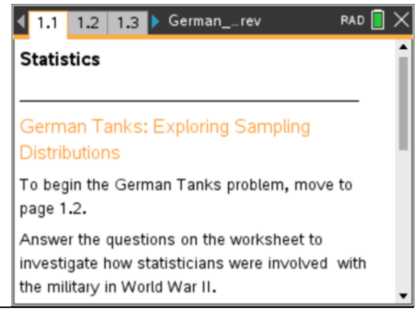
German Tanks: Exploring Sampling Distributions Name ____________
Student Activity Class ____________

Open the TI-Nspire document German_Tanks:_Exploring_Sampling_Distributions.*tns.*

Often real life challenges indicate the importance of what we study. The following activity is based on a genuine problem faced by the military during World War II. The analysis used to solve that problem is developed for you here.



Move to page 1.1.

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

Similar to the challenge faced by statisticians during World War II, you will be provided a sample of numbers representing serial numbers of German war tanks. Assume that the serial numbers were consecutively numbered from 1 to the number of the last manufactured tank. Your goal is to determine the number of manufactured tanks, that is, the largest serial number.

Move to page 1.2.

Follow the directions on Pages 1.2 and 1.3 to seed the random number generator. This will result in different random samples for each member of the class.

Move to page 1.4.

Click the right arrow on Page 1.4 to generate a random sample of five serial numbers in the column labeled **capture** and to see a dotplot of the sample. These values represent your sample of captured serial numbers.

1. a. The smallest serial number would be 1. Did you capture that serial number? The largest serial number is not known. Do you think you captured it? Why or why not?

   b. The goal of this investigation is to create a *rule* to estimate the largest serial number from the population of tanks. Before you compare samples, think about how you might use the serial numbers you have to make an estimate of the number of tanks. Describe in words the rule you would apply to your sample to estimate the number of tanks.

   c. Based on your sample of five numbers, create a formula to estimate the largest serial number that represents the rule you described above.

Move to page 1.5.

Page 1.5 is designed for you to enter your rule into cell A1 of the spreadsheet.

2. Move the cursor over cell A1 and click. A1 will now be highlighted. Type an "=" into cell A1. The equal sign indicates that a formula (or rule) will follow. Type in your rule, move the cursor to a position outside the parentheses, and press enter. The estimate of the largest value based on your rule will be displayed.
   - As an example, consider the rule: "Twice the value of the mean of the sample." This rule would require the following formula to be entered:

$$= 2 \cdot \text{mean(capture)}$$

Note: It is important to refer to the list **capture** when you want values such as the mean, median, max, min, or sum of the sample.

Move to page 1.6.

3. Examine Page 1.6.
   a. What does the **capture** dotplot represent?

   b. What does the **my_estimates** dotplot represent?

4. Share your estimate of the number of tanks with other students. Were any of the estimates very different from yours? What might explain the difference?

Suppose you had a different sample of five serial numbers.

5. Use the right arrow to generate a new sample, assuming all of the tanks were back in use by the German army when this new sample of five tanks was captured. Determine the estimate of the number of tanks from this new sample using your rule. Compare your first and second estimates.

It might be useful to know how your rule will work for different samples.

6. a. Use the right arrow to generate estimates for at least 30 samples.
   b. Examine the dotplot of **my_estimates.** Describe your rule and the distribution of estimates for the total number of tanks using your rule.

Move to page 1.7.

7. The spreadsheet displayed on Page 1.7 contains the estimates for the total number of tanks using four separate rules. The estimates derived from your rule are in Column D and are labeled **my_estimates**. Click in the cell at the top of each column. Describe the rule that generates the estimates in that column.
   Column A:

   Column B:

   Column C:

Move to page 1.8.

8. Four dotplots are displayed on Page 1.8. Each dotplot displays the estimates derived from the four rules you saw on Page 1.7.
   a. The vertical line represents the actual number of tanks in the population for a particular month. Click on the line. How many tanks does this simulation assume the Germans manufactured this month?

   b. Hover over each of the dotplots. Record the minimum and maximum of the distribution for each rule in the table below. (Note that clicking in a white space will deselect a point.)

Rule	Minimum value of sample	Maximum value of sample
maxplus5s		
twicemedians		
multimaxs		
my_estimates		

An estimate generated by a rule or formula from a sample is considered a **statistic**. A rule that consistently generates statistics less than the true value or consistently greater than the true value is an example of **bias**. One outcome of bias would be a dotplot in which most of the dots are located less than the true value. This dotplot indicates that estimates from this rule are consistently less than the true value. Similarly, a dotplot what would have most of the dots located greater than the true value would be another example of bias. Biased rules should be eliminated from consideration when looking for the best estimate.

9. a. Do any of the dotplots suggest the statistic is biased?

   b. Why do you think this rule generates biased results?

As **maxplus5s** demonstrates a bias, we will eliminate it from consideration as the best rule of the four rules under investigation. At first glance, the remaining three rules seem to produce almost the same estimates, and it is hard to tell which ones might be better at estimating the total number of tanks. One way to help decide is to find the "error" in estimating the total number of tanks by subtracting the 245 tanks, or the parameter of the population, from an estimate.

10. Consider subtracting 245 from each of the estimates.
    a. What would a negative difference represent?

    b. What does a value close to zero represent?

    c How might you use the errors to figure out whether a rule is good or bad?

Move to page 1.9.

The spreadsheet on Page 1.9 displays the absolute values of the errors for each rule.

11. If one rule were selected as a better rule for estimating the true value, how would the values displayed in the spreadsheet for this rule compare to the values of a rule that was not as good?

Move to page 1.10.

12. Examine the dotplots of the absolute value of the errors for each rule.
    a. How would you describe the variability displayed by each dotplot?

    b. Knowing that you only have one sample from which to estimate the total number of tanks, which rule—that does not appear biased—would you use? Explain your answer.

Move to page 1.11.

Page 1.11 provides a sum of the absolute value of the errors for each of the rules.

13. Do the sums support your previous conjectures about the better rules? Why or why not?

14. A rule that provides estimates with a small amount of variability around the actual number of tanks would seem to be a good rule to select. Why is a small amount of variability important?

The intelligence reports indicated that the number of tanks manufactured over a certain time period was actually over a thousand, while the statisticians indicated it was closer to 250. (See the table below.) It was verified after the war that the statistician's estimates were more accurate! Many military decisions were made based on an incorrect estimate of the number of tanks.

After the war, production records from the ministry of Albert Speer showed the actual number to be 255. Estimates of the number of tanks for some specific months for the statisticians, the spies in the field, and the German records are below.

Month	Statistical estimate	Intelligence estimate	German records
June 1940	169	1000	122
June 1941	244	1550	271
August 1942	327	1550	342

Source: http://en.wikipedia.org/wiki/German_tank_problem