2002

Sampling Concepts

Paul Boyd, Ph.D.
Johnson & Wales University - Providence, Paul.Boyd@jwu.edu

Sampling Concepts©

Paul Boyd, Ph.D.

The usefulness of any research is dependent upon how well the group studied represents the group about which decisions are to be made or conclusions drawn. That is, it depends upon how well the *sample* reflects relevant characteristics of the *population*. When it is possible to study every member of that group there is no problem, for on these occasions we can easily calculate the exact attribute (*parameter*) of interest for our population.

For example, if we were interested in determining the average number of gallons of gasoline sold to customers at our service station yesterday, we could add up the total number of gallons sold yesterday and divide that by the total number of gas purchases. The result of this *census* would be the exact average number of gallons of gasoline sold per customer (yesterday).

But when it is not possible to study each and every member of the target population, researchers and analysts must estimate characteristics of the group based upon the study of a relatively small number of elements. This smaller group is referred to as a sample, and the process of selecting the individuals who will be included in the sample is called *sampling*.

If we wanted to determine the average number of gallons of gasoline sold to customers over the last year, we would be unlikely (or unwilling) to perform a census of all gasoline sales over such a long period. It would be too time consuming and/or costly to find and drag out the records for each and every day of an entire year and make the calculations.

In this situation we would be more likely to take a sample from among all of the gasoline purchases made during the prior year. One simple method of taking this sample would be to take a random group of, say, 50 days from the past year. The number of gallons sold for each sale over each of the 50 days could then be added together and divided by the total number of individual sales over those 50 days.

The resulting statistic is called the sample mean. The sample mean (along with other information) may then be used to estimate the average number of gallons per purchase for all gasoline purchases during that year. This process of estimating a population parameter from a sample statistic is called *inferential statistics*.

This paper deals with issues surrounding the process of determining the appropriate sample for a research study and the factors you will need to check before you are able to conclude that the sample used is appropriate for both the analysis performed and the management decision you need to make.

I. Populations and Samples

Research is conducted in order to draw conclusions about population parameters: How likely is our target population to purchase our product if we include these features? Or, how satisfied are our customers with our products and services?

A key aspect of making sense out of the research process is to understand the differences between *populations* and *samples*. A population refers to **all** of the entities (orders, stores, customers, sales, units, etc.) about which we want to draw conclusions or make decisions. When researchers measure every element of the population, the study is called a *census*. For various reasons (primarily the time and cost), it is often not feasible to conduct a census. In most situations, a smaller subset of the population is selected for study. This smaller subset is called a *sample*. When a sample is used, only the individual members of the sample are measured.

Why not conduct a census; a study of every individual in the population? There are three factors that limit the ability to conduct a census in research situations:

1.) The individuals in the population may be impossible to identify as a group. For example, there is no list of all individuals over the age of 17 who visit shopping malls at least twice a month. Nor is there a list of all individuals who are thinking about purchasing a 4G telephone.

2.) Even if it were possible to identify all the individuals in a population, the cost of collecting information from them all may be prohibitive. The American Automobile Association may have contact information for all of its 44 million members, but the costs of administering a satisfaction survey to each and every one of them would be astronomical.

3.) Some research destroys the functionality of the item being studied. For example, a study of alkaline battery life would degrade all batteries tested to the point where they were not salable.

In these types of situations it is much more feasible to study a small portion of the population and then use statistical analysis to draw inferences about the population.

If your business is one that has a limited number of customers (say 50) and you are interested in learning their perceptions regarding your customer service, you may want to conduct a census of their opinions by calling each and every one of them and asking. However, if the number of customers you have is quite large, contacting each and every one of them would be impractical. In these situations, you would study only a portion of the total customers; a sample.

The classification of the group as either a population or a sample will depend on the nature of the research and what is being studied. For example, the group of company stores that are located in Connecticut represent the population of stores in that state. However, that same group of stores may comprise a sample for a study of all stores in New England (the population); although it may not be a very "good" sample.

The majority of research is the process of drawing inferences about the population from a sample. Inferential statistics are the quantitative methods (and their results) that allow these inferences.

Both populations and samples have characteristics that we are interested in learning about; that is, measuring. When the group is a population, the measure is called a *parameter*. The measure of a sample is called a *statistic*. Inferential statistics are processes that allow us to draw conclusions about *population parameters* through analyses of *sample statistics*.

For example, we may be interested in estimating the percentage of all of our customers who are satisfied with our services (the parameter), through the analysis of the percentage of a sample of 500 of our customers who are satisfied with our services (the statistic).

Or we may be interested in estimating the average amount of liability insurance carried by all of our franchisees (the parameter) through the analysis of the average amount of liability insurance of 200 franchisees who responded to a survey (the statistic).

Note that both the parameter and the statistic refer to similar concepts. In the first example, they both refer to "the percentage of the group that is satisfied with our services." In the second example, both the parameter and the statistic measure the average amount of liability insurance carried; however, one measures it for the population, the other for the sample.

In order to make it easy to understand whether a parameter or a statistic is being discussed, researchers and statisticians use different symbols to keep them straight. Generally, lower case Greek letters are used to refer to population parameters (e.g., μ, π, σ) and a combination of symbols and letters to refer to sample statistics – $\bar{X}$ (or "x bar"), $\hat{p}$ (or "P hat"), s. Table A presents the research & statistical symbols for the most common characteristics that are measured.

Table A

Symbols for Parameters and Statistics

Measure	Population Parameter	Sample Statistic
Proportion	π (pi) or P*	$\hat{p}$ (p-hat)
Average (Arithmetic Mean)	μ (mu)	$\bar{X}$ (x-bar)
Variance	σ^2 (sigma squared)	s^2 (s squared)
Standard Deviation	σ (sigma)	s or s.d.
Group size	N	n

* Due to possible confusion with the ratio of the circumference of a circle to its diameter (i.e., 3.14159265) which is occasionally used in statistical formulas, the symbol π is generally avoided in research and analysis. *P* is preferred for denoting population proportions.

Whenever you see the symbols in the center column of Table A (Population Parameter) in a formula or a report, a characteristic of the population is being described. When you see the symbols in the right hand column (Sample Statistic), the sample is being discussed. Thus, "N" refers to the number of elements in (the size of) the population, while "n" refers to the sample size.

Despite that fact that parameters and statistics measure similar characteristics, they ARE NOT EQUAL TO ONE ANOTHER; that is, they will not have the same numerical value as one another ($\bar{X} \neq \mu$). They may be close, but it would be wild luck if they were the same, for they do not measure the same thing.

For a descriptive example, suppose a population of 25 orders (N) were placed with a telephone sales business on one day and we were interested in determining (or estimating) the average dollar value of the orders that day (μ). A sample of 10 of these orders (n) may show us that the average order dollar value of the sample was $75 per order ($\bar{X}$ = $75).

This is NOT the average of all 25 total orders for that day. It may be "close" to the average for the entire day or it may not be too close. But in any event, there is little likelihood that it is exactly the same as the population average. If the sample were larger, say 20 orders, the sample average is likely to be closer to the population average, but it still will not be equivalent to it.

Now, think of trying to estimate the number of annual business trips per member for a population of 1.8 million frequent traveler club members from a sample of 300. The sample mean is extremely unlikely to be exactly the value population average.

The question is, how likely is the sample mean to be how close to the true population mean? The answer is largely a function of the size of the sample (n) and the amount of variability in the population (see section IV. below).

A "good" sample is one that accurately represents the population. When the population has diverse characteristics – what researchers call *heterogeneity* – it is important that the any sample used reflects that diversity.

II. Biases Associated with Sampling

Bias is the difference between the truth and how the truth is represented. That is, bias is the difference between what something is, and what it is reported to be. It is easy to see that bias can take on many forms during a research project. And it is the responsibility of effective researchers to guard to the best of their abilities against all of them.

In the sampling process, the extent to which the sample has the same characteristics as the population is the extent to which the sample is accurate as a tool for estimating characteristics of the population and conducting tests of hypotheses. There are two primary sources bias when sampling is undertaken: systematic bias (response bias) and random bias (sampling error). Some bias cannot be eliminated, but it may be controlled.

A. Random Sampling Error

The difference between the *sample statistic* and the population characteristic that it measures (the *population parameter*), is called *sampling error*. The unfortunate aspects of sampling error are that 1) it is always present and 2) you cannot know how much of it there is. However, given various statistical theorems (i.e., the *Central Limit Theorem*), you can calculate the likely *range* of the difference.

As a general rule, sampling error can best be reduced by increasing the sample size. See Section IV below for a comprehensive discussion of sample size.

B. Self Selection Bias

Response bias is a term that is used to describe a situation where respondents do not answer truthfully the questions asked. While this situation is problematic throughout the research process, it is often present during the sampling process.

When individuals choose to participate in research, they may do so based, to some extent, upon their interest or association with the subject being studied. Those who are most interested are most likely to participate; those who are least interested the least. Anytime some respondents can be expected to have a strong opinion – either positive or negative – they can be expected to respond at a higher rate than those who do not maintain that level of interest.

Likewise, some individuals may incorrectly report certain demographic characteristics in the belief that by doing so their opinions will be heard. For example, some individuals will incorrectly report their income in order to participate in research about a high end product.

Also, individuals are reluctant to be perceived as not socially responsible and will tend to be untruthful about past behavior that they believe will put them in a bad light. For example, the proportion of individuals randomly contacted who report that they voted in the previous election always far outweighs the proportion who actually did. Not voting is viewed as socially irresponsible.

III. Who to Use in Your Sample

A sample must be representative of a population if it is to be useful for drawing inferences about it. That is, the sample must have the same characteristics as the population; else the conclusions drawn will be *biased* – not necessarily true for the group of interest. If you are interested in learning what "early adopters" think about your product, you don't want to include "technophobes" in your sample because technophobes, by definition, are not early technology adopters.

Similarly, if you are trying to draw conclusions about all of your customers as a group, then including only those customers within a 50 mile radius of our headquarters in a sample will likely result in inaccurate results. In this situation, differences in regional demographics or climates could lead you to draw conclusions about your entire population that are only true for a subset of that population. The best method for ensuring that a sample has the same characteristics as the population is to make sure that every member of the population has an opportunity to be included in any sample.

A. Random Sampling

Random sampling, also known as "probability" sampling, is the preferred method to ensure that all members of the population have the opportunity to be included in the sample and, thus, that the sample is truly representative of the population.

In its simplest form, each member of the population is assigned a unique, preferably sequential, number and then random numbers (between 1 and the population size) are drawn or assigned until the required sample size is achieved.

For a quick example, let's assume that you have 5000 customers and you want a sample of 600. Each of your 5000 customers would be assigned a unique number from 1 to 5000. Then, 600 random numbers between 1 and 5000 would be selected (possibly through the use of the RAND or RANDBETWEEN functions in Microsoft Excel). The 600 customers with those numbers would be included in the sample and invited to participate in the research.

Alternately, all 5000 could be assigned a random number between 0 and 1, and the those with the lowest (or highest) 600 would be included in the sample.

If you want to ensure that you have sufficient representation of two (or more) segments of your customers you could separate them into segments (groups) and then randomly select from among each group. This process is called *stratified random sampling*.

If, for example, you were interested in the differences between urban, suburban, and rural customers (and still wanted your total sample to be 600), you could separate your customers into these three groups and then randomly select 200 from each.

While random sampling is the best way to ensure that the sample is representative of the population, it is not a perfect solution. The process of randomly drawing a sample does not ensure that the sample is unbiased – that the sample does not over- or under-represent a characteristic of the population – but it goes a long way toward ensuring that the population has a good chance to be fairly represented.

It is not always possible to identify all the members of a target population in order to perform a random selection. There is not, for example, a list of all men who use disposable razors, or people who will buy coffee in Dunkin Donuts stores in the next six months, or the readers of USA Today, or all SUV drivers. In these situations, it is impossible to ensure that all the members of the population have the opportunity to be included in the sample. Consequently, true random sampling is not possible.

When it is impossible to identify every member of the population, the research will be biased. While this is unfortunate, it is not a reason to abandon the research.

The research may proceed with a biased sample so long as two conditions are met:

1) The sample utilized should be as random as is possible, and
2) The extent of the non-randomness should be discussed with the client (and mentioned in any report).

If true random sampling is not possible, researchers must make every effort to ensure that the sample is as representative of the target population as is feasible.

B. "Not Quite Random" Samples: Sampling Frames & Panels

When it is not possible to identify all the members of the "population," it will be difficult to ensure that every member has an opportunity or chance to be included in the sample. In theses cases – such as when the target population of interest is "Cell telephone users" – researchers are responsible for doing their best to ensure that whatever sample is used is selected from a group that, prime facie, maintains the heterogeneity of the target population.

Groups that contain identifiable members of a target population are called *sampling frames*. Sampling frames take on many forms, but the key attribute of such a frame is that researchers have the ability to randomly select from among its members. In general, the larger the number of members of the frame, the more likely that the frame will reflect the heterogeneity of the target population.

The key to assessing the usefulness of any sampling frame is to determine the extent to which the members of the frame represent the *heterogeneity* of the target population. Note that it is possible to use *screening questions* to narrow the sampling frame down in order to make the sample itself more consistent with the target population. However, if the frame is biased to begin with, screening questions will not be able to help.

All sample frames contain at least some *bias.* To some extent; they all fail to represent the population as a whole. It is a useful exercise to try to understand the differences between the members of the target population who appear on the list and those who do not. Typically, the reasons why or why not an individual appears on the list or in the sampled group will indicate the level of bias likely to be encountered.

Sampling frames may take on a wide variety of forms depending on the population of interest:

- <u>Membership lists</u>: Many organizations, representing individuals with specific interests (e.g., high school teachers, sports car enthusiasts, architects, etc.) maintain a list of their members. These organizations may grant access to their lists (for a fee).
- <u>Telephone directories</u>: Given the high penetration of telephones into households, telephone books have a long history as a source of sampling frames for research. If telephone listings are to be used the population of

interest should be geographic based. Of recent concern, however, is the trend to abandon traditional land-based telephones for mobile telephones, for which there are no directories. Additionally, the lack of ability to contact others with transportable telephone numbers in other forms (VOIP) contribute to the some differences between members of directories and the general population.

- <u>Customer lists</u>: Many businesses maintain lists of current and past customers. These lists may be generated in a number of ways; frequent shopper sign up, warranty card registrants, help desk users, etc. While these lists rarely contain all customers, they may sometimes be used as a sampling frame for the entire group.
- <u>Subscribers to a free service or magazine</u>: The process of signing up for a free service often results in the provision of contact information that may be used to generate a sampling frame. Those who have signed-up for a free subscription to a magazine or professional/industry journal, have provided two useful bits of information to researchers; that they are members of a specific population (those with interest in the subject matter) and contact information.

One source of these lists is a profession known as *List Brokers*. List brokers generally maintain lists that are used for marketing purposes (typically, direct marketing). The information they maintain is often simply the name and address of individuals with specific characteristics. Recently, email addresses have been added to these lists. Due to the advent of the "Do Not Call" registry, telephone numbers have become less important to these lists.

In specific situations, the use of a marketing list such as those provided by list brokers may be useful in establishing a sampling frame for business research.

Perhaps the most common source of sampling frames is what is known as a *Research Panel*. A research panel is a collection of identifiable individuals who have consented to participate in research and are contacted from time-to-time to express their opinions on various subjects. To maintain industry ethical standards, panel participants typically *Opt-in* to the panel; that is they actively assert that they are willing to participate in the research. Panel members are often compensated for the time that they spend participating in research studies, although that compensation may be the simple inclusion in a drawing for a cash prize. At the time that panel members opt-in to the panel, they provide basic *demographic* information so that they can be segmented into appropriate sub-populations for research.

Initially, panels were developed for mail and telephone surveys, but the growth of the Internet has resulted in the proliferation of Web-based panels that are easily accessed. Panels are big business in that they can provide easy access to individuals with unique attributes (and therefore, members of unique populations). The most common research panel is a *Consumer Panel*, with members of the general buying public. Consumer panels can be quite large; Harris Interactive, for example, maintains a consumer panel with more than 6 million unique email addresses. Other firms also maintain large panels: NPDOR and Greenfield Online.

There are also *B2B (Business-to-business) panels* that are used for research into business practices. B2B panels typically contain purchasing decision makers.

A growing trend in research is the development of *Custom* or *Specialty Panels*. These panels maintain lists of individuals with specific characteristics who are willing to participate in research. As such, these panels may represent an excellent sampling frame for a specific population of interest. Examples of specialty panels include:

Physicians	Golfers
Lawyers	Corporate buyers
Affluent families	CIOs
Ethnic identity (e.g., Latinos)	Technology "Early Adopters"
Extreme sports enthusiasts	Business travelers

Specialty panels have been developed by large research organizations for sale to other organizations and also by individual companies seeking unique insights into their stakeholders; i.e., customers and potential customers. Wired magazine, for example, maintains a panel of its readers that it surveys from time to time to assist in the development of editorial policy and to provide data for articles. Turnkey Sports & Entertainment, in collaboration with Greenfield Online, has developed a specialty panel of fans of the New York Mets who may be tapped to provide opinions and perspectives of interest to Mets management.

There are some potential drawbacks to the use of research panels that should be understood before using one as a source of data. First, panel members are different from the general public, sometimes in ways that we can't measure. For one thing, panel members actually consent to spend time answering questions about various topics, while most people would not. It is unknown if the cause of this difference appears in other attitudes (say the ones being measured by a survey).

Additionally, despite some very high initial interest in participation in surveys, panel members tend to lose interest and stop responding to invitations to participate, particularly if they fail to win one of the cash drawings. Only a hardcore group of respondents is likely to continue responding. Again, it is unknown how these hardcore respondents differ from the general public, but there is no doubt that they do. It may or may not be important to your specific research.

C. Non-random Sampling

There are many situations where random sampling or even the use of a formal sampling frame is impossible. It may be unlikely that there exist any collection of names of individuals in specific populations that may be of interest to your specific research. There is not, for example, a list of single male parents of children (ages 5 to 8) who shop in Malls on weekdays. When research is required of such a population, a sample must still be identified so that data may be collected.

There are two basic methods for dealing with this situation. The first is to ask a very large number of potential respondents, say, from a panel, if they have the requisite attributes, and then screen out those who do not. This may result in a large number of individuals who are screened out of the survey, which decreases the likelihood that those screened out will participate in research in the future.

The second method of sampling in these situations is called *non-random sampling*, because these methods do not give everyone in the relevant population the opportunity to be included in a sample. Non-random sampling can be performed in a number of ways. But, the extent to which some randomness can be maintained will go a long way to ensuring that bias can be minimized.

In some cases a sample may be selected based upon tangible evidence that a subject has the requisite characteristics. In these situations, someone would have to make a determination that an individual was likely to be a member of the requisite population and solicit them to participate in the research. In the example above, a researcher could look for unaccompanied men with children at a mall between Monday and Friday and invite them to participate in the research. This type of sampling is called both *judgment sampling* (for the selection process) and *intercept sampling* for the public location of the process.

In other cases, a sample will be taken on the basis of the availability of individuals who are at least members of the population, even if the full range of diversity in the population is not present. Sampling of this nature is called *Convenience Sampling*. When professors use students as a sample or a business uses people who walk by the storefront as a sample frame are both examples of convenience sampling.

D. Adding Diversity to a Non-Random Sample

Recall that the purpose of random sampling is to ensure that any diversity in the population is reflected in the sample. When non-random sampling processes are employed, the researcher should make a serious attempt to ensure that any known diversity in the population is included in the sample.

For example, if the population of interest is shopping mall shoppers in Pennsylvania, then sampling in a single mall in King of Prussia will unlikely yield a sample that includes all of the diverse perspectives of shoppers throughout the State. In this case, the sample would be drawn from a number of different malls across the state, making sure to include malls that were visited by individuals with a variety of socio-demographic and economic backgrounds.

Additionally, since it is likely that the characteristics of mall shoppers differ from day to day (say weekends vs weekdays) and hour to hour (mornings, afternoons and evenings), it would be appropriate to take any sample from a mall across the entire time period during which it is open. If the sample selected comprised only the first

100 shoppers in the mall on a Monday morning, then many important perspectives would likely be missed (i.e., the attitudes of those who had to work in the morning).

There are other conditions that should also be considered when selecting such a sample. Are the characteristics of mall shoppers different on rainy or snowy days than they are on sunny days? Are there seasonal differences that may prove important?

To the greatest extent possible – within the time and budgeting requirements of the project – all known sources of differences in the population should be considered and accounted for when a sampling plan is selected. When these known sources of diverse opinions cannot be included in a sample, the researcher should identify the mismatch between the sample and the population and explain any likely consequences of the differences.

E. Screening Questions

Screening questions are questions that appear at the front of a survey that are used to filter out those who are not appropriate subjects for the research. That is, they screen out those who are not members of the target population. Screening questions could also be used to identify individuals with a specific characteristic that may be of interest to the research. They are also effective for controlling self selection bias among potential respondents.

For example, a screening question could be used to identify coffee drinkers from a broader sampling frame. Effective screening questions are worded in such a way that the respondent should not be able to determine which response will result in inclusion in the survey and which will result in *termination*, thus controlling self selection bias.

Screening questions are an excellent way to fine tune a sampling frame. However, researchers should be aware that, if the rate of inclusion is expected to be low (say, if the desired characteristic is recreational water skiers), a very large number of individuals will be screened out. Consequently, a very large number of individuals will need to be invited.

As mentioned above, one downside to using screening questions is the negative effect being screened out of a survey has on panel members. The more often one is screened out of a survey, the less likely respondents are to participate in subsequent studies. This will have the effect of adding bias to future studies using the panel.

F. More Sampling Issues

Control and Experimental Groups

Experimental or "*Causal*" studies are often used to establish extent to which a relationship exists between two or more variables. When experiments are conducted, researchers are typically trying to assess the impact of a change in the value of an *independent variable* on the value of a *dependent variable*.

For example, they might be interested in determining if a change in the corporate logo (independent variable) will lead to an improved image for the organization (dependent variable).

When investigators test such a change in a research setting, they study two groups to see if there is any difference between them. The first group – the *experimental group* – observes or experiences the change in the independent variable. The second group – the control group – does not experience the change in the independent variable (at least during the course of the experiment).

In this example, the experimental group would be shown the new logo and measured as to their perspectives on the new logo. The control group would not be shown the new logo, but they would be asked about their perspectives of the original logo.

In these situations, the researchers must insure that the two sample groups are 1) randomly selected, such that each reflects the characteristics of the broader population, and 2) of sufficiently large size for statistical comparison.

Cluster Sampling

In certain situations, the costs associated with conducting a random sample of the entire population may be prohibitively expensive in terms of either time or resources. However, if the population falls into easily identifiable groups – "clusters" – then it may be possible to randomly select a number of clusters and then measure all (or a random sample) of the members of the selected clusters.

This was the process described at the beginning of the chapter, where the gas station could randomly select a number of days throughout the year and then measure all of the gas purchases on those days (to estimate the average number of gallons purchased per customer for the year). In this case, the days selected represented clusters of customers.

Other examples of clusters are individual shopping malls across the country or *test markets*. The key to the effective use of cluster sampling is to ensure that the clusters selected do not systematically exclude any segments of the population of interest.

Multiple Effect Sampling

When the study includes multiple experimental groups or multiple treatments for the subjects, there would need to be sufficient sub-sample sizes to permit the accurate assessment of all combinations. These designs go by names such as Latin Squares designs, Factorial designs, and Randomized Blocks. These methods are attempts to randomly allocate the sample groups to account for all possible combinations of independent variables.

IV. Sample Size

The *sample size* is the number of objects (people, stores, orders, receipts, etc.) that are measured in order to draw conclusions about the broader population. How large a group should be studied? How many objects should be measured to make the research definitive?

As a general rule, the larger the sample size, the more accurate the estimate of the population parameter. However, there are some practical limits to this process or else every study would be a *census*, the measure of every member of the population.

The appropriate sample size should be determined based upon the level of accuracy desired of the estimation of the population parameter. The accuracy of the estimation of the population characteristic (parameter) may be increased by increasing the sample size.

When testing hypotheses, the sample size is an important component of the concept of *statistical power*, the ability of the research and analysis to detect differences or changes when they actually exist. The larger the sample size, the higher the statistical power of the study.

It may be important to monitor the statistical power of a study. Statistical Power can range from 0 to 1. While there are no objective criteria for determining an acceptable level of statistical power, there are two general rules of thumb:

1.) If the statistical power of a the research is .50 or less, then the study will most likely fail to detect any differences that actually exist (which calls to question the rationale for doing the study in the first place).

2.) If the statistical power is .80 or higher, then the research has a reasonable expectation of being able to identify differences (if they, in fact, exist) [Murphy & Myors, 2004; p.18].

So? How can we tell if the sample employed is large enough to draw accurate conclusions?

The formal method for determining the appropriate sample size is dependent on the value of three measures:

1. The confidence level. (For practical purposes, this is the probability of being correct; drawing the correct conclusion.) While any confidence level from 0 to 99.9999% is possible, most business researchers use the 95% confidence level. If the confidence level increases, the sample size will need to increase.

2. The highest level of error that the study will risk. This is the amount above or below the sample statistic ($\overline{X}$ or $\hat{p}$) that would still result in an acceptable confidence interval. If the allowable level of error decreases, the sample size will need to be increased.

3. For *metric variables*, the variability of the underlying population. Since this is a parameter, it is rarely known. When the population standard deviation is not known, the sample standard deviation (say, from a previous sample) may be used as an estimate if the sample used had more than 30 observations (n>30).

   For *categorical variables*, the third required figure is an estimate of the proportion of the population with the characteristic being studied. Again, this is a parameter and must be estimated, typically by the sample proportion. When this proportion cannot be accurately estimated, this proportion is set to .50, as this will yield the largest possible required sample size and cover all contingencies.

   As the variability of the underlying population increases (or as the population proportion approaches .50), the sample size will need to increase.

There are various formulas for calculating the required sample size based upon whether the data collected is to be of a categorical or quantitative nature (e.g. is to estimate a proportion or a mean). These formulas require the knowledge listed above: the variance or proportion in the population and a determination as to the maximum desirable error, as well as the acceptable Type I error risk (e.g., confidence interval).

But why bother with these formulas? It is possible to use one of them to construct a table that suggests the optimal sample size – given a population size, a specific margin of error, and a desired confidence level. This can help researchers avoid the formulas altogether. Table B presents the results of one set of these calculations. It may be used to determine the appropriate sample size for almost any study.

For most purposes, the first column within the table should suffice (Confidence Level =0.05, Margin of Error = 5%). Simply determine the size of the population down the left most column (use the next highest value if your exact population size is not listed). The value in the next column is the sample size that is required to generate a confidence interval of $\pm$ 5% for any population proportion. Thus, if you have 5000 customers and you want to sample a sufficient number to generate a 95% confidence interval that predicted the proportion who would be repeat customers within plus or minus 5%, you would need responses from a (random) sample of 357 of all your customers.

Table B
Required Sample Size†

Population Size	Confidence = 95%				Confidence = 99%			
	Margin of Error				Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	196	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1067	427	636	827	1119
1,500	306	515	759	1297	460	712	959	1376
2,000	322	563	869	1655	498	808	1141	1785
2,500	333	597	952	1984	524	879	1288	2173
3,500	346	641	1068	2565	558	977	1510	2890
5,000	357	678	1176	3288	586	1066	1734	3842
7,500	365	710	1275	4211	610	1147	1960	5165
10,000	370	727	1332	4899	622	1193	2098	6239
25,000	378	760	1448	6939	646	1285	2399	9972
50,000	381	772	1491	8056	655	1318	2520	12455
75,000	382	776	1506	8514	658	1330	2563	13583
100,000	383	778	1513	8762	659	1336	2585	14227
250,000	384	782	1527	9248	662	1347	2626	15555
500,000	384	783	1532	9423	663	1350	2640	16055
1,000,000	384	783	1534	9512	663	1352	2647	16317
2,500,000	384	784	1536	9567	663	1353	2651	16478
10,000,000	384	784	1536	9594	663	1354	2653	16560
100,000,000	384	784	1537	9603	663	1354	2654	16584
300,000,000	384	784	1537	9603	663	1354	2654	16586

Should more precision be required (i.e., a narrower Margin of Error) or greater confidence desired (0.01), the other columns of the table should be employed. For example, if you have 5000 customers, but you wanted the Margin of Error to be within plus or minus only 2.5% (with a .05 confidence level), you would need responses from a (random) sample of 1176 of all your customers.

As you can see, using the table is much simpler than employing a formula.

Professional researchers typically set a sample size level of about 500 to optimally estimate a single population parameter (e.g., the proportion of likely voters who will vote for a particular candidate). This will construct a 95% confidence interval with a Margin of Error of about ±4.4% (for large populations).

Since there is an inverse relationship between sample size and the Margin of Error, smaller sample sizes will yield larger Margins of Error. For example, a sample size of only 100 will result in a 95% confidence interval with a Margin of Error of almost ±13%, too large a range for estimating the true population proportion with any accuracy.

All of the sample estimates discussed present figures for the largest possible sample size for the desired level of confidence. Should the proportion of the sample with the desired characteristic be substantially different than 50%, then the desired level of accuracy can be established with a smaller sample. However, since you can't know what this percentage is until you actually ask a sample, it is wisest to assume that it will be 50% and use the listed larger sample size.

The number of sub-groups (or "comparison" groups) is another consideration in the determination of a sufficient sample size. Since the parameter must be measured for each sub-group, the size of the sample for each sub-group must be sufficiently large to permit a reasonable (sufficiently narrow) estimation.

For example, if we are interested in learning about the satisfaction of our customers, regardless of where they reside, we could take a random sample of *500 from all of our customers and be reasonably certain that we could draw* accurate conclusions about the overall level of satisfaction. On the other hand, if we wanted to determine if customers who patronize our New York stores are more satisfied than our customers in California, we would need a sample of <u>500 from each group</u> (New York and California) or 1,000 total in order to draw reasonably precise conclusions about the differences.

Note: in order to calculate the required sample size when the parameter to be estimated is the mean (average) as opposed to the proportion, a different set of calculations would need to be employed. Unfortunately, these calculations require an estimate of the population standard deviation to be effective. Since this is generally impossible (since it requires knowledge of the population mean; the parameter to be estimated), the resulting calculations tend to become hypothetical exercises, rather than accurate estimates. If a reliable estimate is not known, it is suggested that the research use Table 5-B for a 3.5% margin of error as an estimate of the required sample size.

If the population standard deviation IS known, use the formula in the Appendix to calculate the require sample size to estimate a population mean.

A. Quotas

There may be times when one of the sub-populations to be studies is relatively rare. In such a case, a random sample would yield either an insufficient number for analysis from the first group, or, more likely, far too many for the other groups.

For example, a situation could arise where an organization was interested in comparing the attitudes of racquet ball players to those who play tennis. Since the incidence of tennis players is significantly higher than that of racquet ball players, a random sample of racquet sports enthusiasts would yield significantly larger numbers of tennis players. If the difference in incidence is four to one, then, randomly, 400 tennis players would be surveyed for every 100 racquet ball players.

If the sample size required for analysis is determined to be 200 for each sub-population, then there will be 800 tennis players in the sample by the time 200 racquet ball players are selected. That is 600 more than is needed for the analysis.

Since there are costs associated with acquiring a sample (list purchase, incentives, etc.) the researcher would want to limit the unnecessary exposure to these costs by enforcing a quota of, say, 200 on the number of tennis players included in the sample. Once 200 tennis playing respondents had been measured, subsequent tennis players (typically determined by *screening questions* early in a survey) are politely excluded from the measurement process.

Establishing quotas is a useful practice to reduce costs while ensuring adequate sample sizes for analysis. Some caution should be employed however when analyzing data that has been subjected to a quota. In these situations, the make-up of the broader sample does not reflect the make-up of the population. In the example above, the sample would contain 50% tennis players and 50% racquet ball players, while the broader population would be 80% – 20%. This discrepancy is often counterbalanced during analysis through a process known as *data weighting*.

B. Response and Completion Rates.

"Sample size" refers to the actual number of observations made (individuals measured), not the number invited to participate. As should be clear to anyone who has refused to participate in a study – as when the telephone rings at 7:00 just as dinner is set on the table – just because someone is selected to become a member of a sample will not insure that data is actually collected from them.

Even if someone begins the research process – say, answers a few questions at the beginning of a survey – it doesn't mean that all the requisite information is collected; that the survey is completed. When respondents self-terminate (*abandon*) a survey prior to its completion, multiple problems for analysis emerge.

The proportion of invitees who actually complete a survey is called the *response rate* or the *completion rate* of the survey. Response rates can vary widely based upon a variety of factors.

Factors that positively influence response and completion rates include:

+ Incenting participation: Despite the appearance of "buying answers," researchers typically compensate respondents for their time and effort in completing surveys. The compensation may take many forms: a monetary payment, entry into a drawing for either a monetary or product related prize, and, depending on the nature of the research, access to selected results of the study. Another form of compensation employed with research panels involves the accumulation of points that may be redeemed for various prizes. Longer time commitments of the respondent during a survey typically require greater compensation.
+ Clear understanding of time requirements: When told of the expected time requirements for their participation in a study they are more apt to complete the study.
+ Short surveys: Surveys that take a long time are often an imposition on respondents. They are more likely to respond to shorter surveys than they are to longer ones. Surveys that are less than 10 minutes long appear to have greater completion rates than longer ones. If the survey takes a considerable amount of time, then the respondent should be forewarned. They failure to do so will result in a high rate respondent drops (abandoning the interview in the middle). This is costly both in terms of the wasted time of the interviewer and the loss of credibility with the respondent.
+ Convenience of participation: Respondents are more likely to participate when they can pick the time that they complete the survey (as well as the location).
+ Personal commitment or obligation: Respondents who feel a personal commitment or obligation to either the researcher or the subject are more likely to participate in research and complete surveys. Customers of some businesses are willing to respond to customer satisfaction studies in the belief that there will be an improvement in the product or service based upon the results of the research.

  Sometimes a commitment or obligation is used or implied in order to increase response rates. For example, some research companies will include 2 dollars in mailed surveys in order instill a feeling of obligation upon those invited to participate in the research. Sometimes researchers will use interviewers rather than web based surveys because some respondents are hesitant to say "no" to an otherwise unobtrusive human being.

+ Respondent effort: When researchers design studies in such a way that the respondent is expected to perform some task (e.g., make sure a set of numbers adds to 100), they are requiring an effort out of the respondent that they may be reluctant to undertake. Likewise, asking bank after bank of similarly worded questions will fatigue the respondent. Not seeing the value, many will abandon the survey.
+ Respect for respondent: Questionnaires that question the intelligence or character of the respondent – even if inadvertently – run the risk of increasing "drops."

V. Bad Samples

A good sample will provide the opportunity to draw representative conclusions about the population. A bad sample will not. Bad samples are those that do not possess the relevant characteristics of the population. The three most common ways for this to occur are:

- ✓ When the sample size is simply too small. How could you possibly draw a conclusion about, say, geographic diversity from a sample of 8? Or a national company to identify the style preferences for uniforms from a sample of 16?
- ✓ When the sample is drawn from a population or frame that (also) includes individuals who are NOT members of the population. If the population of interest is those who eat at Quick Serve Restaurants at least once a week, any consumer panel will have individuals who are part of this population. But it will also include people who are NOT part of that population. Care must be taken to exclude from the sample individuals who are not members of the population.
- ✓ When the sample contains only a (biased) subset of the population. It should be obvious that a department store sample comprised solely of male customers would be biased. The same would be true if the sample came only from New England, or were all in college, or happened to be in the store on Monday morning. All population member types must be included in the sample.

The consequences of a bad sample are biased results. Either the conclusions are spurious or drawn about a population that is different than the one of interest. Collect information from a large, diverse group who are all members of the population and your inferences and inductions will be valid and useful to you and your organization.

References

Boyd, P. (2006, December). By the numbers: A sample size table. *Quirks Marketing Research Review*, 30-31, (Article ID: 20061209).

Murphy, K. R. & Myors, B. (2004*). Statistical Power Analysis (2nd Ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.