## CHAPTER 5

1. When do we use estimation methods?

2. When do we use hypothesis-testing methods?

3. What is the advantage of an interval estimate over a point estimate?

4. Why do we worry about Type I and Type II errors when performing a hypothesis test?

5. How do we reduce our chance of making Type I and Type II errors?

6. What assumptions are necessary when considering inference about one mean using the t-distribution?

## Statistical Inference for One Population Mean

As has been mentioned previously, the main objective of statistics is to conduct statistical inferences concerning parameters of a population. The two methods of statistical inference are: estimation, and hypothesis testing.

Estimation is used to answer the question, "What is the value of the population parameter?"

Hypothesis testing is used to answer the question, "Is the parameter equal to a specific value?"

## Estimation

Types of Estimation
1. Point
2. Interval

Important Terms and Definitions

An **estimator** is a rule that tells us how to calculate the estimate based on sample information.

A **point estimator** of a parameter is a rule that estimates the parameter with a single value.

An **estimate** is a number calculated using an estimator.

An estimator is called **unbiased** if its average value is equal to the parameter being estimated. Otherwise, the estimator is called **biased**. (An unbiased estimator has a bias of zero.)

An **interval estimator** of a parameter is a rule the provides two numbers that form an interval which is likely to contain the parameter. The interval is called a **confidence interval**.

The **confidence coefficient** is the probability that a confidence interval will capture the parameter being estimated.

Point Estimators

Some examples of point estimators are: $\bar{y}$ for the population mean, $s^2$ for the population variance.

A good point estimator is one that is:

1. unbiased, and
2. has a small standard error

Both of the estimators mentioned above are unbiased, and have the smallest standard error for samples taken from a normal distribution.

However, it turns out that the sample standard deviation (s) is a biased estimator for the population standard deviation.

Confidence Intervals

A confidence interval is an interval that is likely to capture the value of the parameter being estimated. The interval consists of two values: the **lower confidence limit (LCL)** and the **upper confidence limit (UCL)**.

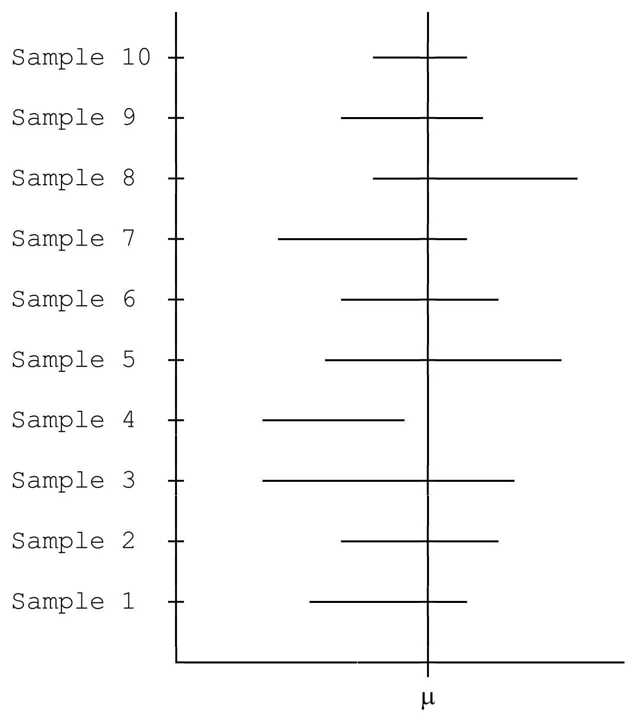
Suppose a 90% confidence interval for $\mu$ was calculated and the result was the interval (12,20).

1. What is the confidence coefficient?

2. What is the LCL? What is the UCL?

3. Which of the following statements expresses what the confidence interval represents?
   a. There is a 90% probability that $\mu$ is between 12 and 20.
   b. There is a 90% probability that the interval from 12 to 20 has enclosed $\mu$.
   c. Both statements are accurate.

**What Might Be Expected to Happen if 90% Confidence Intervals Are Calculated for the Mean**



Using Point Estimators and Confidence Intervals to Make Inferences

When it is operating properly, the Kellogg's raisin scooper will put a mean of 2 scoops of raisins into the boxes of raisin bran. A sample of 36 boxes revealed an average of 1.95 scoops and a standard deviation of .24 of a scoop. The quality control inspector used this data to construct a 95% confidence interval for the mean. The interval was (1.8716, 2.0284).

What is the parameter of interest? Explain.

What is the point estimate used to estimate the parameter of interest?

Based on the confidence interval, do we have enough evidence to conclude that the scooper should be shut down for repairs? Why or why not?

<u>Constructing Large Sample Confidence Intervals</u>

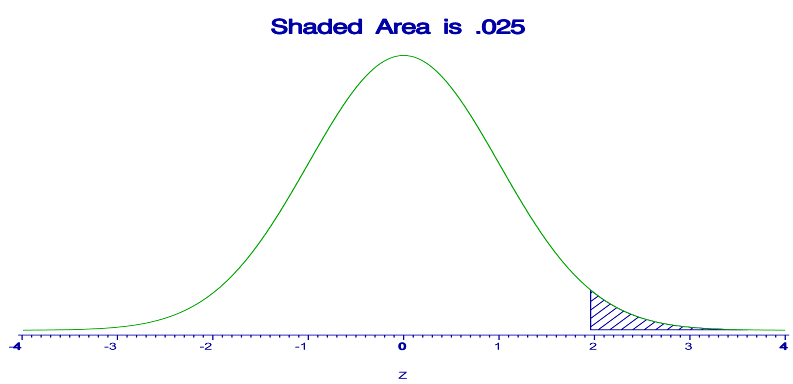
A general expression for the confidence coefficient is $1-\alpha$. So, for a .90 confidence coefficient the value of $\alpha$ is .1. What is alpha for each of the following confidence coefficients?

Confidence coefficient=.95 $\alpha$=

Confidence coefficient=.99 $\alpha$=

Confidence coefficient=.85 $\alpha$=

The **critical value $z_\alpha$** is the standard normal value such that it has an area $\alpha$ to the right of the standard normal curve.




So $z_{.025}$=1.96. Find the following critical z-values.

Find the following values:

$z_{.05}$

$z_{.1}$

$z_{.0668}$

**Large Sample (1-α)100% Confidence Interval Formula for a Single Population Mean**

$$\overline{Y} \pm z_{\mathbf{a/2}} \left( s/\sqrt{n} \right)$$

Example: When it is operating properly, the Kellogg's raisin scooper will put a mean of 2 scoops of raisins into the boxes of raisin bran. A sample of 36 boxes revealed an average of 1.95 scoops and a standard deviation of .24 of a scoop. Construct a 95% confidence interval for the mean number of scoops in a box of raisin bran.

**How Large of a Sample Do We Need?**

Half the width of a confidence interval may be referred to as the "tolerable error", "margin of error," or "sampling error." Frequently, newspapers report the margin of error from their surveys. For example, the paper may say that 52% of voters favor a particular candidate with a margin of error of 1%. Below, we discuss the factors affecting the margin of error and how to determine the sample size needed for a specific margin of error.

For a large sample size the margin of error may be expressed as

$$E = \frac{z_{\alpha/2}\, s}{\sqrt{n}}$$

Factors Affecting the "Margin of Error"

1. The value of the standard deviation (s). We have no control over this value. However, we see that if $z_{\alpha/2}$ and n remain unchanged, then as the variability of the sample increases so does the margin of error.
2. The value of the confidence coefficient ($1-\alpha$). The confidence coefficient affects the value of $z_{\alpha/2}$. The larger the confidence coefficient, the larger the value of $z_{\alpha/2}$ will be. Thus, as the value for the confidence coefficient increases, so does the margin of error.
3. The value of the sample size (n). Since n is in the denominator of the expression for the margin of error, we see that increasing the sample size will decrease the margin of error.

By rearranging the expression for the margin of error, we can determine the sample size that will be needed for a specified margin of error. This formula is

$$n = \frac{\left(z_{\alpha/2}^2\right)\left(s^2\right)}{E^2}$$

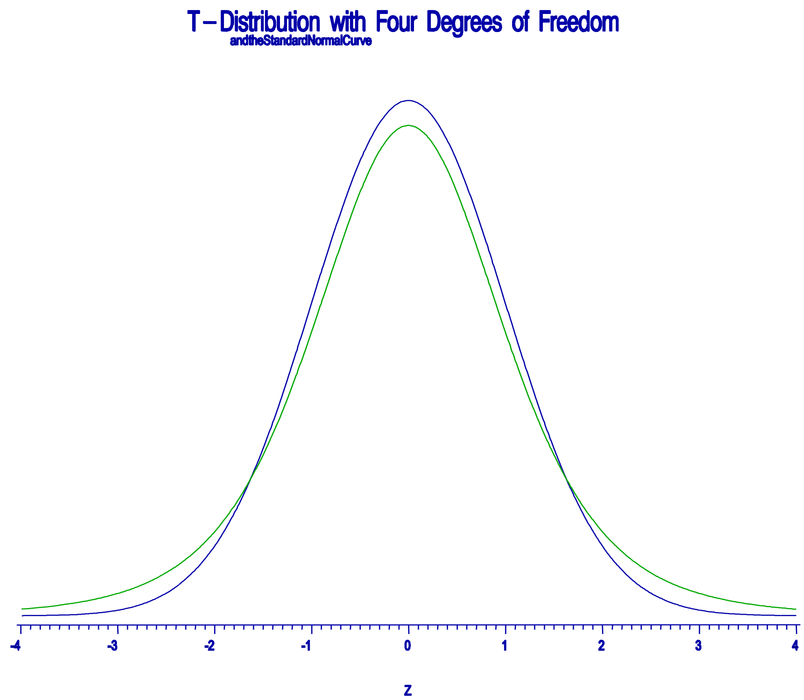
Example: The City of Clemson would like to know the average amount of money a fan spends on a home game-day. The results of a preliminary sample indicated that the sample average was \$23.14 and a standard deviation of \$8.49. Based on this preliminary sample, how many fans should be surveyed for the sample mean to be within \$2 of the actual mean, with 90% confidence?

## The t-distribution and Small Samples

For small sample sizes (less than 30), the t-distribution is used for the construction of confidence intervals and hypothesis testing of the mean.

Properties of the t-distribution

1. The mean is zero.
2. It is symmetric about the mean.
3. It has a variance greater than 1; but the variance approaches 1 as the sample size n becomes large.
4. It is less peaked at the mean and thicker at the tails than the normal distribution.
5. It approaches the standard normal distribution as the sample size becomes large.




Use the t-table in the back of your book to find the following t-values:

$t_{.05,5}$ = $t_{.1,15}$ = $t_{.01,10}$ =

**Small-Sample (1-$\alpha$) Confidence Interval Formula for a Single Population Mean**

$$\bar{y} \pm t_{\alpha/2,(n-1)}\, s/\sqrt{n}$$

Example: A forest manager is interested in determining the average volume of standing timber per acre. Five acres from a 625 acre forest were sampled. The volume in "hundreds of board feet" for the sampled acres were 18, 23, 19, 23, and 15. The mean and standard deviation for this sample were 19.60 and 3.44, respectively. Determine a 95% confidence interval for the average volume per acre.