Chapter 7 presents the beginning of inferential statistics.

Concept: Inferential Statistics

The two major activities of inferential statistics are

1 to use sample data to estimate values of population parameters (proportions, means, and variances), and

2 to test hypotheses or claims made about these population parameters.

Chapter 7

The two major activities of Chaper 7

☞ We introduce methods for estimating values of population proportions,
means, and variances.

☞ We also present formulas for determining sample sizes necessary to
estimate those parameters.

☞ These are applications of the Central Limit Theorem (CLT).

To estimate the value of a population parameter, such as the *true* proportion of adults in the United States who believe in global warming, you can use information from a sample in the form of an estimator.

To estimate the value of a population parameter, such as the *true* proportion of adults in the United States who believe in global warming, you can use information from a sample in the form of an estimator.

Definition
An **estimator** is a rule, expressed as a formula, that tells us how to calculate an estimate based on information in the sample.

To estimate the value of a population parameter, such as the *true* proportion of adults in the United States who believe in global warming, you can use information from a sample in the form of an estimator.

Definition
An **estimator** is a rule, expressed as a formula, that tells us how to calculate an estimate based on information in the sample.

Types of Estimators

✓   **Point Estimation**: Based on sample data, a single number is calculated to estimate the population parameter. The rule or formula that describes this number is called a **point estimator**, and the resulting number is called a **point estimate**.

To estimate the value of a population parameter, such as the *true* proportion of adults in the United States who believe in global warming, you can use information from a sample in the form of an estimator.

Definition
An **estimator** is a rule, expressed as a formula, that tells us how to calculate an estimate based on information in the sample.

Types of Estimators

✓   **Point Estimation**: Based on sample data, a single number is calculated to estimate the population parameter. The rule or formula that describes this number is called a **point estimator**, and the resulting number is called a **point estimate**.

✓   **Interval Estimation**: Based on sample data, two numbers are calculated to form an interval within which the parameter is expected to lie. The rule or formula that describes this calculation is called the **interval estimate** or **confidence interval**.

Example:
In the Chapter Problem we noted that in a Pew Research Center poll, 70% of 1501 randomly selected adults in the United States believe in global warming, so the sample proportion is $\hat{p} = 0.70$. Find the best point estimate of the proportion of all adults in the United States who believe in global warming.

Solution:
Because the sample proportion is the best point estimate of the population proportion, we conclude that the best point estimate of p is 0.70. When using the sample results to estimate the percentage of all adults in the United States who believe in global warming, the best estimate is 70%.

Definition
A **confidence level** is the probability $1 - \alpha$ that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times. (The confidence level is also called degree of confidence, or the confidence coefficient.)

Confidence Level	α	Critical Value, $z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

Definition
The distance (or difference in absolute value) between an estimate and the estimated
parameter is called the **error of estimation**.

In this chapter, we assume that sample sizes are large enough that the unbiased
estimators we study have sampling distributions that can be approximated by the normal
distribution (because of the Central Limit Theorem). Moreover, the Empirical Rule tells us
that for any point estimator with a normal distribution, 95% of the of all point estimates
will lie within two (or more exactly, 1.96) standard deviations of the mean of that
distribution. For unbiased estimators, this implies that the difference between the point
estimator and the true value of the parameter will be less than 1.96 standard deviations or
standard errors (SE), and this quantity, called the margin of error (E), provides a practical
upper bound for the error of estimation.

You may want to change the level of confidence being used to construct your confidence
interval from 95% to $(1 - \alpha) \cdot 100\%$. Then, the margin of error will be within $z_{\alpha/2}$ standard
deviations of the mean of the sampling distribution being used.

7.2 Estimating a Population Proportion

In Section 7.2, we use sample data to construct estimators for a population proportion.

The following assumptions will be used.

- ✓ The sample is a simple random sample.

- ✓ The criteria of the binomial distribution are satisfied. That is, there is a fixed number of independent trials, there are two possible outcomes, and the probabilities of each are the same.

- ✓ There are at least 5 successes and at least 5 failures. That is, $n \cdot p \geq 5$ and $n \cdot q \geq 5$. Since p and q are unknown, use $n \cdot \hat{p} \geq 5$ and $n \cdot \hat{q} \geq 5$. This assumption ensures that the binomial distribution can be approximated by the normal distribution.

The following notation will be used:
The sample proportion is denoted $\hat{p}$ (read as "p hat") where $\hat{p}$ follows the rule

$$\hat{p} = \frac{x}{n} = \frac{\text{the number of successes}}{\text{the total number of trials}}$$

Moreover, $\hat{q} = 1 - \hat{p}$ the sample proportion of failures in a sample of size n.

7.2 Estimating a Population Proportion

When estimating a population proportion, the margin of error (E) is

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The confidence interval used to estimate the population proportion is

$$\hat{p} - E < p < \hat{p} + E,$$

or equivalently,

$$(\hat{p} - E, \ \hat{p} + E)$$