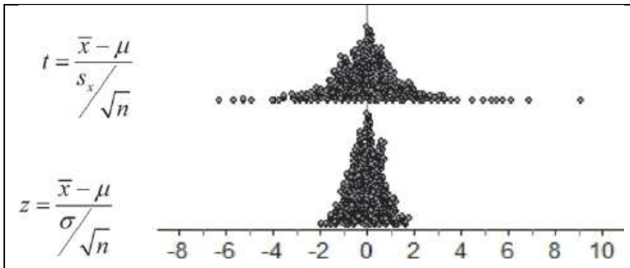
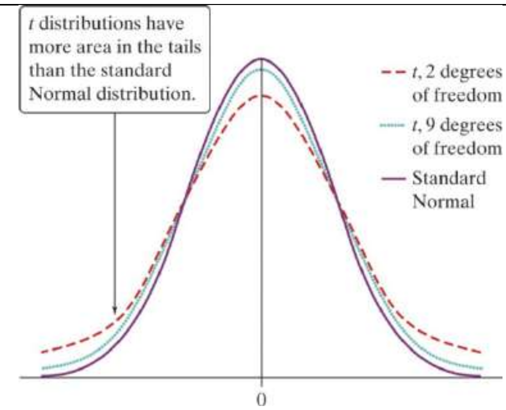
AP Statistics – 8.3b	Name: **KEY**
Goal: Understanding CI for Means when σ is UNKNOWN	Date:

I. **Definitions for: T-Distribution**

1) As we go through section 8.3, make notes on the differences between the normal distribution and the t distribution:





When we do NOT know σ, we can estimate it using the sample standard deviation s_x.

And we get a new statistics: $t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$

This new statistic "t" does not have a Normal distribution! The new ***t*** **distribution** has a ***different shape*** **than the standard Normal curve:**

- ✓ It is symmetric with a single peak at 0,
- ✓ However, it has much more area in the tails.

The new statistic **"t" is a standardized statistic**.
There is a **different t** distribution **for each sample size, specified by its degrees of freedom.**

The statistic will have approximately a t_{n-1} distribution as long as the sampling distribution is close to Normal.

- ✓ The density curves of the *t* distributions are similar in shape to the standard Normal curve.
- ✓ The spread of the *t* distributions is a bit greater than that of the standard Normal distribution.
- ✓ The *t* distributions have more probability in the tails and less in the center than does the standard Normal.
- ✓ As the degrees of freedom increase, the *t* density curve approaches the standard Normal curve ever more closely.

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 non-Normal, especially for larger sample sizes. The t procedures are not robust against outliers, however.

2) **Critical Values: Finding *t****

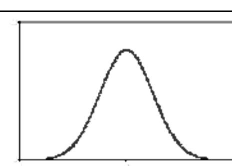
Example #1: Suppose you wanted to construct a 90% confidence interval for the mean of a Normal population based on an SRS of size 10. What critical value t* should you use?

a)Use Table B, to find t*

- **Solution: For n=10, use the line for df = 10 – 1 = 9 and the column with a tail probability of 0.05, the desired critical value is t* = 1.833.**
- **Solution: [2nd] [dist] [4:invT] area=0.05 df = 9 → t* = 1.833.**

a)Use Inverse *t* on the calculator, to find t* for CL's of 70%, 95%, 99%. How do they compare with z*?

- **70% invT(.15,9) t*= 1.10 vs. 70% z*=1.04**
- **95% invT(.025,9) t*= 2.26 vs. 95% z*=1.96**
- **99% invT(.005,9) t*= 3.25 vs. 99% z*=2.58**



II. Constructing a Confidence Intervals for μ

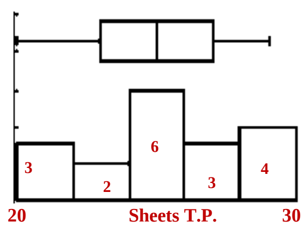
Example #2: ***"Can you spare a square?"*** As part of their final project in AP Statistics, Christina and Rachel randomly selected 18 rolls of a generic brand of toilet paper to measure how well this brand could absorb water. To do this, they poured 1/4 cup of water onto a hard surface and counted how 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]**
  μ = 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 hologram.
I recommend you look at the box plot to check for outliers.



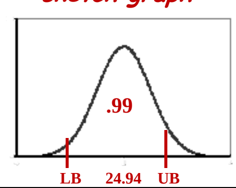
Explanation:

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
  - 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 → df=18-1=17**
    $t^* = \pm 2.898$

Calc. Command: invT(.005,17)= 2.89823

sketch graph



$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} =$$

$$24.9 \pm 2.898 \frac{2.86}{\sqrt{18}} =$$

24.94 ± 1.95 = **(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.**

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

Example #3: *"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.

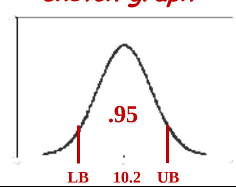
- Check conditions:
  - **Random:** The students were randomly selected. (SRS n=250)
  - **Independent:** Because we are sampling without replacement, we must check the 10% condition. It is reasonable to believe there are more than 10(250) = 2500 students at this large high school.
  - **Normal:** The sample size is large (n = 250) and satisfy CLT so we are safe using t-procedures.

- Construct the confidence interval
  - 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 → df=250-1=249**

    $t^* = \pm 1.97$

  You do not need to show calc. command: [2nd][distr] *invT(.025,249)= -1.9695*

sketch graph



$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} =$$

$$\mathbf{10.2 \pm 1.97 \frac{4.2}{\sqrt{250}}}$$

10.2 ± 0.52 **(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.

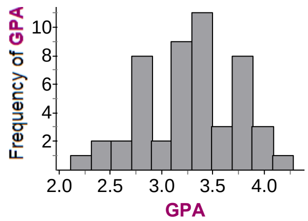
a) What can you conclude about the principal's claim?

- **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.**

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

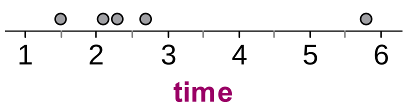
<u>Example #4</u>: *"GPA, Coffee, and SAT scores?"* Determine whether we can safely use a one-sample t interval to estimate the population mean in each of the following settings.

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.



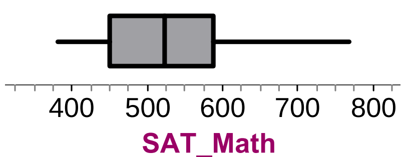
- *Since this is a random sample of 50 students is sufficiently large the t-statistic could be used.*
- *However, this was a convenience sample and a random sample from only your classes; and not from all students at your school.*
- *Therefore, this t-interval can NOT be generalize about the mean GPA for all students at the school.*

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.



- *Since the sample size is small and there is a possible outlier, we should NOT use a t-interval.*

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



- *Since the distribution is only slightly skewed and there is no apparent outlier, it is safe to use a t-interval.*
- *There is no evidence the distribution is NOT normal.*