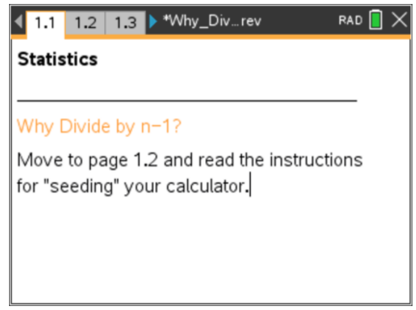
Why Divide by *n*-1?

Student Activity

Name ________________

Class ________________

Open the TI-Nspire document *Why_Divide_by_n-1.tns.*



One of the goals of selecting a sample from an unknown population is to estimate parameters of that population. One of these parameters is variance; the square root of the variance produces the standard deviation. This activity investigates the difference in estimating the population variance calculated using a divisor of *n-1* versus *n*.

.

Move to page 1.2.

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

Tech Tip: Page 1.2 gives instructions on how to seed the random number generator of the Nspire. Page 1.3 is a calculator page for the seeding process. Carrying out this step will prevent everyone from generating identical data. (Syntax: RandSeed #, where # should be a number unique to each student.)

Tech Tip: When you move the cursor over a point in a dotplot, the data value associated with the dot is displayed. Clicking on a vertical line in this activity will display the value plotted on the horizontal axis. To de-select a point, click in a white space on the screen.

Move to page 2.1.

Click on the arrow to set up the activity.

Move to page 2.2.

1. The arrow generates a sample of size three, with replacement, from an unknown but fixed population. Generate a sample, and note the sample variances calculated by dividing by *n* and by *n-1*.
   a. How do the two values differ and why?

b. Record the values from question a into the table below. Then generate three more samples, and record the sample values and variances in the table. Which method do you think will give a better estimate for the population variance and why? (Remember that variance is population squared.)

Sample values	Variance (*n* divisor)	Variance (*n-1* divisor)

Move to page 3.1.

To determine whether one estimator is better than another requires knowing the population and investigating the behavior of the estimators across all possible samples. The population from which the samples in question 1 were drawn is {3,6,9,12,15,18}. The population variance, which is calculated with N=6 as a divisor, is 26.25.

2. The arrow will generate eight possible samples of size three, drawn with replacement from the population, and calculate the variance of each sample. Those variances, calculated using both divisors, are plotted in the dot plots on Page 3.1. Use the arrow to generate a set of eight samples.

   a. What does a data value in each plot represent? Give an example to illustrate your answer.

   b. Make a conjecture about what the vertical line in each plot represents. Then check your conjecture by inspecting the data values in each plot.

   c. Continue to generate groups of samples until you have 104 samples observing what happens to the means of the sample variances calculated with both methods. Describe how the two means compare.

   d Continue to generate samples until you have the sampling distribution of all 216 possible sample variances, calculated using both divisors, for samples of size three drawn with replacement from the given population. How is it possible that some samples have variances of 0 for both methods?

©2011 Texas Instruments Incorporated

e. Use shape, center and spread to compare the sampling distributions for the sample variances found by the two methods. (Remember that the population variance is 26.25)

Move to page 4.1

In many cases, a sample is drawn without replacement because, for a variety of reasons, it is impossible to replace the sample elements. The following questions investigate how to estimate the population variance when a sample is drawn without replacement.

The arrow will generate all possible samples of size three, drawn without replacement from the population in Question 1. The sample variances calculated using both divisors are plotted in the dot plots on Page 4.1. Remember if the divisor were $N-1$, or 5, the result would be 31.5. The population variance, which is calculated with $N=6$ as a divisor, is 26.25.

3. a. Make a conjecture about which method of calculating the variance for a sample will give the best estimate of the population variance. Explain your reasoning.

   b. Use the arrow to generate the sampling distribution of the sample variances, and observe what is happening to the means of the sample variances in each of the distributions. Describe what you observed.

   c. Use shape, center, and spread to compare the sampling distributions for the variances found by the two methods and to check your conjecture in 3a. (Remember that the population variance, which is calculated with $N=6$ as a divisor, is 26.25. If the divisor were $N-1$, or 5, the result would be 31.5.)

4. An *unbiased estimator* for a population parameter is one for which the average of the values of the sample statistic from all possible sample values equals the value of the population parameter. A biased estimator for a parameter is one whose values on average over- or under- estimate the true value of the parameter.

   a. When sampling with replacement, which estimator, the variance calculated by dividing by n or the variance calculated by dividing by $n-1$, will be an unbiased estimator for the population variance? Explain your reasoning.

b. Lin claims that the *n-1* sample variance is better than the *n* sample variance. Do you agree with her? Why or why not.

Move to page 5.1

Using the arrow will select a sample size, *n,* and the display will show 100 random samples generated from a normal population with mean 10 and standard deviation 2 for the sample size you selected. The screens display the simulated sampling distributions of the sample variances calculated for each sample size using divisors of *n* and of *n-1*.

5. a. If the standard deviation is 2, what is the variance?

   b. Make a conjecture about how the means of the two distributions of sample variances will compare as the sample size increases. Explain why your conjecture seems reasonable.

   c. Use the arrow to choose a sample size of 10. Click on the vertical line through the mean in each display and note the values of the means of the two simulated sampling distributions. Repeat this process for at least three other sample sizes greater than 40. Do the results support your conjecture in part a? Why or why not?

6. Describe how sample size seems to affect the difference in approximating the population variance using a divisor of *n* or a divisor of *n-1*.

©2011 Texas Instruments Incorporated