

Chapter 8: Quantitative Sampling

- I. Introduction to Sampling
 - a. The primary goal of sampling is to get a representative sample, or a small collection of units or cases from a much larger collection or population, such that the researcher can study the smaller group and produce accurate generalizations about the larger group. Researchers focus on the specific techniques that will yield highly representative samples (i.e., samples that are very much like the population). Quantitative researchers tend to use a type of sampling based on theories of probability from mathematics, called probability sampling.
- II. Approaches to Sampling: Nonprobability and Probability Sampling Techniques
 - a. Nonprobability Sampling
 - i. A sampling technique in which each unit in a population does not have a specifiable probability of being selected. In other words, nonprobability sampling does not select their units from the population in a mathematically random way. As a result, nonrandom samples typically produce samples that are not representative of the population. This also means that their ability to generalize from them is very limited.
 1. Types of Nonprobability Sampling Techniques
 - a. Haphazard, Accidental, or Convenience Sample
 - i. A sampling procedure in which a researcher selects any cases in any manner that is convenient to be included in the sample. Haphazard sampling can produce ineffective, highly unrepresentative samples and is not recommended. When a researcher haphazardly selects cases that are convenient, he or she can easily get a sample that seriously misrepresents the population. Such samples are cheap and quick; however, the systematic errors that easily occur make them worse than no sample at all.
 - b. Quota Sampling
 - i. Is an improvement over haphazard sampling. In quota sampling, a researcher first identifies relevant categories of people (e.g., male, female; under age of 30, over the age of 30), then decides how many to get in each category. Thus, the number of people in various categories of the sample is fixed.
 - c. Purposive or Judgmental Sample
 - i. Purposive sampling is an acceptable kind of sampling for special situations. It uses the judgment of an expert in selecting cases or it selects cases with a specific purpose in mind. Purposive sampling is used most often when a difficult-to-reach population needs to be measured.
 - d. Snowball Sampling
 - i. Snowball sampling (also called network, chain referral, or reputational sampling) is a method for identifying and sampling the cases in a network. It begins with one or a few people or cases and spreads out on the basis of links to the initial cases.
 - b. Probability Sampling
 - i. A sampling technique in which each unit in a population has a specifiable chance of being selected. The motivation behind using probability sampling is to generate a sample that is representative of the population in which it was drawn. Random sampling does not guarantee that every random sample perfectly represents the

population. Instead, it means that most random samples will be close to the population most of the time, and that one can calculate the probability of a particular sample being accurate.

- ii. The Jargon of Random Sampling
 1. Sampling Element
 - a. A sampling element is the unit of analysis or case in a population that is being measured.
 2. Population
 - a. The large pool of sampling elements in a study is the population or universe. Unfortunately, it's not that simple. The novice researcher must understand that a population is an abstract concept. How can a population be an abstract concept, when there are a given number of people at a certain time? Except for specific small populations, one can never truly freeze a population to measure it. Because a population is an abstract concept, except for small-specialized populations (e.g., all the students in Dominic Little's sociology 497 class in Spring 2000), a researcher needs to estimate the population. As an abstract concept, the population needs an operational definition. This process is similar to developing operational definitions for constructs that are measured. A researcher operationalizes a population by developing a specific list that closely approximates all the elements in the population. This is a sampling frame (to be discussed later).
 - i. Operationalizing the Population
 1. Define the unit being sampled (i.e., Dominic Little's sociology 497 students)
 2. Define the geographical location (i.e., located at CSUN in room SH106)
 3. Define the temporal boundaries (i.e., Spring 2000)
 3. Target Population
 - a. Refers to the specific pool of cases that he or she wants to study and has a working sampling frame.
 4. Sampling Ratio
 - a. The sampling ratio is determined by dividing the sample size by the total population. For example, if a population has 50,000 people, and a researcher draws 5,000 people for the sample, the sample ratio would be .10 (5,000/50,000).
 5. Sampling Frame
 - a. A researcher operationalizes a population by developing a specific list that closely approximates all the elements in the population. This is a sampling frame. He or she can choose from many types of sampling frames: Telephone directories, driver's license records, and so on. Listing the elements in a population sounds simple. But it is often difficult because there may be no good list of elements in a population. A good sampling frame is crucial to good sampling. A mismatch between the sampling frame and the conceptually defined population can be a major source of error. Just as a mismatch between the theoretical and operational definitions of a variable creates invalid measurement, so a mismatch between the sampling frame and the population causes invalid sampling. With a few exceptions sampling frames are almost always inaccurate.
 6. Parameter
 - a. Any true characteristic of a population. Parameters are determined when all the elements in a population are measured.

The population parameter is never known with absolute accuracy for large populations, so researchers must estimate it on the basis of samples. In other words, they use information from the sample to infer things about the population.

7. Statistic
 - a. Within the context of sampling theory and this discussion, a statistic is any characteristic of a sample that may be used to infer about a parameter of a population.
- iii. Why Random?
 1. Random samples are most likely to yield a sample that truly represents the population when compared to nonrandom samples. In other words, it enables researchers to make accurate assumptions or generalizations from the sample to the population under investigation.
 2. Random sampling lets a researcher statistically calculate the relationship between the sample and the population- that is, the size of the sampling error.
 - a. Sampling Error Defined
 - i. A nonstatistical definition of the sampling error is the deviation between sample results and a population parameter due to random processes.
 3. Fewer Resources are Necessary: Time and Cost
 - a. If properly conducted, a random sample can produce results that can be used to accurately predict parameters within the population at a fraction of the cost of measuring the entire population. For example, how much time and money would it cost to survey the entire U.S. population? Compare that figure to what it would cost in time and money to survey a sample of 2000 U.S. residents.
 4. Accuracy
 - a. The results of a well-designed, carefully executed probability sample will produce results that are equally if not more accurate than trying to reach every single person in the whole population.
- iv. Types of Probability Sampling Techniques
 1. Simple Random
 - a. In simple random sampling, a researcher develops an accurate sampling frame, selects elements from the sampling frame according to a mathematically random procedure, and then locates the exact element that was selected for inclusion in the sample.
 2. Systematic Sampling
 - a. Elements are randomly selected using a sampling interval. The sampling interval (i.e., K^{th} is some number) tells the researcher how to select elements from a sampling frame by skipping elements in the frame before selecting one for the sample. For example, a researcher would have a list of 1,000 elements in her or his population. Let's assume the sample size is 100. In this case, the researcher would select every 10th case. There are two tricks to this that must be followed: first, the sample frame must have the elements ordered in a random way and second; the starting point (the point at which the first element is selected for inclusion into the sample) must be determined randomly.
 3. Stratified Sampling
 - a. In stratified random sampling, a researcher first divides the population into subpopulations (strata: defined as a characteristic of the population. For example, female and male.) on the basis of supplementary information. After dividing the population into strata, the researcher draws a random sample from each

subpopulation. In general, stratified sampling produces samples that are more representative of the population than simple random sampling if the stratum information is accurate.

4. Cluster Sampling

- a. Cluster sampling addresses two problems: Researchers lack a good sampling frame for a geographically dispersed population and the cost to reach a sampled element is very high. Instead of using a single sampling frame, researchers use a sampling design that involves multiple stages and clusters. A cluster is a unit that contains final sampling elements but can be treated temporarily as a sampling element itself. In other words, the researcher randomly samples clusters, and then randomly samples elements from within the selected clusters; this has a big practical advantage. He or she can create a good sampling frame of clusters, even if it is impossible to create one for sampling elements. Once the researcher gets a sample of clusters, creating a sampling frame for elements within each cluster becomes more manageable. A second advantage for geographically dispersed populations is that elements within each cluster are physically closer to one another. This may produce a savings in locating or reaching each element.

III. Random Digit Dialing

- a. Random-digit-dialing (RDD) is a special sampling technique used in research projects in which the general public is interviewed by telephone. Here is how RDD works in the United States. Telephone numbers have three parts: a three-digit area code, a three-digit exchange number or central office code, and a four-digit number. In RDD, a researcher identifies active area codes and exchanges, and then randomly selects four digit numbers. After finding and calling a working residential number, a second stage of sampling is necessary, within household sampling, to select the person to be interviewed. Remember that the sampling element in RDD is the phone number, not the person or the household.

IV. How Large Should a Sample Be?

- a. The best answer to this question is, "It depends!" What does it depend on?
 - i. The kind of data analysis the researcher plans (descriptive, multiple regression).
 - ii. On how accurate the sample has to be for the researcher's purposes (acceptable sampling error).
 - iii. On population characteristics (homogenous or heterogeneous, large or small). On principle for sample sizes is, the smaller the population, the bigger the sampling ratio has to be for an accurate sample. Larger populations permit smaller sampling ratios for equally good samples. This is because as the population size grows, the returns in accuracy for sample size shrink. For small populations (under 1,000), a researcher needs a large sampling ratio (about 30%). For moderately large populations (10,000), a smaller sampling ratio (about 10%) is needed to be equally accurate. For large populations (over 150,000), smaller sampling ratios (about 1%) are possible to be very accurate. To sample from very large populations (over 10,000,000), one can achieve accuracy using tiny sampling ratios (0.025%).

V. Drawing Inferences

- a. The purpose of sampling is to enable a researcher to draw inferences from the sample to the population. The thing to remember is: probability samples are more likely when compared to nonprobability samples to yield representative samples of the population. In other words, a researcher, who wants to draw inferences about the population from his or her sample, should always try to produce a sample that is similar to the population. If the sample is not similar or representative of the population in which it was drawn, the ability to make accurate inferences is highly impaired.

VI. So, Should I Always Use A Probability Sampling Technique?

- a. NO! The answer is a little more complicated than that. Besides, if it were that easy to determine the sampling technique for a study why would there be so many to choose from? The short answer to this question is: It depends. It depends on numerous factors. This

question will be better answered in class but here's a rule of thumb: Choose a technique that is well suited for your study, budget and time. But always remember, no matter how much hard work goes into research design, data collection, pilot testing or preparation, a poorly drawn sample may be useless when attempting to generalize to larger populations.