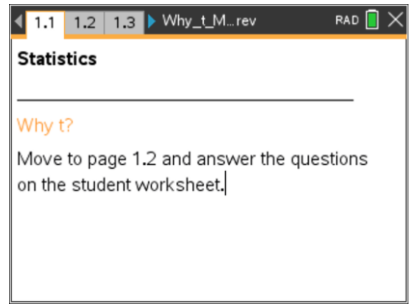
Why t?
Student Activity

Name ________________
Class ________________

Does having knowledge of a population's standard deviation affect the variability of the standardized statistic for the mean of a sample? This lesson examines the difference between knowing and not knowing the population standard deviation, comparing what actually happens during sampling to what we might expect to happen. The goal is to learn from any differences we observe.



A number of environmental disasters have occurred during the past few decades. Over the years, companies responsible for these problems have begun to display an awareness that they need to take substantial corrective action. After large accidents, part of "making it right" is to set aside a fund to pay business owners for revenue lost due to the accident. One way to determine a fair payment for lost business is to examine revenues for a sample of days during pre-accident times in order to estimate the mean pre-disaster daily revenue. Revenue during the time of reduced business levels following the accident can then be compared to the pre-disaster figure to determine a reasonable payout to the affected business owner.

To study the behavior of a common statistical solution, we will turn to a vastly oversimplified hypothetical example. A large international company, Oops, Inc., has just caused a major local disaster, resulting in loss of business for a small diner in a fishing village. Assume that the diner is open only for lunch, from 11:00 am until 3:00 pm, with daily revenues distributed approximately normally with mean $800 and standard deviation of $150. The distribution of revenue does not vary with the season of the year, etc. The owner of the diner will tell Oops that the mean is $800, but of course, neither the mean nor standard deviation is actually known to Oops. Oops needs to be convinced that the owner's estimate is correct before handing over large sums of money! This is the job of a statistician and why sampling must be done.

In the interest of simplicity, so that the key statistical idea will be clear, this activity will use very small samples. In practice, larger samples would be used. In fact, this activity will show one way in which using larger samples is a good idea!

Open the TI-Nspire document *Why_t.tns.*
Move to page 1.2.

Press ctrl ▶ and ctrl ◀ to navigate through the lesson.

Note: This activity involves generating a number of random samples from a population. In order to avoid having your results be identical to those for another student in the room, it is necessary to "seed" the random number generator. Read the instructions on Page 1.2 for seeding your random number generator, and carry out that seeding on Page 1.3.

Move to page 2.1.

The graph on Page 2.1 is a portion of the normal distribution curve having mean 800 and standard deviation 150, representing the idealized population of daily revenue from the diner.

1. Suppose you were to select days randomly, one at a time, and record the revenues for the diner for those days. Remembering what you know about normal distributions, write a sentence describing the values you would consider “typical” daily revenue values and values you would consider “unusual.” Explain your reasoning.

2. Click on the right arrow on Page 2.1 to generate a random day’s revenue. (The left arrow resets the screen.) Is the value you obtain typical or unusual? Compare values across the class to see how many unusual values occurred.

3. Repeat the process until you get what you think is an unusual revenue value. Leave that value displayed on your screen. How many days (clicks) did you have to examine before obtaining an unusual value? (Note the counter on the arrow.)

Move to page 2.2.

The upper panel of the page is exactly the same as page 2.1. The lower panel displays the standardized test statistic (z-score) calculated from the randomly-selected day’s revenue. Recall that this statistic is calculated using the formula $z = \frac{x-\mu_0}{\sigma}$.

4. a. Record the value of the standardized test statistic for your unusual sample from Question 3 into the table below. How does your standardized test statistic compare to those of your classmates?

b. Generate four more revenue values, and record the results in the table. Decide whether each revenue value is unusual or typical, and note the associated standardized test statistic value. Write a sentence to summarize how revenue value is related to unusual standardized test statistic values.

Revenue	Standard Statistic	Unusual or Not

Move to page 3.1.

The curve on this page is the same as that for the population on previous pages. However, the right arrow now generates a sample of revenues from three randomly selected days and displays the three sample values and the sample mean.

5. Write a sentence to describe the values that represent typical means for samples of size three. What mean values you would consider unusual? Compare your new description of “typical” to the one you wrote in Question 1 and briefly explain any differences.

6. Use the arrow to generate samples, observing whether the sample mean appears to be typical or unusual. Stop when you have a sample mean that seems somewhat unusual, and leave that sample displayed on your screen.
   a. How many samples did you have to examine before obtaining an unusual mean?

   b. Did the variability in the sample means that you observed seem consistent with your description in Question 5 of typical means? Explain briefly.

Move to page 3.2.

The top work area of this page is exactly the same as Page 3.1. The lower panel displays the standardized test statistic (z-score) associated with the sample mean. Recall that the standardized test statistic for sample mean is calculated using the formula $z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$.

7. a. Record the value of the standardized test statistic for your unusual sample from Question 6 in the table below. How does your standardized test statistic compare to those of your classmates? How do the standardized test statistic values for the class compare to those you observed in Question 4a?

Mean Revenue	Standard Statistic	Unusual or Not

   b. Generate three more samples, and add the results to the table on the previous page. In each sample, decide whether the mean revenue value is unusual or typical and note the associated standardized test statistic value. Write a sentence to summarize how samples are related to unusual standardized test statistic values.

   c. Compare your table from Question 4 to this new table. How are decisions about "unusual or not" alike? How are they different?

Look back at the formula before Question 7. Notice that the calculation of the standardized test statistic (z-score) requires knowing the population mean, the sample mean, the population standard deviation, and the sample size. In reality, the actual population mean and standard deviation of daily revenues are not known to Oops, and Oops does not want to just take the diner owner's word for it.

8. The owner's claimed value for the mean becomes the null hypothesis of a hypothesis test, so it is available for use in the z-score calculation. However, not knowing the population standard deviation makes the calculation of the z statistic impossible, and makes it difficult to determine if the value is unusual or not.
   a. Explain how you might be able to get a reasonable approximation for the population standard deviation to use in the calculation of the standardized test statistic in this situation.

b. Write a sentence predicting how calculating the standardized test statistic using the standard deviation of the sample will affect the set of standardized test statistic values that are judged as unusual.

Move to page 3.3.

The top work area of the page is again exactly the same as Page 3.1. The bottom work area displays the new standardized test statistic associated with the sample mean, now adjusted to account for the fact that the population standard deviation is unknown. This adjusted standardized test statistic for sample mean is calculated using the formula $z\_adj = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$.

9. The last sample you selected in Question 7 is still displayed.
   a. Identify the change you see in the formula for the standardized score. Why is it necessary?

   b. Compare the new standardized test statistic to the one displayed on Page 3.2. Would you call the new value unusual? Explain briefly.

   c. How does your new standardized test statistic compare to those of your classmates? How do the new standardized test statistic values for the class compare to those you observed in Question 7?

10. Generate more samples, trying to get as large (positive or negative) a standardized test statistic as you can, keeping track of the number of clicks it takes to get it. Stop when you think you have an extremely unusual value.
    a. How does your new unusual standardized test statistic compare to the unusual values your class got in Question 7? How many clicks did it take you to find it?

    b. Compare this new “unusual sample” to those you wrote in your table in Question 7. Is this new sample’s mean noticeably farther from the population mean than those in your table? Check samples from your classmates, too.

c. Consider both how often “unusual” values of the new statistic appear and how “unusual” they are. Describe how well the Empirical Rule for normal distributions seems to relate to the distribution of these new standardized test statistics.

Move to page 3.4.

11. The left work area of Page 3.4 shows your most recent samples (without their means), stacked with newer samples at the top and older samples at the bottom. The right panel displays the corresponding values of the adjusted standardized test statistic. Your most recent sample—the unusual one—is displayed at the top in each panel. Study this display and make a conjecture about what causes unusual values of the adjusted standardized test statistic. Remember, the new formula for the standardized test statistic is $z\_adj = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$. Use the formula to explain your conjecture.

Move to page 4.1.

The top work area of Page 4.1 is the same as Page 3.1. The bottom work area of this page displays the standard normal distribution—a normal distribution with mean 0 and standard deviation 1. The right arrow generates and displays a sample of size $n = 3$, together with its mean, and adds that information into the lower work area. The lower work area then shows the simulated sampling distribution of the standardized test statistic calculated using the standard deviation from the sample as an approximation for the standard deviation of the population as was done for Questions 9 -11.

12. Generate a number of samples, and watch the standardized test statistic value accumulate. Describe how well the standard normal curve predicts this simulated sampling distribution.

Definition: The distribution for this new statistic is called the ***t* distribution.**

Name _______________
Class _______________

Move to page 4.2.

Page 4.2 is identical to Page 4.1 except that the bottom work area displays the *t* distribution with 2 degrees of freedom.

13. Generate a number of samples, and watch the standardized test statistic value accumulate. Describe how well the t distribution predicts this simulated sampling distribution.

14. Suppose that you have been hired as a statistical consultant to Oops as they process claims such as the one from the diner. Each business owner will submit a claim stating their average revenue, and Oops will have to use a sample from past business records to decide whether that claimed average seems consistent with (typical for) that sample. Which distribution (normal or t) would you recommend that they use in making that decision. Why?