# Chapter Seven

Sampling and Sampling Distributions

**Summary**

• Through the process of sampling, researchers attempt to generalize the characteristics of a large group (the population) from a subset (sample) selected from that group. The term parameter, associated with the population, refers to the information we are interested in finding out. Statistic refers to a corresponding calculated sample statistic.

• A probability sample design allows us to estimate the extent to which the findings based on one sample are likely to differ from what we would find by studying the entire population.

• A simple random sample is chosen in such a way as to ensure that every member of the population and every combination of N members have an equal chance of being chosen.

• In systematic sampling, every Kth member in the total population is chosen for inclusion in the sample after the first member of the sample is selected at random from the first K members in the population.

• A stratified random sample is obtained by (1) dividing the population into subgroups based on one or more variables central to our analysis and (2) then drawing a simple random sample from each of the subgroups.

• The sampling distribution is a theoretical probability distribution of all possible sample values for the statistic in which we are interested. The sampling distribution of the mean is a frequency distribution of all possible sample means of the same size that can be drawn from the population of interest.

• According to the central limit theorem, if all possible random samples of size N are drawn from a population with a mean μY and a standard deviation σY, then as N becomes larger, the sampling distribution of sample means becomes approximately normal, with mean and standard deviation σY/√N.

• The central limit theorem tells us that with sufficient sample size, the sampling distribution of the mean will be normal regardless of the shape of the population distribution. Therefore, even when the population distribution is skewed, we can still assume that the sampling distribution of the mean is normal, given a large enough randomly selected sample size.

**Outline**

* Aims of Sampling
  + The population is a group that includes all the cases (individuals, objects, or groups) in which the researcher is interested
  + The sample is a subset of cases selected from a population
  + The term parameter, associated with the population, refers to measures used to describe the distribution of the population we are interested in
  + The term statistic refers to a corresponding characteristic calculated for the sample
  + The major objective of sampling theory and statistical inference is to provide estimates of unknown parameters from sample statistics that can be easily obtained and calculated
* Some Basic Principles of Probability
  + A probability is a quantitative measure of the chance that an event will occur
  + It is expressed as a ratio of the number of times a desired outcome will occur relative to the set of all possible and equally likely outcomes
  + The term probability is often applied to events where the outcome is uncertain
  + The relative frequency method is used when we use information from past events to help us predict the likelihood of future events
  + The normal distribution can be used to estimate the probability of occurrences of certain observations
* Probability Sampling
  + Probability sampling is a method that enables the researcher to specify for each case in the population the probability of its inclusion in the sample
  + Although accurate estimates of sampling error can be made only from probability samples, social scientists often use nonprobability samples because they are more convenient and cheaper to collect
  + A simple random sample is a sample design chosen in such a way as to ensure that (1) every member of the population has an equal chance of being chosen and (2) every combination of N members has an equal chance of being chosen
    - Researchers usually use computer programs or tables of random numbers in selecting random samples
  + Systematic random sampling is a method of sampling in which every Kth member in the total population is chosen for inclusion in the sample after the first member of the sample is selected at random from among the first K members in the population
  + Stratified random sampling is a method of sampling obtained by (1) dividing the population into subgroups based on one or more variables central to our analysis and (2) then drawing a simple random sample from each of the subgroups
    - A proportionate stratified sample is one in which the size of the sample selected from each subgroup is proportional to the size of that subgroup in the entire population
    - In a disproportionate stratified sample, the size of the sample selected from each subgroup is deliberately made disproportional to the size of that subgroup in the population
  + A Closer Look 7.1 Disproportionate Stratified Samples and Diversity
* The Concept of Sampling Distribution
  + The sampling distribution helps estimate the likelihood of our sample statistics and, therefore, enables us to generalize from the sample to the population
  + Sampling error is the discrepancy between a sample estimate of a population parameter and the real population parameter
  + The sampling distribution is a theoretical probability distribution of all possible sample values for the statistic in which we are interested
* The Sampling Distribution of the Mean
  + The sampling distribution of the mean is a theoretical distribution of sample means that would be obtained by drawing from the population all possible samples of the same size
* The Mean of the Sampling Distribution
  + The sampling distribution can be described in terms of its mean and standard deviation
  + The standard deviation of the sampling distribution is also called the standard error of the mean
* The Central Limit Theorem
  + One of the most important statistical principles underlying statistical inference
  + If all possible random samples of size N are drawn from a population with a mean μY and a standard deviation σY, then as N becomes larger, the sampling distribution of sample means becomes approximately normal, with mean and standard deviation
  + The central limit theorem tells us that with a sufficient sample size the sampling distribution of the mean will be normal regardless of the shape of the population distribution
  + Although there is no hard-and-fast rule, a general rule of thumb is that when N is 50 or more, the sampling distribution of the mean will be approximately normal regardless of the shape of the distribution
* Statistics in Practice
  + The central limit theorem