



10- Percent Rule

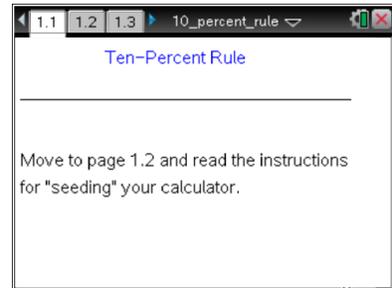
Student Activity

Name _____

Class _____

Open the TI-Nspire document *10_Percent_Rule.tns*.

This lesson focuses on sampling distributions of sample means, paying special attention to the standard deviation of those sampling distributions.



Formal statistical inference—hypothesis tests and confidence intervals—on means requires values for both the mean and the standard deviation of the sampling distribution of the sample mean. We will consider how the standard deviation of the sampling distribution of sample means change when samples are taken without replacement instead of with replacement.

Move to pages 1.2 and 1.3.

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

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, then carry out that seeding on Page 1.3.

Move to page 2.1.

- Page 2.1 displays the calculations of the **mean** and **standard deviation** of a population containing exactly six elements, namely {3, 6, 9, 12, 15, 18}. Record these two values for use throughout this activity.

Note: The displayed mean is exact, but the standard deviation has been rounded to three decimal places.

Mean: _____ Standard Deviation: _____

Move to page 2.2.

The arrows on the left side of Page 2.2 will generate samples of size three from the population on Page 2.1. (The arrows on the right side of the page reset the page, erasing all samples, in case you want to start over.)

- For this problem, samples are taken with replacement. Thus, for example, the ordered triple (3, 3, 15) could be a sample, as could (3, 3, 3)



10- Percent Rule Student Activity

Name _____

Class _____

2.
 - a. Click once on an arrow on the left side of Page 2.2 to generate a sample of size three from the given population, chosen with replacement. What do the vertical line in the upper plot and the single dot in the lower plot represent?
 - b. Click about 10 more times on the same arrow to generate more random samples of size three from the given population. Describe the resulting plots.
 - c. Recall that the sampling distribution of a statistic is the distribution of that statistic from all possible samples. You have only seen a few values from the sampling distribution of means of samples of size three from the given population, so your distribution in the lower work area of Page 2.2 is very incomplete. Use your answers from Question 1 and formulas from earlier in the course to calculate the exact mean and standard deviation of the complete sampling distribution of sample means of samples of size three.

Move to page 3.1.

Page 3.1 continues building the sampling distribution of sample means from samples of size three from the given population. Again, samples are taken with replacement, so triples such as (3, 3, 15) might occur. Also, as before, the arrows on the left side of the page generate samples and the arrows on the right side of the page reset so that you can start over.

3.
 - a. Click the arrow once and describe the information displayed in the plot.
 - b. Choose the highest or lowest of the displayed values (dots) in the plot on Page 3.1. State the displayed value, and identify at least one triple (sample) that could have produced it. (You might want to refer back to question 1 to recall the population from which the samples are being drawn.)
 - c. Compare the mean and standard deviation displayed below the dot plot to the values you calculated in Question 2c and explain any differences you see.



10- Percent Rule Student Activity

Name _____

Class _____

4. a. Continue clicking the arrow on Page 3.1 until the entire sampling distribution of sample means is displayed. Describe the shape of the sampling distribution, and explain how you know that it is not a normal distribution.
- You will know the entire distribution is displayed when repeatedly clicking the arrow does not change the distribution anymore.
- b. How do the mean and standard deviation of the sampling distribution compare to those you calculated in Question 2c?
5. a. Consider the three or four most extreme sample means (dots) displayed on Page 3.1. Identify the samples from which they came.
- b. For each sample that you identified in part a, what fraction of the entire population was included in that sample?
6. a. Suppose now that the samples from the population were taken without replacement. That is, no element of the population is allowed to appear more than once in any given sample. Describe how you think that change in sampling procedure would change the extreme values in the sampling distribution of sample means.
- b. What fraction of the population is included in any sample of size three taken without replacement?



7. Standard deviation indicates something about the differences among members of the distribution it describes. Based on your work in Questions 5 and 6, discuss differences in the standard deviation of the sampling distribution of sample means that you expect between samples taken with replacement and samples taken without replacement.

Move to page 4.1.

Page 4.1 again builds a sampling distribution of sample means from samples of size three from the same population. However, this time samples are taken without replacement, so triples such as (3, 3, 15) can never occur. Note that only the means of the samples are displayed, not the elements that make up each sample.

8. a. Click the arrow on Page 4.1 until the entire sampling distribution of sample means is displayed. Describe the shape of the sampling distribution, and explain how you know that it is not a normal distribution.
- b. How do the mean and standard deviation of the sampling distribution compare to those you calculated in Question 2c?
- c. How do the standard deviations from the sampling distributions on Pages 3.1 and 4.1 compare? Are the results consistent with what you predicted in Question 7?

The idea of **residual** runs through all of statistics. Remember that a residual is the difference “Observed – Expected.” Here we can think of the actual standard deviation of the sampling distribution

as the “Observed” value and the value predicted by the formula $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ as the “Expected” value.

Since a difference of five units can be “large” or “small” in different contexts, one useful way to judge the size of a residual is to express it as a “relative difference” or fraction of the Expected value. That is,

$$\text{compute Relative Difference} = \frac{\text{Observed} - \text{Expected}}{\text{Expected}} .$$



9. Compute the relative difference for the standard deviation of the sampling distribution you generated on Page 4.1, and interpret what it means.
10. Based on your answers in Questions 8 and 9, name at least one advantage and one disadvantage to sampling without replacement.

Move to page 4.2.

Page 4.2 behaves exactly as Page 4.1 did, except that now the center set of arrows will change the sample size to any values you prefer. As with Page 4.1, samples are again taken without replacement.

11. a. Record the standard deviation information from Page 4.1 into the table below.

Sample Size	SD by Formula	Actual SD	Relative Difference
1			
2			
3			
4			
5			
6			

- b. Set the sample size to a value other than 3, then generate the corresponding sampling distribution of sample means. Record the standard deviation information for that distribution in the table above.
- c. Complete the summary table for sample size and standard deviation of sampling distributions by choosing other sample sizes and generating their corresponding sampling distributions. Describe how sample size is related to the “agreement” of the actual standard deviation of

sampling distributions and the standard deviation predicted by the formula $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$.



Move to page 5.1.

12. Page 5.1 displays a new, larger population. The arrow allows you to change to a new population. Using the table below, record descriptions of each possible population.

Population Size	Mean	Standard Deviation	Shape

Move to page 5.2.

Page 5.2 behaves exactly as Page 4.2 did. The population from which samples will be taken is the one you selected on Page 5.1.

13. a. Page 5.2 generates samples without replacement and plots the means of those samples. Thus, as before, you will build a simulated sampling distribution. However, now the population is larger, so it will no longer be possible to generate all possible samples. That is, you will have only an approximate sampling distribution. How do you think that will affect your standard deviation calculations?
- b. In carrying out this kind of simulation, what might we do to improve our results?

At this point, your teacher will split your class up into groups and assign each group a specific population to investigate. Within your group, divide up the work however you think will be most efficient.

Tech Tip: Sample means will be calculated and displayed in groups of 100 at a time. Because of the amount of calculation going on in generating those samples, there will be several seconds of delay before your display is updated. Be sure to watch for the plot to update before your next click. Be patient.



10- Percent Rule

Student Activity

Name _____

Class _____

14. Return to Page 5.1, and use the arrow to set the population to the one assigned to your group.
- Then move back to Page 5.2, and use the lower arrow on the left side of the page to select a sample size for which to generate the simulated sampling distribution. (Each person in your group can use a different sample size.)
 - Click the upper arrow on the left side of the page to generate sample means.
 - Continue until you have 1500 samples. (Remember, wait for the plot to update between clicks.)
 - Record the information for your group’s simulations in the table below. Describe any patterns you see.

Pop Size	Sample Size	SD by Formula	Simulated SD	Real Difference
	2			

15. Using the results of your group’s work in Question 14 (investigating additional sample sizes if needed), determine the largest sample size for which the simulated standard deviation and the “formula” standard deviation differ by no more than 5%.

Move to page 6.1.

16. a. The left work area of Page 6.1 shows two columns of a spreadsheet. Collect the class results from Question 15 into this page, entering the appropriate sample size in the cell next to the population from which it was obtained. Exactly what does the first pair of values (20, ___) tell you about sampling distributions?
- b. Describe the plot in the right work area.
- c. Click the arrow in the scatterplot to display the graph and equation of the least squares regression line for your class data. Interpret that equation and, if possible, relate your interpretation to the title of this activity.



Extension

Theory says that the standard deviation of the sampling distribution of sample means for samples

taken without replacement is actually $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \sqrt{1 - \frac{n-1}{N-1}}$.

This is the usual formula but with an additional factor: the quantity $\sqrt{1 - \frac{n-1}{N-1}}$ is called the **Finite Population Correction (FPC)** factor.

17. Return to your data table in Question 14, and verify the accuracy of the new formula by carrying out the indicated calculations and comparing the results to your simulation data.

18. Use algebra on the FPC factor to find an exact rule for when the relative difference in standard

deviations will be more than 5% of $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$.