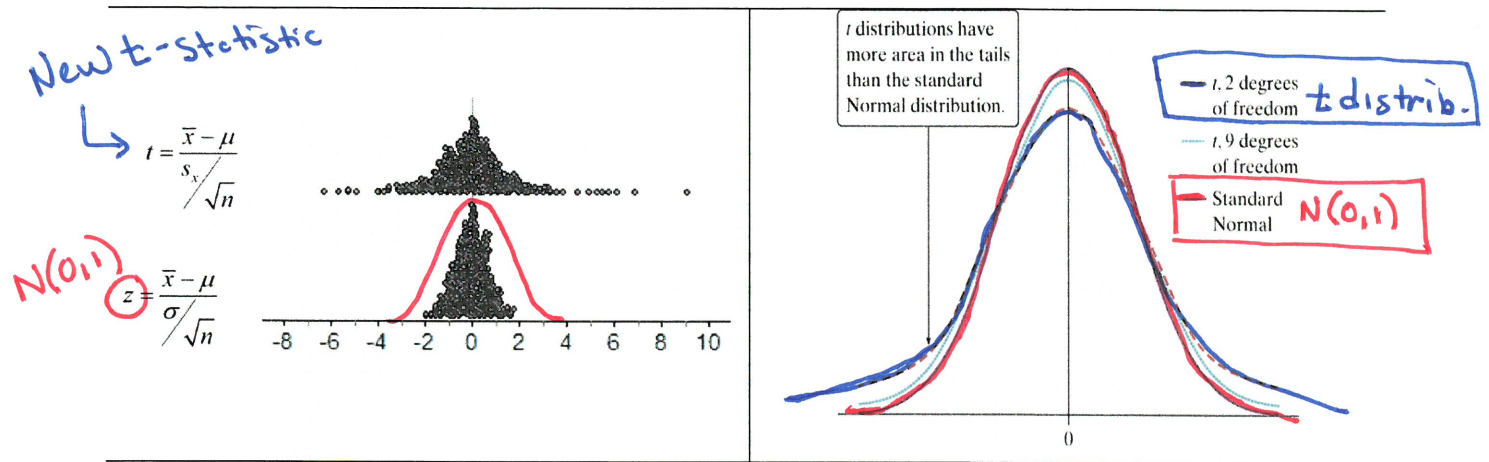


AP Statistics – 8.3b	Name: 2020 KEY
Goal: Understanding CI for Means when σ is UNKNOWN	Date:

I. What you need to know about the t-Distribution



- The density curves of the t-distributions is similar in shape to the standard Normal curve. They both have SYMMETRIC shapes with a SINGLE peak at zero (0).
- The spread of the t-distribution is GREATER (WIDER) than that of the standard Normal distribution.
- The t-distributions has MORE area in the tails and less in the center than the Normal.
- As the sample size (n) or df increase, the t-distribution approaches the NORMAL distribution.
- As the sample size increases, the t-distribution tails narrow and the center becomes taller.
- There is a different t-distribution for each sample size, specified by DEGREES OF FREEDOM (DF).
- We use the “t-statistic” to estimate POPULATION MEAN (μ) because we do NOT know the POPULATION STANDARD DEVIATION (σ) as long as the sampling distribution is close to THE SHAPE OF THE NORMAL DISTRIBUTION.
- For the “t-statistic,” we estimate population standard deviation (σ) with sample S.D. (s_x).
- The formula for the “t-statistic” is:

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$$

with degrees of freedom df = n - 1.

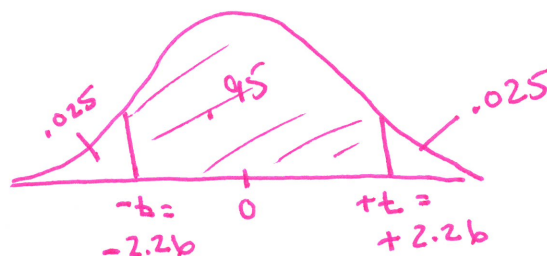
How to find Critical Values for the t-Distribution: Finding t*

Example #1: Suppose you wanted to construct a 95% confidence interval for the mean of a Normal population based on an SRS of size 10. What critical value t* should you use? [Use Calc. InvT]

$$n = 10$$

$$df = 10 - 1 = 9$$

$$t^* = \text{InvT}(.025, 9)$$



$$t^* = 2.26$$

$$\sqrt{5}$$

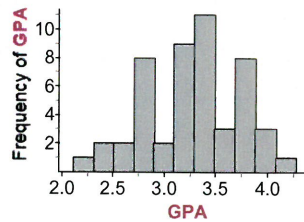
$$z^* = 1.96$$

II. Determine when a one sample t Interval can be done

- **Definition:** An inference procedure is called ROBUST if the probability calculations involved in the procedure remain fairly accurate when a condition for using the procedures is violated.
- The t-procedures are relatively ROBUST when the population is NOT Normal, especially for large ($n > 30$) sample sizes.
The t-procedures are NOT robust against outliers or extremely skewed small datasets.

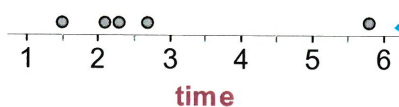
Example #2: "GPA, Coffee, and SAT scores?" Determine whether we can safely use a one-sample t interval to estimate the population mean in each of the following settings.

- a) To estimate the average GPA of students at your school, you randomly select 50 students from classes you take. Here is a histogram of their GPAs.



$\mu =$ true mean GPA AT THIS SCHOOL
 $n = 50$ (large dataset)
 Random condition violated. This is NOT AN SRS, BUT A CONVENIENCE Sample. We should NOT use the t-interval because we can not GENERALIZE TO THE HS.

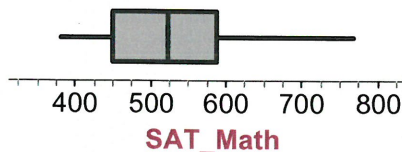
- b) The dot plot below shows the amount of time it took to order and receive a regular coffee in 5 visits to a local coffee shop.



$\mu =$ true mean time to order coffee.
 $n = 5$ (small dataset)
 potential outlier

We should not use the t-interval. This is a small dataset with an outlier.

- c) The boxplot below shows the SAT math score for a random sample of 20 students at your high school.



$\mu =$ true SAT math score at this HS.
 $n = 20$ (small dataset)

Since the dataset is slightly skewed with NO outliers, it would be safe to use the t-interval.

Normal Condition: To use the t-distrib. for μ is checked if 1 of the following applies:
 1) It was stated population is normal.
 2) Large sample ($n \geq 30$), there CLT applies.
 3) Small samples - check the graph of the distribution to confirm it is roughly symmetric with NO outliers.

III. Constructing a Confidence Intervals for μ

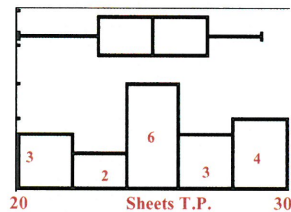
Example #3: “Can you spare a square?” As part of their final project in AP Statistics, Christina and Rachel randomly selected 18 rolls of a generic brand of toilet paper to measure how well this brand could absorb water. To do this, they poured 1/4 cup of water onto a hard surface and counted how many squares it took to completely absorb the water. Here are the results from their 18 rolls:

29 20 25 29 21 24 27 25 24
 29 24 27 28 21 25 26 22 23

Construct and interpret a **99% confidence interval**.

- Define the parameter of interest: **[remember it’s the population parameter]**
 $\mu =$ **the mean number of squares of generic toilet paper needed to absorb 1/4 cup of water**
- What inference method will you use?
Name the interval (or give the formula): 1-sample t interval for μ
- Check conditions:
 - Random:** A random selection of rolls of generic toilet paper was taken. [or SRS $n=18$ rolls TP]
 - Independent:** (Since we are sampling without replacement, we must check the 10% condition.) It is reasonable to believe there are more than $10(18) = 180$ rolls of generic toilet paper produced.
 - Normal:**
NOTE: Since the sample size is small ($n = 18$) and we aren’t told that the population is normally distributed; we need to check the shape of the distribution to confirm the normal condition. To do this, sketch the histogram and provided an explanation.

You must sketch the histogram.
 I recommend you look at the box plot to check for outliers.



Explanation:

Based on the histogram, the graph looks symmetric and I checked the box blot to determine there were no outliers. So it is reasonable to that the distribution is approximately Normal.

- Construct the confidence interval

<ul style="list-style-type: none"> The sample mean for these data is $\bar{x} = 24.94$ The sample standard deviation is $s_x = 2.86$ Find the critical value t^*: 99% CL, $n=18 \rightarrow df=18-1=17$ $t^* = \pm 2.898$ <p>Calc. Command: <code>invT(.005,17)= 2.89823</code></p>	<p><i>sketch graph</i></p>
$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} =$ $24.9 \pm 2.898 \frac{2.86}{\sqrt{18}} =$ $24.94 \pm 1.95 = \quad (22.99, 26.89)$	

CHECK with Calc: [stat] [tests] [8: TInterval] [Data] List: L1 Freq:1 C-Level=.99
Calc. Provides: (22.991, 26.897)

- Interpret the interval in context:
 - We are 99% confident that the interval from about 23 squares to 27 squares captures the true mean number of squares of generic toilet paper needed to absorb 1/4 cup of water.

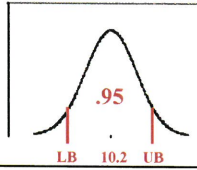
IV. Using a Confidence Intervals to test a claim (hypothesis)

Example #4: “How much homework?” The principal at a large high school claims that students spend at least 10 hours per week doing homework on average. To investigate this claim, an AP Statistics class selected a random sample of 250 students from their school and asked them how many long they spent doing homework during the last week. The sample mean was 10.2 hours and the sample standard deviation was 4.2 hours.

Problem:

- (a) Construct and interpret a 95% confidence interval for the mean time spent doing homework in the last week for students at this school.
- (b) Based on your interval in part (a), what can you conclude about the principal’s claim?

- Define the parameter of interest:
 μ = the mean time spent doing homework in the last week for students at this school
- What inference method will you use?
 Construct a 1-sample t interval for μ with 95% confidence.
- Check conditions:
 - **Random:** The students were randomly selected. (SRS $n=250$)
 - **Independent:** Because we are sampling without replacement, we must check the 10% condition. It is reasonable to believe there are more than $10(250) = 2500$ students at this large high school.
 - **Normal:** The sample size is large ($n = 250$) and satisfy CLT so we are safe using t -procedures.
- Construct the confidence interval

<ul style="list-style-type: none"> ▪ The sample mean for these data is $\bar{x} = 10.2$ ▪ The sample standard deviation is $s_x = 4.2$ ▪ Find the critical value t^*: $n=250 \rightarrow df=250-1=249$ $t^* = \pm 1.97$ <p><i>You do not need to show calc. command: [2nd][distr] invT(.025,249) = -1.9695</i></p>	<p><i>sketch graph</i></p> 
$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} =$ $10.2 \pm 1.97 \frac{4.2}{\sqrt{250}}$ $10.2 \pm 0.52 \qquad (9.68, 10.72)$	

CHECK with Calc: [stat] [tests] [8: TInterval] [Stats] xbar=10.2 Sx=4.2 n=250 C-Level=.95

Calc. Provides: (9.6768, 10.723)

- Interpret the interval in context
 We are 95% confident that the interval from 9.7 hours to 10.7 hours captures the true mean number of hours of homework that students at this school did in the last week.
- a) What can you conclude about the principal’s claim?
- The confidence interval (9.7-10.7) contains 10 hours but it also contains plausible values for μ that are less than 10 hours.
 - Therefore, the interval does not provide convincing evidence to support the principal’s claim that students spend at least 10 hours on homework per week, on average.