Objectives:

- To learn to apply a confidence interval procedure for estimating a population mean.
- To recognize similarities of structure and interpretation between confidence interval procedures for a population mean and for a population proportion.
- To become familiar with the ***t-distribution*** for finding critical values.
- To discover the roles played by sample size and sample standard deviation as well as by sample mean concerning confidence intervals.
- To see how to check whether the ***t-procedure*** can be validly applied to a given sample of data.
- To continue to recognize the utility of graphical analyses and the limitation of summarizing a distribution of data exclusively by its mean.

Again…So far we have seen…

We have seen that over the long run sample means fall into a particular pattern. This pattern is directly in line with what we refer to as the normal distribution! Thus we can use our knowledge of the Central Limit Theorem to calculate probabilities that a sample statistic falls within a certain range of a population mean.

Problem…

When we want to estimate the population mean, we often do not know the standard deviation of the population – which as we have seen the normal distribution relies greatly on this value – not only do not usually know, there is often no way of actually knowing it. (and if we did we would know the mean already and thus no need for an estimate)

Problem Solved…

For this situation we can apply the ***t-distribution.*** As you will see, this is very similar to a normal distribution but more uncertain and the value of this uncertainty depends on the value of *n*.

Standard error, this is what we call an estimate of the standard deviation of sample means. We use the following formula calculate this.

$$\text{standard error of } \bar{x} = \frac{s}{\sqrt{n}}$$ *(Where s is the standard deviation of the sample)*

t-distribution a distribution used as an alternative when we do not know the standard deviation of a population. The use is very similar to the normal distribution except that since the uncertainty is higher, the confidence intervals need to encompass a larger range of values to be used for analysis. This distribution is very dependent on the sample size *n*.

Characterizations of the t-distributions:

- ***mean = 0 (same as the normal curve)***
- ***degrees of freedom (abbreviated d.f.)*** based on the size of *n* and defines the spread of the mound. This takes the place of the Z-score. Smaller *n*'s yield a smaller *d.f.* and have a wider mound. As *n* increases the *d.f.* increases then the mound gets closer and closer to the shape of the normal curve (as *n* infinite the *t-distribution* gets closer and closer to the normal distribution)

Degrees of Freedom (d.f.) for a t-distribution is found by finding ***n-1 (where n is the size of the sample)***

In future Topics - *There will be other uses of the d.f., which will have different ways of finding it.*

Confidence Interval for a population μ:

$\bar{x} \pm (t^*_{n-1})\frac{s}{\sqrt{n}}$	where (t^*_{n-1}) is the appropriate critical value from the t-distribution with ***n-1*** degrees of freedom (round down if exact value is not found in the table)

Technical Conditions:
- SRS sample from the population of interest
- CLT conditions must be satisfied (n ≥ 30 or normally distributed population)

Be aware:

The reasoning behind and the interpretation of these confidence intervals are the as previous topics: If one were to repeatedly take samples from a population and create a 95% confidence interval from each one, in the long run 95% of the intervals generated will contain the true population mean. We are 95% confident that a particular interval will contain the true population mean.

Again… it is important to realize that this does not imply that there is a 95% probability that an interval will contain the true mean, since probability is based on random events, where a population has a definite value for mean – we just don't know what it is. This is different (though sometimes difficult to grasp) than what is stated at the bottom of page 431. Page 431 states that if we took a whole a bunch of samples and found the intervals as described, wrote these on a strip of paper, pulled one out of the bag, there would be a 95% probability that it would contain the true mean.

Factors that affect Interval Widths:
- ***Smaller samples produce wider intervals***
- ***Samples with more variability will produce a wider interval (larger s.d. yields wider intervals.)***
- ***Higher confidence levels produce wider intervals (95% interval is wider than a 90% interval.)***

Confidence Intervals estimate population means,
NOT the values of individual observations in a population