1 Samples and Sampling Distributions

1.1 Objectives

- Explain common sample biases.
- Describe various sampling methods.
- Describe the sampling distribution of the sample mean (proportion).
- Explain the importance of the Central Limit Theorem.
- Determine if the sampling distribution of sample means (proportions) is normally distributed.
- Calculate, or infer, the expected value and standard error of the sampling distribution of sample means (proportions)
- Find probabilities for sample means (proportions) using properties of the sampling distribution of sample means (proportions).

1.2 Sampling

A **sampling frame** is a group of accessible subjects from which to draw a sample.

Example: Frequent shopper list - a store does not have a list of all of its regular shoppers but it does have a list of shoppers who signed up for the "frequent shopper" program.

Caution: When the sampling frame differs from the population, which it almost always does, you must consider and deal with the differences.

- Are the opinions of those who signed up as frequent shoppers different from the rest of the shoppers?
- What about customers who signed up for the frequent shopper program but who don't shop there anymore?

Basic Definitions and Concepts	
Definition	**Example**
A **population** consists of all items of interest in a statistical problem. A **sample** is a subset of the population. A valid sample contains information that is representative of the population.	Population – all American high school teachers Sample – 30 high school teachers who, as a group, represent the population
Representative – the participants of a study include subjects that the study is going to be about.	A survey about home water treatment systems selects participants who own homes.
Random – the selection of subjects, the questions in a survey, and/or the process in a study involve no favoritism	The selection of manufacturers for a survey about future expectations does not favor participants based on type of manufacturing, location, size, or sales revenue.
Biased - the selection of subjects, the questions in a survey, and/or the process in a study show favoritism	A survey about an upcoming election selects participants who answer calls from unknown caller IDs.

1.2.1 Bias

Bias refers to the tendency of sample statistic to systematically over- or underestimate a population parameter.

Types of Bias	
Definition	**Example**
Selection bias refers to a systematic underrepresentation of certain groups from consideration for a sample, i.e., portions of the population are excluded from the sample.	An Internet poll asks respondents to enter the number of computers they have in their homes. People without a computer or with limited access to a computer are excluded from the survey.
Nonresponse bias respondents differ in meaningful ways from non-respondents.	An email survey asked people how receptive they are to receiving email solicitations. People who don't like email solicitations will not respond.
Response bias (also called survey **bias**) is the tendency of a person to answer questions on a survey untruthfully or misleadingly.	The leading nature of the following survey question may pressure respondents to answer 'no.' "Many people believe this playground is too small and in need of repair. Do you think the playground should be repaired and expanded even if that means imposing an entrance fee to the park?"

1.2.2 Sampling Methods

Bad Sampling Methods	
Definition	**Example**
Convenience sample – only subjects who are convenient are included in the study or survey. Biased toward those who are available.	A manager wants to know what people think about the selection of stores and restaurants in the mall and surveys people who frequent the mall. What do people who don't go to the mall think about the stores and restaurants?
Voluntary response sample – a large group of subjects are invited to participate, and those who do are counted.	A request that travelers who have used the local airport visit a survey site to report on their experiences is much more likely to hear from those who had long waits, cancelled flights, and lost luggage than from those who had flights that were on time and carefree. These samples tend to be biased toward those with strong opinions or who are strongly motivated – especially from those with negative opinions.

Valid Sampling Methods		
Definition	**Example**	**Visual**
Simple Random Sample (SRS) Randomly select *n* observations from the population (sampling frame) such that the resulting group had the same probability of being selected from the population as any other sample of *n* observations.	A population contains 10 members under the age of 25 and 20 members over the age of 25. The sample will include six people chosen at random, without regard to age.	
Stratified Random Sampling 1. Define mutually exclusive and collectively exhaustive groups, called strata, based on members' shared attributes or characteristics. 2. Divide the population (sampling frame) into the strata. 3. Randomly select observations from each stratum that are proportional to the stratum's size. 4. Combine the selected outcomes into one stratified sample.	A population contains 10 members under the age of 25 and 20 members over the age of 25. The sample will include two people chosen at random under the age of 25 and four people chosen at random over 25. Bias can occur if information from the sample overemphasizes a particular stratum of the population.	
Cluster Sampling 1. Define mutually exclusive and collectively exhaustive groups, called clusters, based on members' shared attributes or characteristics. 2. Divide the population (sampling frame) into the clusters. 3. Randomly select one or more clusters and collect all observations from each selected cluster. 4. Combine the selected outcomes into one cluster sample.	A population contains 5,000,000 members divided approximately equally among 5 Northeastern states. The sample will include all members from one or two randomly selected states.	
Systematic Sample Shuffle the database to make sure certain groups are not excluded and take every nth subject.	A sample consists of every 10^{th} employee from a population of 500 employees.	