Chapter 7
SAMPLING PROCEDURES IN RESEARCH

Researchers must answer many questions when developing a project to collect information about a population of interest. Consider the following questions involving sampling:

- Should we take a census (complete canvas) or a sample?
- What kind of sample should be taken?
- What size should the sample be?

Answering these questions depends upon the application of statistical inference. This chapter first considers the selection of sample types and the process of sample planning. It then describes different kinds of samples. Finally, related to design, but a unique decision area in its own right, is the determination of the size of the sample.

PLANNING THE SAMPLE

Two broad objectives are fundamental to the use of samples in survey research projects:

- Estimation of information about a population based on a sample from the population,
- Hypothesis testing to compare the relationships between data items for some selected population groups.

Each involves making inferences about a population on the basis of information from a sample. The precision and accuracy of project results are affected by the manner in which the sample has been chosen. However, as Exhibit 7.1 illustrates, precision is only a reflection of sampling error and confidence limits and has nothing to do with accuracy.

Exhibit 7.1 Precision versus Accuracy in Sampling

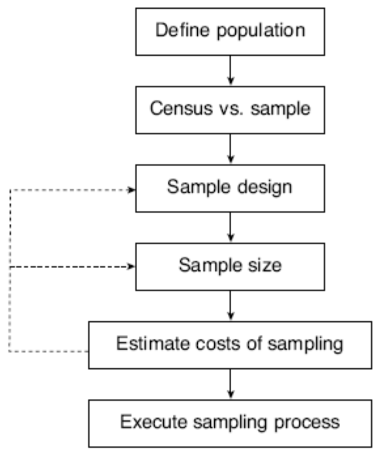
There often is confusion between precision and accuracy in marketing research. When a researcher speaks of sampling error and the size of the confidence limits placed on an estimate, that researcher is speaking of precision and not accuracy.

Accuracy is affected by nonresponse bias, memory error, misunderstanding of questions, problematic definition of terms, and processing errors, and not by sample size (Semon, 2000). The researcher must constantly ask: "Have we asked the right questions?" and "Have we measured what we want to measure properly and correctly?"

The difference between precision and accuracy is analogous to the difference between errors of degree and errors of kind. Accuracy, like errors of kind, are clearly more serious. If a researcher has made the correct measurements, precision will show the degree of error; but if the right measurements have not been taken, precision is probably meaningless (Semon, 2000). The quality of a sample depends upon the quality of design and execution of a research project at every stage of the process.

Consequently, strict attention must be paid to the planning of the sample. It must also be recognized that sample planning is only one part of planning the total research project. The process of selecting a sample follows the well-defined progression of steps shown in Figure 7.1, and will be discussed in turn.

Figure 7.1 Steps in Sample Planning



Defining the Population

The first step in sample panning is to define the population to be investigated. A population, also known as a universe, is defined as the totality of all units or elements (individuals, households, organizations, etc.) to which one desires to generalize study results. While seemingly an easy task, an imprecise research problem definition often leads to an imprecise population definition.

Specifying a population involves identifying which elements (in terms of kind) are included, as well as where and when. For example, a group medical practice that is considering expanding into sports medicine might acquire information from any or all of the distinct population groups listed in Table 7.1. The population element is the unit of analysis, and may be defined as an individual, household, institution, patient visit, and so on.

Table 7.1 Possible Population Choices for a Research Study by a Medical Practice

Group Members	*Where*	*When*
All patients	Designated group practice	Last 12 months
Patients who have had orthopedic work	Designated group practices	Last 12 months
All people	Specified geographic area	Last 12 months
All people who have had orthopedic work	Specified geographic area	Last 12 months

The second and third columns of Table 7.1 define population choices in terms of its location and timeframe. One useful approach is to first define the ideal population to meet study objectives, and then apply practical constraints to ultimately limit and define the study population. However over defining the population should be avoided unless it is absolutely necessary. Over defining can limit the extent to which findings can be generalized, and greatly increase the operational cost and difficulty of finding population elements (Sudman, 1976).

Census, or Sample?

Once the population has been defined, the investigator must decide whether to conduct the survey among all members of the population, or only a sample subset of the population. The desirability and advantages of using a sample rather than a census depend on a variety of factors such as geographic location, the absolute size of the population, and the sample size required for results sufficiently accurate and precise to achieve the required purposes.

Two major advantages of using a sample rather than a census are speed and timeliness. A survey based on a sample takes much less time to complete than one based on a census. Frequently, the use of a sample results is a notable economy of time, money and effort, especially when a census requires hiring, training and supervising many people.

In other situations, a sample is necessary because of the destructive nature of the measurement, such as in product testing. There is a related problem over surveying human populations when many different surveys need to be conducted on the same population within a relatively short period of time. Nonprobability sampling techniques explicitly protect against this problem.

In still other situations, a sample may control non-sampling errors. Samples (a smaller number of interviews compared to a census) may result in better interviewing, higher response rates through more call backs, and better measurement in general. The total amount of sampling and non-sampling error of a sample may actually be than the non-sampling error alone would be for a census. In Chapter 2, we emphasized the importance of minimizing total error.

However, under certain conditions a census may be preferable to a sample. When the population is small, the variance in the characteristic being measured is high, the cost of error is high, or the fixed costs of sampling are high, sampling may not be useful. In addition, if the characteristic or attribute of interest occurs rarely in the population, then a census of a tightly defined population might be desirable (for example, people with a rare genetic disorder). In this case, it would be necessary to sample a relatively large proportion of the general population to provide statistically reliable information. Obviously, the practicality of this depends upon the absolute size of the population and the occurrence rate for the characteristic of interest.

Sample Design

> *Sample design is "the theoretical basis and the practical means by which data are collected so that the characteristics of a population can be inferred with known estimates of error" (Survey Sampling, Inc.).*

Operationally, sample design is at the heart of sample planning. Sample design specification, including the method of selecting individual sample members, involves both theoretical and practical considerations (such as cost, time, labor involved, organization). The following checklist is suggested to obtain a sample that represents the target population (Fink, 2003):

1. Are the survey objectives stated precisely?
2. Are the eligibility criteria for survey respondents or experimental subjects clear and definite? Exclusion criteria rule out certain people.
3. Are rigorous sampling methods chosen? This involves selecting an appropriate probability or nonprobability sampling method.
4. Further questions to be answered in this section include:

- What type of sample should be used?
- What is the appropriate sampling unit?
- What is the appropriate frame (that is, list of sampling units from which the sample is to be drawn) for the particular design and unit decided upon?
- How are refusals and nonresponse to be handled?

Type of Sample

Much of the sampling in the behavioral sciences and in marketing research, by its nature, involves samples are selected by the judgment of the investigator, convenience, or other nonprobabilistic (nonrandom) processes. In contrast, probability samples offer the promise of bias-free selection of sample units and permits the measurement of sampling error. Nonprobability samples offer neither of these features. In nonprobability sampling one must rely on the expertise of the person taking the sample has selected respondents in an unbiased manner, whereas in probability sampling the sampling of respondents is independent of the investigator.

> **Example:** A dog food manufacturer tested consumer reactions to a new dog food by giving product samples to employees who own dogs and eliciting their responses about a week later. The employees' dogs liked the food and the pet food manufacturer actually introduced the new dog food product. However when it hit the market the product was a flop… dogs simply would not eat it. As managers scrambled to find out what went wrong, research showed that employees were so loyal to the company's products that their dogs had little variety and would eat anything for a change. In the broader market dogs were used to a greater variety of dog foods including table scraps and just did not like the taste of the new dog food. In this case, a biased sample was erroneously assumed to conform to the general population of dogs and dog owners.

A researcher choosing between probability and nonprobability sampling implicitly chooses the probability sample's relative size of sampling error against the nonprobability sample's combined sampling error and selection bias. For a given cost, one can usually select a larger nonprobability sample than probability sample, meaning that the sampling error should be lower in the nonprobability sample, but that the nonrandom process used for selecting the sample may have introduced a selection bias.

The Sampling Unit

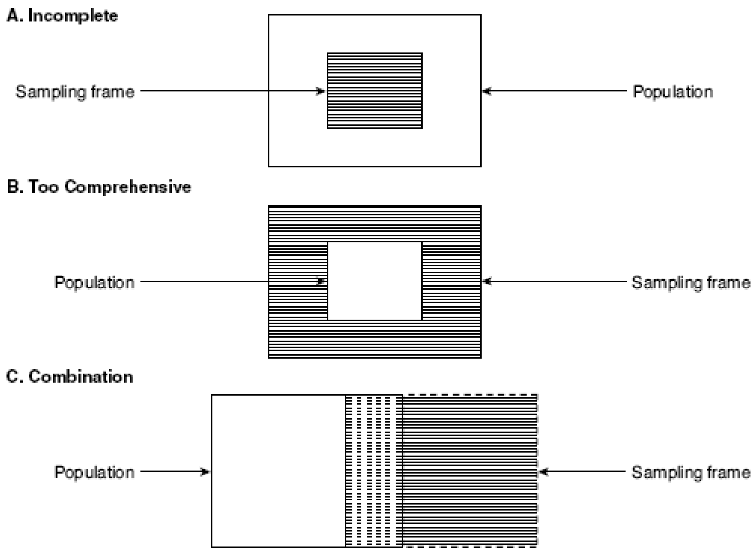
The sampling unit is the unit of the population to actually be chosen during the sampling process. The sampling unit may contain one or more elements describing the population. For instance, a group medical practice may be interested in surveying past patient behavior of the male wage earner *or* his entire household. In either case, it may be preferable to select a sample of households as sampling units.

The Sampling Frame

A sampling frame is a means of identifying, assessing and selecting the elements in the population. The sampling frame usually is a physical listing of the population elements. In those instances where such a listing is not available, the frame is a procedure producing a result equivalent to a physical listing. For instance, in a consumer survey where personal interviews are conducted with a mall intercept method is used to obtain data, the sample frame includes all those people who enter the mall during the study period.

Ideally, the sample frame should identify each population element once, but only once, and not include elements not in the defined population. Such a perfect frame is seldom available for marketing research purposes. As shown in Figure 7.2, a sampling frame may be incomplete, too comprehensive, or a combination of both. In addition, the frame may include individual population elements more than once. Any of these situations can lead to coverage error.

Figure 7.2 Sampling Frame – Population Relationship



Perhaps the most widely used frame in survey research for sampling human populations is the telephone directory. Use of such a frame, however, may lead to frame error from the exclusion of non-subscribers (no home phone or cell phone only), voluntarily unlisted subscribers, and involuntarily unlisted subscribers. For example, more than one in five American homes (20.2%) has only wireless telephones (National Center for Health Statistics, 2009). Combined with the fact that in many major cities more than 50% of home phones are unlisted, the potential for frame errors is huge. Similar frame errors can arise from such other sampling frames as city directories, maps, trade association membership lists, or any other incomplete or dated listing or representation of a population. The National Do Not Call Registry provisions do not cover calls from political organizations, charities, telephone surveyors, or companies with which a consumer has an existing business relationship, nonetheless, resident push back leads to further frame error.

Even though today's panel management software makes compiling and managing sample lists relatively easy, sampling is the single most important problem in e-mail and Internet-based surveys. Dillman (2000) suggests asking the following five questions about any potential sampling list:

- Does the list contain everyone in the survey population?
- Does the list include names of people who are not in the study population?
- How is the list maintained and updated?
- Are the same sample units included on the list more than once?
- Does the list contain other respondent information that can be used to customize the survey or direct survey logic?

Whether the researcher uses their own list, or one purchased from a commercial source, these questions are equally applicable.

Sample Size

Sample size, as part of sample planning is related to precision. Four traditional approaches can determine the appropriate sample size for any given research project:

1. Arbitrarily or judgmentally determined
2. Minimum cell size needed for analysis
3. Budget-based
4. Specifying desired precision in advance of sampling and then applying the appropriate standard error formula to determine the sample size.

For approaches 1-3, the precision could be measured after compiling the data and then applying the appropriate standard error formula or formulas if a probability design is being used.

Costs of Sampling

The sample plan must account for the estimated costs of sampling. Such costs include **overhead costs**, which are relatively fixed for a sampling procedure, and **variable costs**, which depend on the number of respondents contacted and interviewed in the study. In reality, costs from all aspects of data collection are usually considered together. The dashed line shown in Figure 7.1 when the sample design and sample size are considered, the estimated costs may be so great that other sample designs and/or smaller-sized samples may be considered.

Execution of the Sampling Process

The last step in sample planning is to execute the sampling process. A sample is chose that is thought to be both representative and to adequately mirror the various patterns and subclasses of the population (Exhibit 7.2). The sample should be of sufficient size to provide confidence in the stability of its characteristics.

Exhibit 7.2 Samples Are Not Always an Exact Match

Many researchers feel that the best way to assess the validity of a sample is to compare its demographic profile (i.e., distributions of the key demographic characteristics) with a national or otherwise known profile. This alone does not necessarily guarantee a good sample. At the same time, a poor fit does not necessarily mean that the obtained sample is bad.

When the fit is not good, some researchers tend to develop a weighting scheme. However many studies have shown that large differences in demographic characteristics may to translate into small differences in the variable of interest, whether it is a behavior or an attitudes, interests, or opinions (AIO) measure. Of course, there could be serious distortion if specific segments are of concern. But the differences generally have to be much greater than we would think for a significant effect.

This, in turn, requires a measure of precision, which necessitates using probability-based design. From this discussion, one might conclude that an ideal sample is obtained with a probability process. In general, this is preferable. However, recognize that it is more important to avoid distorted samples than to be able to measure sampling error. There may be a tendency to ignore the existence of potential bias when using probability designs.

This chapter is primarily concerned with probability sampling, but before discussing this topic it is useful to describe some of the procedures by which nonprobability samples are taken in marketing research.

NONPROBABILITY SAMPLING PROCEDURES

Nonprobability sampling is distinguished from probability sampling in that nonprobability sample elements do not have a known, nonzero chance of being selected for the sample. As such, sampling error is generally not able to be measured. Nonprobability samples are widely used in exploratory research, but are also valuable for non-exploratory research.

The Quota Sample

The quota sample is the most commonly employed nonprobability sampling procedure in marketing research. Quota samples are collected to reflect proportions in the various subclasses (or strata) of the population of interest. This might be, for example, the proportion of the adult population who fall into various age-by-gender-by-education groupings.

To prepare a quota sample, the subclass proportions are first estimated from some outside source, such as census data. Next, if an interviewer has a total number of, say, 600 interviews to obtain, the age-gender-education proportions in the population are applied to the 600 total interviews to determine the appropriate quotas.

However, in quota sampling the interviewer is *not* required to *randomly* select the respondents necessary to fill each quota category. The lack of random selection is the major distinction between quota sampling and stratified random sampling.

In quota sampling, the interviewer's judgment is relied upon to select actual respondents within each quota. Therefore, many sources of selection bias are potentially present. For example, the interviewer may not bother to call back if the first call results in a "not-at-home." Interviewers may go to selected areas where the chances are good that a particular type of respondent is available. Certain houses may be skipped because the interviewer does not like the appearance of the property, and certain people may be allowed to pass in a mall-intercept approach because they do not "look right." Still other interviewer habits and biases exist that can influence their selection of respondents within the quota.

The advantages of quota sampling are the lower costs and greater convenience provided to the interviewer when selecting respondents to fill each quota. Quota sampling is quite close to traditional probability sampling when selection is tightly controlled (Sudman, 1976).

Quota sampling is discussed rather extensively in Stephan and McCarthy (1958). A critique of the method is offered by Deming (1960). A comparison of quota and probability sampling was the subject of a seminar held in 1994 at the Survey Methods Centre in the United Kingdom (Survey Methods Centre, 1995).

The Judgment Sample

A somewhat representative sample may be provided through the use of purposive, or judgment, sampling. Sound judgment or expertise and an appropriate strategy, leads one to carefully and consciously choose the elements so as to develop a suitable sample. The intent is to select respondents representative of the population in such a way that errors of judgment in the selection will cancel each other out. The relative advantages of judgment sampling are that it is inexpensive, convenient to use, less time-consuming, and provides results as good as probability sampling. One weakness of this approach is that without an objective basis for making the judgments, there is no way of knowing whether the so-called typical cases are, in fact, typical.

The Convenience Sample

Convenience sampling is a generic term that covers a wide variety of ad hoc procedures for selecting respondents. Convenience sampling means that the sampling units are accessible, convenient and easy to measure, cooperative, or articulate. An illustration of convenience sampling is “intercept” interviews among shopping-mall customers or in other areas where large numbers of consumers may congregate. In this case, the researcher needs to select respondents with care. People frequent malls for different reasons and stay in malls different lengths of time, which leads to a biasing effect if a probability sample is attempted. A sample is said to be *length biased* if the probability of observing an individual at a particular site is dependent on the individual’s length of stay at the site.

Firms may also authorize samples to be taken from such intact groups as Parent-Teacher Associations, church groups, philanthropic organizations, and so on. Again, the purpose is to obtain a relatively large number of interviews quickly from a cooperating group of respondents. Usually the sponsoring organization receives a donation from the interviewing firm for the help and cooperation of its members. However depending on the purpose of the research, many potential sources of selection bias may exist in that only certain members may respond, who may be disproportionately different on one of many demographic, attitudinal or behavioral dimensions.

Snowball Sampling

Snowball sampling (also known as multiplicity or chain-referral sampling) is the rather colorful name given to a procedure in which initial respondents are selected randomly, but where additional respondents are then obtained from referrals or other information provided by the initial respondents. One major advantage of snowball sampling is the ability to estimate various characteristics that are rare in the total population. A more complete discussion of this technique is given by Zinkhan, Burton, and Wallendorf (1983).

> **Example**: A study of international tourism, required researchers to interview respondents in the United Kingdom, France, and Germany who visited the United States in its bicentennial year. Given the likelihood of finding a qualified adult respondent was less than two percent, stratified probability methods were used to select initial respondents. A referral procedure (up to two referrals per qualified respondent) was then used to obtain a second group of qualified respondents. (However, this particular study did not obtain subsequent referrals from this second group of respondents.)

In other types of snowball sampling, referrals from referrals are obtained, and so on, thus leading to the “snowballing” effect. Even though a probability-based procedure may be used to select the initial group of respondents, the overall sample is a nonprobability sample.

Major advantages of this type of sampling over conventional methods are the substantially increased probability of finding the desired characteristic in the population, and lower sampling variance and costs.

PROBABILITY SAMPLING PROCEDURES

The best-known type of probability sample is no doubt the simple random sample. However, many occasions in marketing research require more specialized sampling procedures than those of simple random-sampling methods. Statisticians have developed a variety of specialized probability-sampling designs that, although derived from simple random-sampling principles, can be used to reduce sampling error and cost. Five major modifications can be made to the basic selection process, as shown in Table 7.2.

These techniques are discussed in turn, following a review of simple random sampling. Our purpose is to describe the major characteristics of each technique, rather than to present a detailed mathematical exposition of its procedures. Many excellent statistics and research books review the mathematical aspects of these sampling techniques.

The Simple Random Sample

In a simple random sample, each sample element has a known and equal probability of selection, and each possible sample of *n* elements has a known and equal probability of being the sample actually selected. It is drawn by a random procedure from a sample frame—a list containing an exclusive and exhaustive enumeration of all sample elements. One widely used process for generating a simple random sample is: upload the elements of the sample frame to a spreadsheet, number them, and then use the random-number generator function to select the sample members.

Simple random samples are not widely used in consumer research, for two reasons. First, it is often difficult to obtain a sampling frame that will permit a simple random sample to be drawn and second, one may not want to give all sample units an equal probability of being selected. Consumer research usually requires people, households, stores, or areas to be the basic sampling units. While a complete representation of areas is available through maps, there normally is no complete listing of persons, the households in which they live, or the stores available. When persons, households, or stores are to be sampled, some other sample design must be used.

In business-to-business (B2B) research, there is a greater opportunity to apply simple random sampling. In this case purchasing agents, companies, or areas are the usual sampling units and the population under study is often relatively small. One is therefore in a better position to develop a complete list of respondents or sample frame.

Table 7.2 Selection Methods for Probability Samples

Probability Samples	*Nonprobability Samples*
I. *Equal probability* for all elements a. Equal probabilities at all stages b. Equal overall probabilities for all elements obtained through compensating unequal probabilities at several stages	*Unequal probabilities* for different elements; ordinarily compensated with inverse weights a. Caused by irregularities in selection frames and procedures b. Disproportionate allocation designed for optimum allocation
II. *Element Sampling*: single stage, sampling	*Cluster Sampling*: sampling units are clusters of elements

unit contains only one element	a. One-stage cluster sampling b. Sub-sampling or multistage sampling c. Equal clusters d. Unequal clusters
III. *Unstratified Selection*: sampling units selected from entire population	*Stratified Sampling*: separated selections from partitions, or strata, of population
IV. *Random Selection* of individual sampling units from entire stratum or population	*Systematic Selection* of sampling units with selection interval applied to list
V. *One-Phase Sampling*: final sample selected directly from entire population	*Two-Phase (or Double) Sampling*: final sample selected from first-phase sample, which obtains information for stratification or estimation

SOURCE: Adapted from Kish, 1965, p. 20.

The Systematic Sample

Systematic sampling involves only a slight variation from simple random sampling. In a systematic sample, each sample element has a known and equal probability of selection. The permissible samples of size *n* that are possible to draw each have a known and equal probability of selection, while the remaining samples of size *n* have zero probability of being selected.

The mechanics of taking a systematic sample are rather simple. If the population contains *N* ordered elements and a sample size *n* is desired, one merely finds the ratio of *N*/*n* and rounds to the nearest integer to obtain the sampling interval. For example, if there are 600 members of the population and one desires a sample of 60, the sampling interval is 10. A random number is then selected between 1 and 10, inclusively; suppose the number turns out to be 4. The analyst then takes as the sample elements 4, 14, 24, and so on.

Essentially, systematic sampling assumes that population elements are ordered in some fashion—names in a telephone directory, a card index file, or the like. Some types of ordering, such as an alphabetical listing, will usually be uncorrelated with the characteristic (say, income level) being investigated. In other instances, the ordering may be related to the characteristic under study, as when a customer list is arranged in decreasing order of annual purchase volume.

If the arrangement of the elements of the sample is itself random with regard to the characteristic under study, systematic sampling will tend to give results close to those provided by simple random sampling. We say “close” because in systematic sampling, all combinations of the characteristic do not have the same chance of being included. For example, it is clear that in the preceding example, the fifth, sixth, and so on items have zero chance of being chosen in the particular sample after the first item has been determined.

Systematic sampling may increase the sample’s representativeness when items are ordered with regard to the characteristic of interest. For example, if the analyst is sampling a customer group ordered by decreasing purchase volume, a systematic sample will be sure to contain both high- and low-volume customers. On the other hand, the simple random sample may yield, say, only low-volume customers, and may thus be unrepresentative of the population being sampled if the characteristic of interest is related to purchase volume.

It is also possible that systematic sampling may decrease the representativeness of the sample in instances where the items are ordered in a cyclical pattern. For example, a sample interval of 7 for a systematic sampling of daily retail-store sales figures would reflect the same day of data for each week and would not reveal the day-of-the-week variations in sales.

The Stratified Sample

It is sometimes desirable to break the population into different strata based on one or more characteristics, such as the frequency of purchase of a product, type of purchase (e.g., credit card versus non-credit card), or the industry in which a company competes. In such cases, a separate sample is then taken from each stratum. Technically, a stratified random sample is one in which a simple random sample is taken from each stratum of interest in the population. In practice, however, systematic and other types of random samples are sometimes taken from each of the strata. In this case, the resulting design is still referred to as a stratified sample.

Stratified samples are generally conducted according to the following procedure:

- The entire population is first divided into an exclusive and exhaustive set of strata, using some external source, such as census data, to form the strata.
- A separate random sample is selected within each stratum.
- From each separate sample, some statistic (such as a mean) is computed and properly weighted to form an overall estimated mean for the whole population.
- Sample variances are also computed within each separate stratum and appropriately weighted to yield a combined estimate for the whole population.

The two basic varieties of stratified samples are proportionate and disproportionate. In proportionate stratified sampling, the sample drawn from each stratum is made proportionate in size to the relative size of that stratum in the total population. In disproportionate stratified sampling, one departs from the preceding proportionality by taking other circumstances, such as the relative size of stratum variances, into account.

> **Example**: A company is interested in estimating the average purchases of consumers of hot cereal. The researcher may be willing to assume that, although average consumption would vary markedly by family size, the variances around the means of the strata would be more or less equal among family sizes. If so, the researcher would make use of proportionate stratified sampling.

More generally, however, both means and variances will differ among strata. If this is the case, the researcher would make use of disproportionate stratified sampling. In this instance, the number of families included in each stratum would be proportionate to (the product of) the relative size of the different family-sized strata in the population and the standard deviation of each family class. This requires, of course, that the researcher be able to estimate (from past studies) the within-group standard deviation around the average purchase quantity of each purchasing stratum. Formulas for computing sampling errors in stratified samples are briefly discussed later in this chapter and can be found in standard texts on sampling.

As intuition would suggest, the increased efficiency of stratified sampling over simple random sampling depends on how different the means (or some other statistic) really are among strata, relative to the within-stratum variability. The greater the within-stratum homogeneity and among-stratum heterogeneity, the more efficient the stratified sampling is relative to simple random sampling.

Two final considerations need to be addressed regarding stratified sampling. First, a sample size needs to be calculated for each stratum or subgroup. Second, the process of sample selection can be time-consuming and costly to carry out if many subgroups are to be used.

Nevertheless, this method continues to be widely used due to the segmentation of markets that companies routinely engage in.

The Cluster Sample

The researcher will ordinarily be interested in the characteristics of some elementary element in the population such as individual family attitudes toward a new product. However, when larger primary sampling units are desired, cluster sampling may be used. For example, the researcher may choose to sample city blocks and interview all the individual families residing therein. The blocks, not the individual families, would be selected at random. Each block consists of a cluster of respondents. The main advantage of a cluster sample relative to simple random sampling is in lower interviewing costs rather than in greater reliability.

The Area Sample: Single Stage and Multistage

As the name suggests, area sampling pertains to primary sampling of geographical areas—for example, counties, townships, blocks and other area descriptions. A single-stage area sample occurs when only one level of sampling takes place (such as a sampling of blocks) before the basic elements are sampled (the households). If a hierarchy of samples within the larger area is taken before settling on the final clusters, the resulting design is usually referred to as a multistage area sample.

> Example: Consider the sample design used by the Gallup Organization for taking a nationwide poll. Gallup draws a random sample of locations as the first stage of the sampling process. Blocks or geographic segments are then randomly sampled from each of these locations in a second stage, followed by a systematic sampling of households within the blocks or segments. A total of about 1,500 persons are usually interviewed in the typical Gallup poll.

METHODS FOR DETERMINING SAMPLE SIZE

There are several ways to classify techniques for determining sample size. Two that are of primary importance are the following:

- Whether the technique deals with fixed or sequential sampling
- Whether its logic is based on traditional or Bayesian inferential methods

Other than the brief discussion of sequential sampling that follows, this chapter is concerned with the determination of a fixed sample size with emphasis on traditional inference, such as Neyman-Pearson, rather than Bayesian inference.[1]

Although the discussion will focus on the statistical aspects of setting sample size, it should be recognized that nonstatistical dimensions can affect the value of a research project. Such things as study objectives, the length of a questionnaire, budget, time schedule, and the requirements of data analysis procedures, all have a direct effect on sample size decisions.

Fixed Versus Sequential Sampling

As the name implies, in fixed-size sampling the number of items sampled is decided in advance. The size of the sample is chosen to achieve a balance between sample reliability and sample cost. In general, all sample observations are taken before the data are analyzed.

In sequential sampling, however, the number of items is not preselected. Rather, the analyst sets up in advance a decision rule that includes not only the alternative of stopping the sampling process (and taking appropriate action, based on the sample evidence already in hand) but also the possibility of collecting more information before making a final decision. Observations may be taken either singly or in groups, the chief novelty being that the data are analyzed as they are assembled and sample size is not predetermined.

In general, sequential sampling will lead to smaller sample sizes, on average, than those associated with fixed-size samples of a given reliability. The mathematics underlying sequential sampling are, however, more complex and time consuming. In addition, the problem may be such that it is less expensive to select and analyze a sample of many items at one time than to draw items one at a time (or in small groups) and analyze each item before selecting the next.

Sampling Basics: Terminology

Intuitively we would expect that when we increase the size of the sample, our estimate of the population parameter should get closer to the true value. Also, we would expect that the less dispersed the population's characteristics are, the closer our sample estimates should be to the true parameter. After all, the reason why we sample in the first place is to make some inference about the population. These inferences should be more reliable when the sample is larger and when there is less variability in the population variables measured. To see this in action, consider the example in Exhibit 7.3.

When you read about samples in newspapers or other documents, the researcher often reports the margin of error or confidence interval for the statistical findings reported in the study. The "**margin of error"** or "**confidence interval"** is the plus-or-minus figure that represents the accuracy of the reported. Consider another example:

> A Canadian national sample showed "Who Canadians spend their money on for Mother's Day." Eighty-two percent of Canadians expect to buy gifts for their mom, compared to 20 percent for their wife and 15 percent for their mother-in-law. In terms of spending, Canadians expect to spend $93 on their wife this Mother's Day versus $58 on their mother. The national findings are accurate, plus or minus 2.75 percent, 19 times out of 20.

In this example, if 82% of your sample indicates they will "buy a gift for mom" and you use a confidence interval of 2.75%, you can be "95% confident that for ALL CANADIANS, somewhere between 79.25% (82%-2.75%) and 84.75% (82%+2.75%) would have picked that answer.

The "**confidence level"** tell you how confident you are of this result. It is expressed as a percentage of times that different samples (if repeated samples were drawn) would produce this result. The 95% confidence level means that if 20 different samples were drawn, 19 times out of 20, the results would fall in this - + confidence interval. A 99% confidence level would mean that 99 out of 100 times, the results would fall into the stated -+ confidence interval. The 95% confidence level is the most commonly used.

When you put the confidence level and the confidence interval together, you can say that you are 95% (19 out of 20 times) sure that the true percentage of the Canadian population that will "buy a gift for mom" is between 79.25% and 84.75%.

Wider confidence intervals increase the certainty that the true answer is within the range specified. These wider confidence intervals are associated with smaller sample sizes and of course produce larger sampling errors. When the costs incurred from making an error are extremely high (you are betting your company, or a multi-million dollar decision is being made) the confidence interval should be kept small. This can be done by increasing the sample size to reduce the sampling error.

Exhibit 7.3 How Do Election Polls Work?

The following is an Edited version of "Inside the paper's election polls", an article by Elsa McDowell that appeared in The Charleston Post and Courier:

The beauty of... election polls is that they are straightforward. They use statistical formulae to estimate how many people will vote one way and how many will vote another. No spin. No qualifying clauses to muddy the picture. The difficulty of.. election polls is that they are not always straightforward. How else could you explain that a poll done by one candidate shows him in the lead and that a poll done by his opponent shows her in the lead? Statisticians say there are ways to twist questions or interpret answers to give one candidate an advantage over another.

One reader took issue with a recent poll results run in The Post and Courier. He questioned whether the methodology was described in enough detail, whether the sample size was adequate. He was right about one point. The story did not make clear who was polled. It said "voters" and failed to elaborate. It should have indicated that the people polled were registered and likely to vote in the November elections. His next point is debatable. He said the sample size of 625 likely voters was insufficient for a state with nearly 4 million residents and suggested at least 800 should have been polled.

Brad Coker, the researcher responsible for the study responded that "the standard sample size used by polling groups nationally is 625. It produces, as the story stated, a margin of error of plus-or-minus 4 percent. Increasing the sample size to 800 would have produced a margin of error of plus-or-minus 3.5 - more accurate, but not so much more accurate to justify the additional cost."

"Many people do not understand how sample sizes work. They believe that, the larger the pool, the larger the sample size needs to be. It's not like that. You can take a drop of blood from a 400-pound person and it will contain the same data you would get if you took it from a 100-pound person," he said.

The reader's next concern was that the margin of error of plus-or-minus 4 applies only to the group viewed in its entirety. "If 'minorities' constituted 27 percent of the total sample, then only 169 were sampled. The margin of error then skyrockets into double digits." Coker said the reader is right and wrong. The margin of error (also known as a confidence interval) does jump for subgroups, but does not reach double digits. In this case, it moves from plus-or-minus 4 to plus-or-minus 6 to 8.

Two days before The Post and Courier ran their poll, another short story was run about a poll commissioned by MSNBC. That poll indicated incumbent Gov. Jim Hodges with a 45-43 percent lead. (Our) poll indicated challenger Mark Sanford was ahead 45 to 41 percent. When the margin of error is considered, both polls show the race is still a toss-up.

Controlling the Size of the Confidence Interval

Sampling theory teaches us that the accuracy of a sample estimate is dependent on such factors as the dispersion and skewness of the population's responses, the sample size, and the size of the population. Controlling these variables contributes to the incidence (and elimination) of sampling error. Note that "non-sampling" errors, such as bad question design or selection of a "bad" sample frame are not controlled by sample size.

<u>Sample Size</u>

Larger sample sizes generally produce a more accurate picture of the true characteristics of the population. Larger samples tighten the size of the confidence interval, making your estimate much more accurate. This relationship is not linear as shown in Table 7.3. Increasing sample size from 500 to 1000 reduces the confidence interval from +- 4.38 to +- 3.1.

<u>Dispersion</u>

The accuracy of an estimate also depends on the dispersion and skewness of the population on the question being asked. A sample of individuals registered for the republican political party would likely give a less dispersed evaluation of former president George W. Bush than would a sample of democrats. Likewise, a sample of Catholic priests would have less variability on the issue of abortion than would a survey of the general population. Accuracy of the sample estimate increases as the dispersion in the population decreases. Depending on the method of sample size calculation, dispersion is expressed as sample variance or as a proportion holding alternative positions (favorable or unfavorable toward an issue).

When using a proportion to compute sample size or estimate confidence intervals, it is easy to cover all eventualities by assuming maximum variability (50-50 proportions). Likewise, once your data has been collected, the observed proportion and final sample size can be used to obtain a more accurate estimate of the actual confidence interval.

<u>Population Size</u>

The size of the population also influences the size of the confidence interval, but not as much as you might expect. For a sample of 1000 respondents from a population of 100,000, the confidence interval is +- 3.08%. However, if the population were instead 1 million, the confidence interval widens to only +- 3.1%. The confidence interval is far more sensitive to changes in the sample size than to the size of the total population.

Non-sampling errors cannot be compensated for by increased sample size. Often, larger samples accentuate non-sampling errors rather than reduce them. Non-sampling errors come from samples that are not truly random, bad scales, misleading questions, incomplete surveys, etc.

Table 7.3 Proportion Based Confidence Intervals Computed at the 95% Confidence Level

Sample Size	Variability Proportions					
	50/50%	40/60%	30/70%	20/80%	90/10%	95/5%
25	20	19.6	18.3	16	12	8.7
50	14.2	13.9	13	11.4	8.5	6.2
75	11.5	11.3	10.5	9.2	6.9	5
100	10	9.8	9.2	8	6	4.4
150	8.2	8	7.5	6.6	4.9	3.6
200	7.1	7	6.5	5.7	4.3	3.1
250	6.3	6.2	5.8	5	3.8	2.7
300	5.8	5.7	5.3	4.6	3.5	2.5
400	5	4.9	4.6	4	3	2.2
500	4.5	4.4	*4.1	3.6	2.7	2
600	4.1	4	3.8	3.3	2.5	1.8
800	3.5	3.4	3.2	2.8	2.1	1.5
1000	3.1	3.0	2.8	2.5	1.9	1.4
1500	2.5	2.5	2.3	2.0	1.5	1.1
2000	2.2	2.2	2.0	1.6	1.2	0.96
2500	2	1.9	1.8	1.6	1.2	0.85
5000	1.4	1.4	1.3	1.1	.83	0.6

*Example Interpretation: In a product usage study where the expected product usage incidence rate is 30%, a sample of 500 will yield a precision of +/- 4.1 percentage points at the 95% confidence level.

This table is compute using the following formula:

$$\frac{(\text{Number of Standard Errors})^2 * ((\text{proportion})*(1-\text{proportion})) / (\text{Accuracy})}{(1+((\text{Number of Standard Errors})^2 * ((\text{proportion})*(1-\text{proportion})) / (\text{Accuracy})-1) / (\text{the population size})}$$

This formula is easily entered into a spreadsheet, to compute a sample size determination table.

Sampling Basics: Sampling Distributions and Standard Errors

The reader will recall from elementary statistics the concept of a sampling distribution. For a specified sample statistic (e.g., the sample mean) the sampling distribution is the probability distribution for all possible random samples of a given size *n* drawn from the specified population. The standard error of the statistic is the standard deviation of the specified sampling distribution. We shall use the following symbols in our brief review of the elementary formulas for calculating the standard error of the mean and proportion (under simple random sampling):

μ = population mean
π = population proportion regarding some attribute
σ = standard deviation of the population
s = standard deviation of the sample, adjusted to serve as an estimate of the standard deviation of the population
$\overline{X}$ = arithmetic mean of a sample
p = sample proportion
n = number of items in the sample

We here identify eight important properties associated with sampling distributions:

1. The arithmetic mean of the sampling distribution of the mean ($\overline{X}$) or of the proportion (*p*) for any given size sample equals the corresponding parameter values μ and π, respectively.

2. The sampling distribution of the means of random samples will tend toward the *normal distribution* as sample size *n* increases, regardless of the original form of the population being sampled.

3. For large samples (e.g., $n \geq 100$ and for π fairly close to 0.5) the normal distribution also represents a reasonable approximation of the binomial distribution for sample proportions.

4. In the case of finite universes, where the sample size *n* is some appreciable fraction of the total number of items in the universe, *N*, the standard error formulas should be adjusted by multiplication by the following *finite multiplier*:

$$\sqrt{\frac{N-n}{N-1}}$$

   For practical purposes, however, use of the finite multiplier is not required unless the sample contains an appreciable fraction, say 10% or more, of the population being sampled. At a 10% sampling fraction, taking it into account will reduce the random sampling error by 5%. If the sampling fraction is 5%, 2%, or 1%, not ignoring it will reduce error very slightly—2.5%, 1.0%, and 0.5%, respectively.

5. Probabilities of normally distributed variables depend on the distance (expressed in multiples of the standard deviation) of the value of the variable from the distribution's mean. If we subtract a given population mean μ from a normally distributed variable X_i and divide this result by the original standard deviation σ, we get a standardized variable Z_i that is also normally distributed but with zero mean and unit standard deviation. In symbols this is as follows:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

   Table A.1 in Appendix A at the end of this book presents the standardized normal distribution in tabular form. Note further that the original variate may be a sample mean, $\overline{X}$. If so, the denominator is the *standard error* (i.e., standard deviation of the sampling distribution). We can then define *Z* as some number of standard errors away from the mean of the sampling distribution

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

   where $\sigma_{\overline{X}}$ denotes the standard error of the mean. (A similar idea is involved in the case of the standard error of the proportion.)

6. The formulas for the standard error of the mean and proportion of simple random samples are, respectively, the following: Mean Proportion

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \qquad \sigma_{\overline{P}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

7. If the population standard deviation σ is not known, which is often the case, we can estimate it from the sample observations by use of the following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_1 - \overline{X})^2}{n-1}}$$

We can consider *s* to be an *estimator* of the population standard deviation, σ. In small samples (e.g., less than 30), the *t* distribution of Table A.2 in Appendix A is appropriate for finding probability points. However, if the sample size exceeds 30 or so, the standardized normal distribution of Table A.1 is a good approximation of the *t* distribution.
In cases where σ is estimated by *s*, the standard error of the mean becomes

$$est.\sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

where $est.\sigma_{\overline{X}}$ denotes the fact that σ is estimated from *s*, as defined in the preceding equation.

8. Analogously, in the case of the standard error of the proportion, we can use the sample proportion *p* as an estimate of π to obtain

$$est.\sigma_{\overline{P}} = \sqrt{\frac{p(1-p)}{n}}$$

as an estimated standard error of the proportion. Strictly speaking, *n*–1 should appear in the denominator. However, if *n* exceeds about 100 (which is typical of the samples obtained in marketing research), this adjustment makes little difference in the results.

Methods of Estimating Sample Size

In our discussion of sample planning we pointed out that there are four traditional approaches to determining sample size.

- First, the analyst can simply select a size either arbitrarily or on the basis of some judgmentally based criterion. Similarly, there may be instances where the size of sample represents all that was available at the time—such as when a sample is composed of members of some organization and data collection occurs during a meeting of the organization.

- Second, analysis considerations may be involved and the sample size is determined from the minimum cell size needed. For example, if the critical aspect of the analysis required a cross-tabulation on three categorical variables that created 12 cells (2 categories x 3 categories x 2 categories = 12 cells), and it was felt that there should be at least 30 observations in a cell, then the absolute minimum sample size needed would be 360.

- Third, the budget may determine the sample size. If, for example, the research design for an online survey of physicians, the cost of each interview was estimated to be \$50, and the budget allotted to data collection was \$10,000, then the sample size would be 200.

It may appear that these methods are for nonprobability samples. While this certainly is true, these methods are also applicable to probability samples and have occasionally been used for such samples. For probability samples, the precision must be determined after the fact.

- In a fourth approach to sample size determination, we specify the desired precision in advance and then applying the appropriate standard error formula to calculate the sample size. This approach using traditional inference (Neyman-Pearson), relies on either of two major classes of procedures available for estimating sample sizes. The first and better known of these is based on the idea of constructing confidence intervals around sample means or proportions. This can be called the confidence-interval approach. The second approach makes use of both type I (rejecting the true null hypothesis) and type II (accepting a false null hypothesis) error risks and can be called the hypothesis-testing approach. We discuss each of these approaches in turn.

Before doing this, however, two points must be made. First, as with the other approaches, the analyst must still calculate the standard error after data collection in order to know what it is for the actual sample that provided data. Second, the size of sample that results from traditional inference refers to the completed (or resulting) sample. Depending on the data collection method used, the original survey delivery may have to be much larger. For example, suppose that the size of the desired sample was 582. A mail survey is used for data collection and past experience has shown that the response rate would be around 25%. The original sample size in this case would have to be 2,328 so that 582 responses would be obtained.

The Confidence Interval Approach

It is not unusual to construct a confidence interval around some sample-based mean or proportion. The standard error formulas are employed for this purpose. For example, suppose a researcher sampled 100 student consumers and noted that their average per capita consumption of specialty fruit/energy drinks was 2.6 pints per week. Past studies indicate that the population standard deviation σ can be assumed to be 0.3 pint.

With this information, we can find a range around the sample mean level of 2.6 pints for which some prespecified probability statement can be made about the process underlying the construction of such confidence intervals.

For example, suppose that we wished to set up a 95% confidence interval around the sample mean of 2.6 pints. We would proceed by first computing the standard error of the mean:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{100}} = 0.03$$

From Table A.1 in Appendix A we find that the central 95% of the normal distribution lies within ±1.96 *Z* variates (2.5% of the total area is in each tail of the normal curve).
With this information we can then set up the 95% confidence interval as

$$\overline{X} \pm 1.96\sigma_{\overline{X}} = 2.6 \pm 1.96(0.03)$$

and we note that the 95% confidence interval ranges from 2.54 to 2.66 pints.
Thus, the preassigned chance of finding the true population mean to be within 2.54 and 2.66 pints is 95%.

This basic idea can be adapted for finding the appropriate sample size that will lead to a certain desired confidence interval. To illustrate, let us now suppose that a researcher is interested in estimating annual per capita consumption of specialty fruit/energy drinks for adults living in a particular area of the United States. The researcher knows that it is possible to take a random sample of respondents in the area and compute a sample mean. However, what the researcher really wants to do is be able to state with, say, 95% confidence that the population mean falls within some allowable interval computed about the sample mean. The researcher wants to find a sample size that will permit this kind of statement.

The Case of the Sample Mean

Let us first assume that the allowable error is 0.5 gallon of energy drinks per capita and the level of confidence is 95%. With this in mind, we go through the following checklist:

1. *Specify the amount of error (E) that can be allowed*. This is the maximum allowable difference between the sample mean and the population mean. $\overline{X} \pm E$, therefore, defines the interval within which μ will lie with some prespecified level of confidence. In our example, the allowable error is set at *E*, or 0.5 gallon per year.

2. *Specify the desired level of confidence*. In our illustrative problem involving specialty fruit/energy drink consumption, the confidence level is set at 95%

3. *Determine the number of standard errors (Z) associated with the confidence level*. This is accomplished by use of a table of probabilities for a normal distribution. For a 95% confidence level, reference to Table A.1 indicates that the Z value that allows a 0.025 probability that the population mean will fall outside *one* end of the interval is Z = 1.96. Since we can allow a *total* probability of 0.05 that the population mean will lie outside *either* end of the interval, Z = 1.96 is the correct value for a 95% confidence level.

4. *Estimate the standard deviation of the population*. The standard deviation can be estimated by (a) judgment; (b) reference to other studies; or (c) by the use of a pilot sample. Suppose that the standard deviation of the area's population for specialty fruit/energy drink consumption is assumed to be 4.0 gallons per capita per year.

5. *Calculate the sample size using the formula for the standard error of the mean*. One standard error of the mean is to be set equal to the allowable error (*E* = 0.5) divided by the appropriate *Z* value of 1.96.

$$\sigma_{\overline{X}} = \frac{E}{Z} = \frac{0.5}{1.96} = 0.255$$

   This will assure us that the interval to be computed around the to-be-found sample mean will have a 95% preassigned chance of being ±0.5 gallon away from the population mean.

6. Neglecting the finite multiplier, we then solve for *n* in the formula

$$\sigma_{\overline{X}} = \frac{E}{Z} = \frac{\sigma}{\sqrt{n}} \qquad \text{or} \qquad \sigma_{\overline{X}} = 0.255 = \frac{4.0}{\sqrt{n}}$$

Hence, $n \cong 246$ (rounded)

7. In general, we can find *n* directly from the following formula:

$$n = \frac{\sigma^2 Z^2}{E^2} = \frac{16(1.96)^2}{(0.5)^2} \cong 246$$

If the resulting sample size represents a significant proportion of the population, say 10% or more, the finite population multiplier is required and the sample size must be recalculated using the following formula, where *N* is the size of the population:

$$n = \frac{N(\sigma^2 Z^{2)}}{NE^2 + \sigma^2 Z^2}$$

The Case of the Sample Proportion

Suppose that, in addition to estimating the mean number of gallons of specialty fruit/energy drinks consumed per capita per year, the researcher is also concerned with estimating the proportion of respondents using one or more specialty fruit/energy drinks in the past year. How should the sample size be determined in this case?

The procedures for determining sample size for interval estimates of proportions are very similar to those for interval estimates of means. In this case the following checklist would be used:

1. *Specify the amount of error that can be allowed*. Suppose that the desired reliability is such that an allowable interval of $p - \pi = \pm 0.05$ is set; that is, the allowable error *E* is 0.05, or 5 percentage points.
2. *Specify the desired level of confidence*. Suppose that the level of confidence here, as in the preceding problem, is set at 95%.
3. *Determine the number of standard errors* Z *associated with the confidence level*. This will be the same as for the preceding estimation; *Z* = 1.96.
4. *Estimate the population proportion,* (π). The population proportion can again be estimated by *judgment*, by *reference to other studies*, or by *the results of a pilot sample*. Suppose that π is assumed to be 0.4 in this case; that is, the researcher assumes that 40% of the population used one or more specialty fruit/energy drinks last year.
5. *Calculate the sample size using the formula for the standard error of the proportion*. One standard error of the proportion is to be set equal to the allowable error (*E* = 0.05) divided by the appropriate *Z* value of 1.96.

$$\sigma_{\bar{P}} = \frac{E}{Z} = \frac{0.05}{1.96} = 0.0255$$

6. Neglecting the finite multiplier, we then solve for *n* in the following formula:

$$\sigma_{\bar{P}} = \frac{E}{Z} = \sqrt{\frac{\pi(1-\pi)}{n}} = 0.0255 = \sqrt{\frac{0.4(0.6)}{n}}$$

Hence, $n \cong 369$ (rounded)

7. In general, we can find *n* directly from the formula

$$n = \frac{\pi(1-\pi)Z^2}{E^2} = \frac{0.4(1-0.4)(1.96)^2}{(0.05)^2} \cong 369$$

Once again, if the resulting *n* is 10% or more of the population size, the finite population multiplier is required and the sample size can be computed from the following:

$$n = \frac{N\pi(1-\pi)Z^2}{NE^2 + \pi(1-\pi)Z^2}$$

Determining Sample Size When More Than One Interval Estimate Is to Be Made from the Same Sample

The usual case when collecting sample data for estimation of various parameters is that more than one estimate is to be made. The sample size for each of the estimates will usually be different. Since only one sample is to be chosen, what size should it be?

A strict adherence to the allowable error and the confidence levels specified in the calculation of the sample sizes for the individual estimation problems leaves no choice but to take the largest sample size calculated. This will give more precision for the other estimates than was specified but will meet the specification for the estimate for which the size of sample was calculated. In the specialty fruit/energy drink consumption problem, for example, the sample size would be 369 (the sample size calculated for estimating the proportion of users) rather than 246 (the sample size calculated for estimating the mean amount used). Remember, the sample sizes determined in this manner are for obtained samples. In order to determine the size of the original sample the researcher must estimate the rate of response expected. For example, if a mail survey was to be conducted and it was believed that only a 20% response would be obtained, a desired obtained sample of 250 would need an original sample of 1,250.

Devices for Calculating Sample Size

In practice, manual devices (paper nomographs) and online sample size calculators can be used to easily compute a sample size helpful in rough-guide situations where the researcher is not sure of either allowable error levels or population standard deviations. One such calculator is located at http://marketing.byu.edu/samplesizecalculator.html .

The Hypothesis-Testing Approach

As indicated earlier, sample sizes can also be determined (within the apparatus of traditional statistical inference) by the hypothesis-testing approach. In this case the procedures are more are more elaborate. We shall need both an assumed probability of making a type I error—called the alpha risk—and an assumed probability of making a type II error—called the beta risk. These risks are, in turn, based on H_0: the null hypothesis, and H_1: the alternate hypothesis.

In hypothesis testing the sample results sometimes lead us to reject H_0 when it is true. This is a type I error. On other occasions the sample findings may lead us to accept H_0 when it is false. This is a type II error. The nature of these errors is shown in Table 7.4.

A numerical example should make this approach clearer. We first consider the case for means and then the case for proportions. Before doing this, however, a few words on the relationship between the Type I and Type II errors are in order. The relationship between these two errors is an inverse. The ability of a sample to protect against the type II error is called statistical power.

Table 7.3 Sample Size When Estimating Population Mean and Proportion (Selected Samples)

A. Mean

Population size, N	*Reliability, r*	*Z-value, Z*	*Standard deviation, s*	*Precision, d*	*Sample size, n*
400	95%	1.96	1.0	± .25	53
400	90%	1.645	1.0	± .25	39
400	90%	1.645	s	± .25s	39
400	95%	1.96	s	± 1/3s	32
400	95%	1.96	1.0	± .33	32
400	99.7%	3.0	s	± 1/3s	67
200	95%	1.96	s	± 1/3s	30
1600	95%	1.96	s	± 1/3s	34
$N \to \infty$	95%	1.96	s	± 1/3s	35
$N \to \infty$	90%	1.645	s	± .25s	43

B. Proportion

Population Size N	*Reliability r*	*Z-value Z*	*Standard Deviation s*	*Precision d*	*Sample Size n*
500	95%	1.96	0.5	± 10%	81
500	95%	1.96	.3 or .7	± 10%	69
$N \to \infty$	95%	1.96	0.5	± 10%	96
$N \to \infty$	95%	1.96	0.5	± 5%	384
$N \to \infty$	90%	1.645	0.5	± 10%	68
$N \to \infty$	99%	2.58	0.5	± 5%	666
$N \to \infty$	99.7%	3.00	0.5	± 5%	900
$N \to \infty$	99%	2.58	0.5	± 1%	16,641
$N \to \infty$	99.7%	3.00	0.5	± 1%	22,500
400	90%	1.645	.2 or .8*	± 10%	39
200	90%	1.645	.2 or .8*	± 10%	36

*As the difference between *p* and 0.5 increases, the sampling distribution for *p* becomes more skewed and may deviate from the normal approximation. Thus, the data should be interpreted with care for small samples as *p* approaches either 0.0 or 1.0. (The corresponding percentage would be 0.0 or 100.0.)

SOURCE: Tatham, 1979, p.b.

Table 7.4 Types of Error in Making a Wrong Decision

Action	**H_o is true**	**H_o is false**
Accept H_o	No error	Type II error (β)
Reject H_o	Type I error (α)	No error

When the hypothesis is one of difference, a type II error occurs when what is really chance variation is accepted as a real difference. Taking statistical power into account often indicates that larger (and, thus, more costly) samples are needed. Sample size is affected by the effect size—the needed or expected difference. When it is large, say more than 15 percentage points, statistical power is usually not a problem. Very small effect sizes (e.g., two percentage points) require such large sample sizes as to be impractical; surveys cannot really reliably measure such small differences or changes. As an illustration, if a researcher wants to discern an effect size of, say, 10 points at a 95 % confidence level and a desired power to detect the superiority of one alternative over another of 80%, the approximate sample size needed would be 310. If the desired power is raised to 95%, the sample size needed would be about 540 (Semon, 1994).

The researcher and the manager as well, must consider the relative business risks when setting appropriate levels of protection against the two types of errors (Semon, 1990).

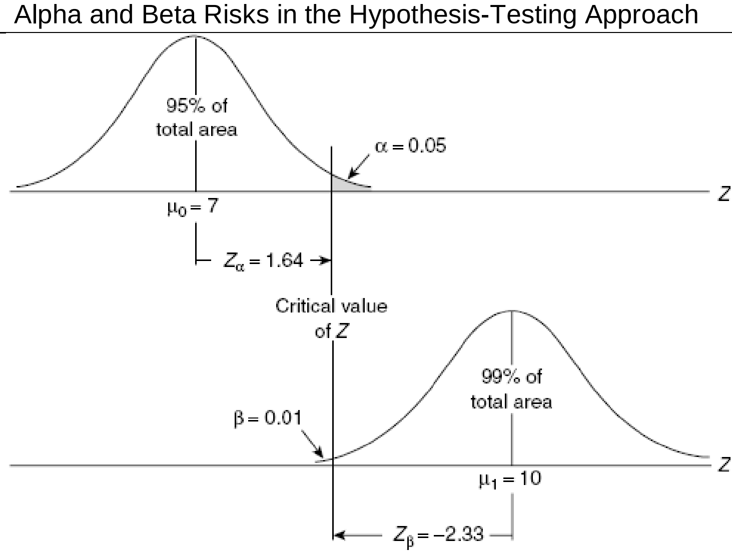
The Case Involving Means

As an illustrative example, let us assume that a store test of a new bleaching agent is to be conducted. It has been determined earlier that if the (population) sales per store average only 7 cases per week, the new product should not be marketed. On the other hand, a mean sales level of 10 cases per week would justify marketing the new product nationally. Using methods of traditional inference, how should the number of sample stores for the market test be determined? The procedures are similar to those for interval estimation problems but are somewhat more complicated. Specifically, we go through the following checklist:

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population means μ_0 and μ_1, respectively. (By convention, the null hypothesis is the one that would result in no change being made, if accepted.) In the bleach market-introduction problem, the values are set at H_0: μ_0 = 7 cases per week, and H_1: μ_1 = 10 cases per week.

2. Specify the allowable probabilities (α and β, respectively) of type I and type II errors. The type I error is the error of rejecting a true null hypothesis. The type II error is made when the alternate hypothesis is rejected when it is true. α and β are the allowable probabilities of making those two types of errors, respectively. They are shown graphically in Figure 7.3, where we assume that in the bleach-introduction problem the allowable probabilities of error are assigned as α = 0.05 and β = 0.01.

Figure 7.3 Alpha and Beta Risks in the Hypothesis-Testing Approach



3. Determine the number of standard errors associated with each of the error probabilities α and β. For a one-tailed test the *Z* values for the 0.05 and 0.01 risks, respectively, are found from Table A.1 in Appendix A to be Z_α = 1.64 and Z_β = 2.33. These are shown in figure 7.3. Note that in the figure we affix a minus sign to the value of Z_β since the critical values lies to the left of μ_1 = 10.

4. Estimate the population standard deviation σ. In the case of the new bleach the standard deviation of cases sold per store per week is assumed to be five cases.

5. *Calculate the sample size that will meet the α and β error requirements*. Because *two* sampling distributions are involved, a simultaneous solution of two equations is required to determine the sample size and critical value that will satisfy both equations. These equations are the following:

$$\text{critical value} = \mu_0 + Z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$\text{critical value} = \mu_1 - Z_{\beta\alpha} \frac{\sigma}{\sqrt{n}}$$

6. Setting the right-hand side of these two equations equal and solving for *n* gives

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

In the bleach problem the desired sample size is

$$n = \frac{(1.64 + 2.33)^2 5^2}{(10 - 7)^2} \cong 44 \text{ stores (rounded)}$$

Having solved for *n*, the sample size, we can then go on to solve for the critical value for the mean number of cases by means of the following substitution:

$$\text{critical value} = \mu_1 - Z_\beta \frac{\sigma}{\sqrt{n}}$$

$$= 10 - (2.33)\frac{5}{\sqrt{44}} = 8.24 \text{ cases}$$

Alternatively, we could find the critical value from the first of the two equations:

$$\text{critical value} = 7 + (1.64)\ \frac{5}{\sqrt{44}} = 8.24$$

The decision rule then becomes the following: Take a sample of 44 stores for the controlled store test. If the mean number of cases of the new bleach sold per week in the sample stores is less than or equal to 8.24 cases, do not introduce the product. If the mean number of cases of bleach sold per week is greater than 8.24 cases, introduce the product.

The Case Involving Proportions

For sample-size determination involving proportions, the following analogous steps are required:

1. Specify the values of the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population proportions π_0 and π_1, respectively.

2. Specify the allowable probabilities (α and β, respectively) of type I and type II errors.

3. Determine the number of standard errors associated with each of these error probabilities (Z_α and Z_β).

4. Calculate the desired sample size *n* from the formula:

$$n = \left[\frac{Z_\alpha \sqrt{\pi_0(1-\pi_0)} + Z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2$$

This formula is appropriate for relatively large samples ($n \geq 100$), where the normal distribution is a good approximation to the binomial. To illustrate its application, suppose that a researcher is interested in the true proportion of residents in a large city who would be willing to pay over $400 for a portable refrigerator-bar combination if it were commercialized.

Assume that the marketing researcher would recommend commercialization of the firm's refrigerator-bar combination if the true proportion of consumers who would pay over $400 for this class of goods is 70%. If the proportion is only 60%, the researcher would not recommend commercialization. The hypotheses are:

$$H_0\text{: } \pi_0 = 0.6$$

$$H_1\text{: } \pi_1 = 0.7$$

The alpha risk associated with the null (status quo) hypothesis is selected by the researcher to be 0.05 if the true proportion π is equal to 0.6. Moreover, the researcher is willing to assume a beta risk of 0.1 if the true proportion is equal to 0.7. With these assumptions it is possible to obtain the approximate sample size by using the preceding formula:

$$n = \left[\frac{Z_\alpha \sqrt{\pi_0(1-\pi_0)} + Z_\beta \sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0} \right]^2$$

where $Z_\alpha = Z_{0.05} = 1.64$, $Z_\beta = Z_{0.1} = 1.28$, $\pi_0 = 0.6$, and $\pi_1 = 0.7$. The solution follows:

$$n = \left[\frac{1.64\sqrt{0.6(0.4)} + 1.28\sqrt{0.7(0.3)}}{0.7 - 0.6} \right]^2 \cong 193 \text{ (rounded)}$$

Accordingly, in this example the sample size to take is 193. The critical value can be found analogously as follows:

$$\text{critical value} = \pi_1 - Z_\beta \sqrt{\frac{\pi_1(1-\pi_1)}{n}}$$

$$= 0.7 - (1.28)\sqrt{\frac{0.7(0.3)}{193}} = 0.658$$

In this case the decision rule is the following: Take a sample of 193 residents. If the sample proportion who would pay over $400 is less than or equal to 0.658, do not commercialize the refrigerator-bar combination. If the sample proportion exceeds 0.658, commercialize the product.

DETERMINING SAMPLE SIZE FOR OTHER PROBABILITY–SAMPLE DESIGNS

Thus far we have discussed only the determination of *simple* random-sample sizes using the methods of traditional statistical inference. How are the sizes for other types of random-sample designs—systematic, stratified, cluster, area, and multistage—determined?

The answer to this question is that the same *general* procedures are used to determine the overall sample size, but the formulas for the standard errors differ. The formulas become more complex and difficult to estimate as one considers stratified sampling, cluster sampling, or the other more elaborate designs. This is because the standard error for these designs is partially a function of the standard deviation (or proportion) of each stratum or cluster included in the design. For a multistage sample consisting of several strata in one stage followed by clusters in another and systematic sampling in a third, the standard error formula can become very complex indeed. And once the overall sample size is determined; it must be apportioned among the strata and clusters, which also adds to the complexity.

Appropriate formulas for estimating standard errors and sample sizes for other random-sample designs are available elsewhere (Sudman, 1976; Kish, 1965). In general, as compared with the size of simple random samples, systematic samples may be the same (for purposes of calculating the standard error, the assumption is typically made that the systematic sample is a simple random sample). Stratified samples are usually smaller, and cluster samples will usually be larger in size to provide the same reliability as a simple random sample. If used properly, stratification usually results in a smaller sampling error than is given by a comparable-size simple random sample. Consequently, the advantages of a stratified sample design over a simple random sampling design are as follows (Sangren, 2000, p. 67):

- For the same level of precision, one would need a smaller sample size in total, and this leads to a *lower cost*
- For the same total sample size, one would gain a greater precision for any estimates that are to be made

EVALUATION OF THE TRADITIONAL APPROACH

If one were to devise the ideal method of determining sample size, at a minimum one would want it to meet the criteria of being:

1. Logically complete
2. Adaptable to a wide range of sampling situations
3. Simple to use

If the traditional (Neyman-Pearson) approach to sample-size determination were to be rated on these criteria, the rating would be low for logical completeness and high for both adaptability and simplicity.

The traditional approach is logically incomplete because sample size is specified as being a function only of the conditional probabilities of making errors. The conditional costs of wrong decisions, prior probabilities, nonsampling errors, and the cost of sampling are not considered in the model. More advanced texts, however, do consider traditional sample-size determination by means of formulas that include the costs of sampling (see Cochran, 1963, Chapter 4).

The fact that these variables are excluded implies that somehow they must be taken into account outside the model. However, the only way that accommodation can be made is through adjustment of either the specified confidence level or the assigned alpha and beta risks.

SUMMARY

One of the most difficult problems in research design is the one concerned with the size of sample to take. We discussed the determination of sample size from the standpoint of traditional inferential methods and all aspects of sample size determination were covered. Estimation and hypothesis-testing applications were discussed. We concluded the chapter with an evaluation of this traditional approach to determining sample size.

REFERENCES

Cochran, W. G. (1963). Sampling techniques (2nd ed.). New York: Wiley.

Deming, W. E. (1960). Sample designs in business research. New York: Wiley.

Dillman, D. A. (2000). Mail and internet surveys: The tailored design method (2nd ed.). New York:Wiley.

McDowell, Elisa (2002), "Inside the paper's election polls", The Charleston Post and Courier Accessed August 21, 2009 http://www.accessmylibrary.com/coms2/summary_0286-2181489_ITM

Kish, L. (1965). Survey sampling. New York: Wiley.

National Center for Health Statistics (2009), Data: Cell Phone-Only Households State-by-State, Associated Press, as reported on Fox News, March 11, 2009 http://www.foxnews.com/story/0,2933,508926,00.html;

http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200905.htm

http://marketing.byu.edu/samplesizecalculator.html

http://www.qualtrics.com/wiki

Sangren, S. (2000, April). Survey and sampling in an imperfect world: More second-best solutions. Quirk's Marketing Research Review, pp. 16ff.

Semon, T. T. (1990, January 8). Don't panic when sample isn't quite what you expected. Marketing News, 24, 31.

Semon, T. T. (1994, January 3). Save a few bucks on sample size, risk millions in opportunity loss. Marketing News, 28, 19.

Semon, T. T. (2000, June 9). Quality sampling, research means avoiding sloppiness. Marketing News, 34, 11.

Sudman, S. (1976). Applied sampling. New York:Academic Press.

Survey Methods Centre (1995). Quota versus probability sampling. Survey Methods Centre Newsletter, 15(1).

Tatham, E. (1979). A report on sample size selection. Shawnee Mission, KS: Customer Satisfaction Institute.

Unlisted phone numbers on the rise (1989) http://findarticles.com/p/articles/mi_m3065/is_n8_v18/ai_7812479/

Zinkhan, G., Burton, S., & Wallendorf, M. (1983). Marketing applications for snowball sampling. In W.R. Darden et al. (Eds.), Research methods and causal models in marketing. Chicago: American Marketing Association.

This page is intentionally blank