Learning Goals:

- Construct a confidence interval for the population mean when conditions are met. Interpret the confidence interval in context.
- Interpret the meaning of a confidence level associated with a confidence interval.
- Explain how the margin of error is affected by changes in sample size and by changes in confidence level.

Introduction:

In this activity we will learn to construct a confidence interval that provides a range of estimates for the population mean. As before, the confidence interval is based on a normal model for the distribution of sample statistics, in this case, sample means.

1) Recall the general formula for a 95% confidence interval is…

   Sample statistic ± margin of error
   Sample statistic ± 2(standard error)

   Given what we have learned about the distribution of sample means, this becomes

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$

   where z_c is the tabulated z-value for the desired confidence level, for example 1.96 (rounded to 2 by our rule-of-thumb) for 95% confidence.

2) What conditions have to be met to use this formula to estimate a population mean?

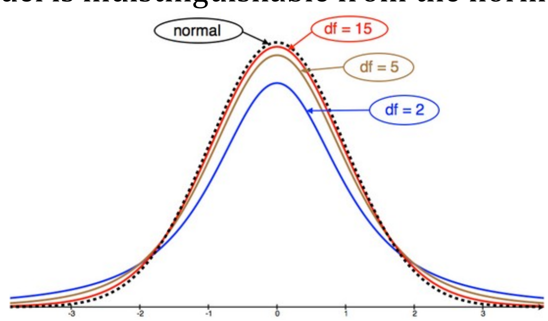
For this confidence interval, we have to know the population standard deviation, σ, based on previous studies. It may occur to you that if you do not know μ, it is unlikely that you know σ. So, we are forced to take a different approach. **We estimate σ using the sample standard deviation *s*.**

We will use the sample standard deviation s, instead of population standard deviation σ, in the calculation of the standard error.

This is the same type of adjustment we used in when substituting the sample proportion $\hat{p}$ for population proportion p in estimating the standard error in sample proportions.

But this adjustment is not as straightforward as our work with proportions. This estimate for σ introduces more uncertainty in the process. The problem is worse with smaller samples because the sample standard deviations vary more. Therefore, for small samples, s is a worse approximation for σ. Unfortunately, this makes the normal model a bad fit and inappropriate for determining critical values. For this reason, we use what is called a *T-model* instead of the normal model.

The T-model, like the normal model, is bell-shaped. However, the shape of a normal model is affected by its mean and standard deviation while a T-model is affected by the sample size, or more specifically by the *degrees of freedom*. The degrees of freedom are one less than the sample size (n-1). Therefore, each sample size has its unique T-model. As sample sizes increase, the T-model is indistinguishable from the normal model.



What is the formula for the confidence interval if we use the sample standard deviation to estimate the population standard deviation?

The confidence interval formula changes from

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} \pm t_c \frac{s}{\sqrt{n}}$$

So how do we determine the T-score we need for our confidence interval?
We will use an applet or we will let StatCrunch do this for us.

What are the conditions for use of the T-model?
The conditions for use of a T-model are similar to the conditions for use of a normal model; however, we can use the T-model for smaller samples in some situations.

- Must have a random sample (as with all statistical inference procedures.)
- If a normal model fits the population's distribution, then use the T-model.
- If a normal model does not fit the population's distribution (or you can't determine if this), then, with sample sizes of 30 or greater, use the T-model.
- For a sample of less than 30, plot the data. If the distribution of data in the sample is not heavily skewed, then we assume the population's distribution is also not heavily skewed and we use the T-model.

1) **Example:** Community college students often work and have family responsibilities in addition to attending college. Busy schedules can result in less sleep. In a project for her statistics class, a student randomly selects 20 students and determines that this sample sleeps on average 6.2 hours a night with a standard deviation of 0.7 hours.

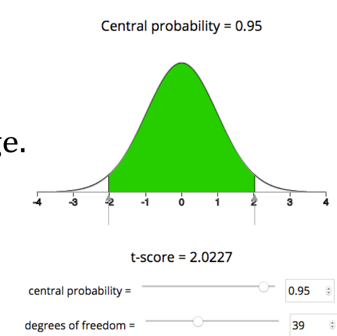
   Here is the distribution of sleep hours in her sample.

   Estimate the mean hours of sleep for students at this college with a confidence interval.

   a) Verify that conditions are met for use of the confidence interval formula involving the T-model.

   b) Why does this student need to use a T-score instead of a Z-score in her confidence interval?

   c) Determine the t-score using the OLI applet image.



   d) Find and interpret the confidence interval.

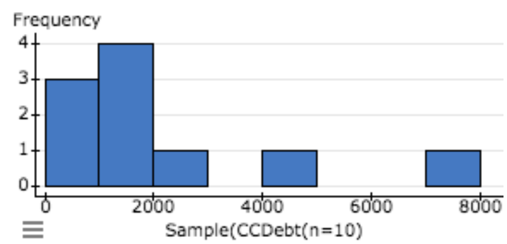
Group work:

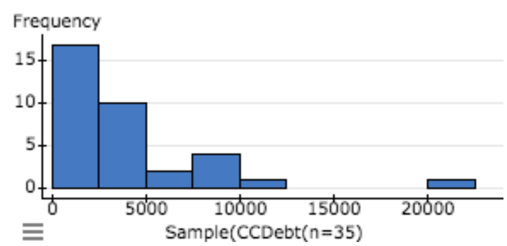
1) We want to estimate the mean amount of credit card debt ($) owed by students at our college.

   Here are two different samples randomly selected from the population of students at our college. Can we use the T-model to find a confidence interval using the sample mean from either of these samples? Why or why not?

Summary statistics:

Column	n	Mean	Std. dev.
Sample(CCDebt(n=10)	10	2239.8	2077.2873
Sample(CCDebt(n=35)	35	3910.8	4338.716



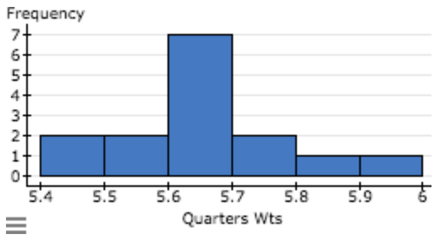


2) To discourage the use of counterfeit coins, vending machines can be set to reject unusually light coins and unusually heavy coins. Such settings are based on an estimated mean weight of the coins. What is the mean weight of quarters?

   As with most manufacturing processes, we can assume that a normal model is a good fit for the distribution of weights.

   We collect 15 quarters from various drawers in our house and car and weigh them.

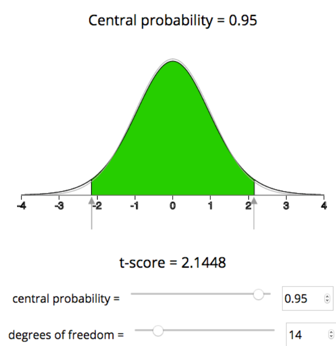
Here is information about the sample:

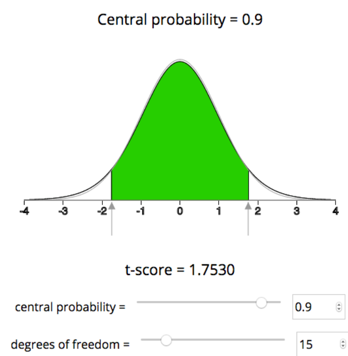


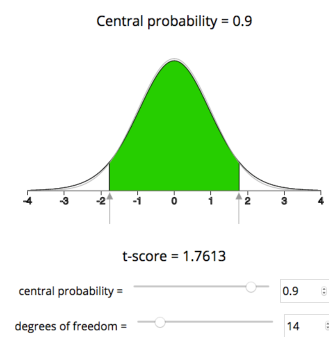
Column	n	Mean	Std. dev.
Quarters Weights (grams)	15	5.65	0.12

a) Explain why we can use the T-model despite the small sample size and slight skew in the data.

b) Pick the image with the correct T-score for a 90% confidence interval for this situation and calculate the 90% confidence interval.







c) According to Wikipedia, quarters are manufactured with a weight of 5.67 grams. Does this seem reasonable given your confidence interval? Why or why not?

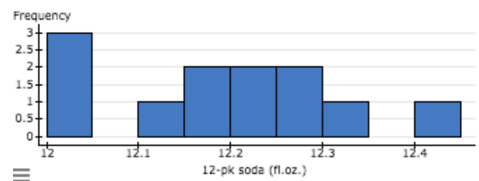
3) A soda is labeled with a volume of 12 fluid ounces. Quality control processes are set to slightly overfill bottles so that customers do not receive less than 12-fl.oz. There is inevitable variability in the filling process; but, as with most manufacturing processes, we can assume that a normal model is a good fit for the distribution ofvolumes.

What is the mean volume of soda in a can marked 12 fluid ounces?

To answer this question, suppose that we measured the volume of 12 sodas in a 12-pack and got a mean of 12.19 fl.oz. Below is information about the sample of 12 sodas.

Summary statistics:

Column	n	Mean	Std. dev.
12-pk soda (fl.oz.)	12	12.190833	0.11988315



Below is the StatCrunch print-out for the confidence interval based on the T- model.

One sample T confidence interval:
μ : Mean of variable

95% confidence interval results:

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.114663	12.267003

a) The StatCrunch T confidence interval is based on the T-model. Explain why use of the T-model is appropriate here despite the small sample size and the slight skew in thedata.

b) What does the standard error tell us?

c) Interpret the confidence interval in the context of soda volume.

4) Below are two T confidence intervals based on the same sample of 12 sodas. One is the 90% confidence interval and the other is the 99% confidence interval. Which is which? How do you know?

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.128683	12.252984

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
12-pk soda (fl.oz.)	12.190833	0.034607285	11	12.08335	12.298317

5) Below are two 95% confidence T intervals based two different samples. One sample was a 6-pack, the other a 12-pack. Which is which? How do you know?

Variable	Sample Mean	Std. Err.	L. Limit	U. Limit
Sample 1	12.190833	0.034607285	12.114663	12.267003

Variable	Sample Mean	Std. Err.	L. Limit	U. Limit
Sample 2	12.106667	0.037475918	12.010332	12.203002