Statistical Inference:

- The aim is to obtain information about a population from information contained in a sample.
- A population is the set of all the elements of interest under consideration. (p.258)
- Numerical characteristics of a population, such as the mean and standard deviation, are called parameters. (p.258)
- A sample is a subset of the population selected for analysis. (p.258)
- The sample results provide only estimates of the values of the population characteristics.
- The way in which we select our sample is important
  - Keep in mind we typically take a sample from a population to learn more about the population
  - Our sample should be representative of the population at large
  - Knowing whether a sample is representative of the population is difficult since it is our limited knowledge of the population that leads us to sample in the first place
An approach to gathering our sample must be developed to safeguard against possible biases.

### 7.2 Simple Random Sampling

**From a Finite Population**

- A random sample from a finite population of size $N$ is a sample selected such that each possible sample of size $n$ has the same probability of being selected. (p.260)
- Replacing each sample element before selecting subsequent elements is called sampling with replacement.

**From an Infinite Population**

- A random sample from an infinite population is a sample selected such that the following conditions are satisfied:
  - Each element comes from the same population
  - Each element is selected independently

### 7.3 Point Estimation

- We use the data from the sample to compute a value of a sample statistic that serves as an estimate of a population parameter.
  - Another way to say this is that we want to infer the population characteristics (e.g., mean) from the sample data
• A sample statistic is a real-valued function \( T = r(X_1, X_2, X_3, \ldots, X_n) \) of the random variables. It is a formula showing how to combine the sample data to form a point estimate of the population parameter. It is called an estimator.

• We refer to the sample mean \( \bar{x} \) as the point estimator of the population mean \( \mu \). (p.265)

• \( s \) is the point estimator of the population standard deviation \( \sigma \). (p.265)

• \( \bar{p} \) is the point estimator of the population proportion \( p \).

**Sampling Error**

• The absolute difference between an unbiased point estimate and the corresponding population parameter is called the sampling error.

• Sampling error is the result of using a subset of the population, not the entire population.

• The sampling errors are:
  
  \[
  | x - \mu | \quad \text{for sample mean}
  
  | s - \sigma | \quad \text{for sample standard deviation}
  
  | p - P | \quad \text{for sample proportion}
  
**7.4 Introduction to Sampling Distributions**
• The sample statistic T is a function of $X_1, X_2, X_3, \ldots, X_n$ and since $X_1, X_2, X_3, \ldots, X_n$ are random variables, it follows that T is also a random variable.
  o The value of T will vary from sample to sample since the values assumed by $X_1, X_2, X_3, \ldots, X_n$ will vary from sample to sample.
• The value our sample mean (for example) takes on depends on the actual sample selected.
• For each different possible sample we could generate a value of T.
• Just as many different samples can be drawn from a population, many different sample means can be formed.
• **This idea is very important**
  o It suggests that sample statistics have probability distributions showing the different possible values the sample statistic can assume, along with the associated probability.
  o The probability distribution for a sample statistic is called a sampling distribution.

7.5 Sampling Distribution of $\bar{x}$
• The sampling distribution of \( \bar{x} \) is the probability distribution of all possible values of the sample mean \( \bar{x} \). (p. 270)
  o As with other sampling distributions \( \bar{x} \) has an expected value, a standard deviation, and a characteristic shape or form.

• When the expected value of a point estimator equals the population parameter, we say the point estimator is unbiased. (p. 270)

• Form of the Sampling Distribution of \( \bar{x} \)
  o Population has a normal distribution – In many situations it is reasonable to assume that the population from which we are selecting a simple random sample has a normal or near normal distribution. (p.272)
  o Population does not have a normal distribution – When the population from which we are selecting a simple random sample does not have a normal distribution, the central limit theorem is helpful in identifying the shape of the sampling distribution. (p.272)

• The central limit theorem says the sampling distribution of \( \bar{x} \) can be approximated by a normal probability distribution when we use a large (\( n \geq 30 \)) random sample.
7.6 Sampling Distribution of $\bar{p}$

- Remember that the sample proportion $\bar{p}$ is the point estimator of the population proportion $p$.
- The sampling distribution of $\bar{p}$ is the probability distribution of all possible values of the sample proportion $\bar{p}$.
- Standard Deviation of $\bar{p}$
  - Formula depends on whether the population is finite or infinite.
- Form of the Sampling Distribution of $\bar{p}$
  - The sampling distribution of $\bar{p}$ can be approximated by a normal distribution whenever $np \geq 5$ and $n(1-p) \geq 5$. (p. 281)

7.7 Properties of Point Estimators

- Since several different sample statistics can be used as point estimators of different population parameters general notation will be used in this section:
  - $\theta =$ the population parameter of interest
  - $\theta =$ the sample statistic or point estimator of $\theta$

- Unbiased
  - If the expected value of the sample statistic is equal to the population parameter being estimated, the
sample statistic is an unbiased estimator of the population parameter.

- **Efficiency**
  - Given the choice of two unbiased estimators, we prefer the point estimator with the smaller standard deviation.

- **Consistency**
  - A point estimator is consistent if the values of the point estimator become closer to the population parameter as the sample size becomes larger.