policy decisions.

Most marketers developing and administering surveys are not statisticians, so attem guide survey sampling with complicated equations is not useful to most marketers. Presented is research on sampling that is understandable to those who seek sample not have a statistical background.

Don't forget 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.

INDICATION OF A LARGE SAMPLE

- The decisions have very serious consequences.
- Reliance on the data for critical decision making.
- 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
- Costs and timing are not an issue.
- Time and resources are readily available

INDICATION OF A SMALL SAMPLE

- 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.

Simple Sample Design

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 determine who is eligible and who is not.

LIST OF TARGET POPULATION

Develop the list from which the sample will be drawn. Use random number assignments in choosing the participants from the population. (Random number assignments are available in Excel and Access.)

SAMPLE SELECTION METHODS

Sampling methods are classified as either probability or nonprobability.

In probability samples, each member of the population has an equal probability (chance) of being selected.

In nonprobability sampling members are selected from the population in some nonrandom manner or based on subjective judgment.

PROBABILITY SAMPLING METHODS

STRATIFIED RANDOM

Stratified 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. After the require sample size has been calculated every N^{th} record is selected.

STRATIFIED SAMPLING

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

The advantage of probability sampling is that sampling error can be calculated.

NONPROBABILITY SAMPLING

CONVENIENCE AND JUDGMENT

Convenience and Judgment 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.

QUOTA SAMPLING

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

In nonprobability sampling, the degree to which the sample differs from the population remains unknown

SELECTING THE SAMPLE

What is the optimum percentage of a population to accurately predict from the sample the characteristics of the population?

Five percent of a population of 1,000 is not enough however, for a population of 10,000, it is correct. For a population of 100,000, a five-percent sample is five times larger than necessary. (%5 = 5,000 and for a 95% confidence level only 370 are required) The key is to sample just enough people to assure confidence in the results, but no more. (Why waste resources 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. This is standard with a sample as small as 370 respondents from a population of 10,000. Doubling the size of the sample (to 770 respondents) only reduces the margin of error (sampling 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. **Sample sizes are based on a 95 percent confidence level.**

TABLE TO SIMPLY DETERMINE CORRECT SAMPLE SIZE

How to read this table:

	± 3% Sampling error		± 5% Sampling error		± 10% Sampling error	
Split-→ Population	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
100,000,000	1,067	683	384	246	96	61

−*Table was developed using statistically valid methodologies by Priscilla Salant and Don A. Dillman, authors "How to Conduct a Survey"*

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.

If you can tolerate a larger sampling error at ± 5%, then a sample of only 278 are required.

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, it is best to be conservative.

RESPONSE RATES

A survey's response rate is the proportion of persons included in the sample who complete the process. 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.

Obviously, 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 sample.

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, budget 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.

CONFIDENCE LEVEL AND MARGIN OF ERROR

CONFIDENCE LEVEL

The purpose of taking a random sample from a population and computing a statistic, such as the mean, is to approximate the mean of the population. How well the sample statistic estimates the underlying population is the issue.

Confidence levels are stated as a percentage, such as 95 %. What does this mean? It means that if the same population is sampled on numerous occasions, the results would reflect the true population approximately 95 % of the time.

MARGIN OF ERROR OR SAMPLING ERROR

The margin of error (confidence *interval*) measures the maximum amount by which the survey results are expected to differ from the actual population. Results from a survey are an estimate the entire population.

The margin of error in a sample = 1 divided by the square root of the number in the sample.

With survey results you have a very educated guess about what the larger population thinks. If your samples size has a 5 percent margin of error that means that if you implemented the survey 100 times — asking a different sample of people each time — the overall percentage of people who responded the same way would remain within 5 percent of your original results at least 90 times.

(WARNING: Math Geek Stuff)

The margin of error is what statisticians call a confidence *interval*. The math behind it is much like the math behind the standard deviation. So you can think of the margin of error at the 90 percent confidence *interval* as being equal to two standard deviations in your sample.

Confidence interval and confidence level are NOT the same.

COMBING MARGIN OF ERROR AND CONFIDENCE LEVEL

This all boils down to understanding what it means when we say: *The results have a margin of error of ± 5 percent at a 95% confidence level.* This means that you can be assured that if 50 percent of the respondents answered "yes" to a particular question, you can be 95 percent certain if the question was asked of everyone within the population that 45% to 55% of the respondents would also answer "yes".