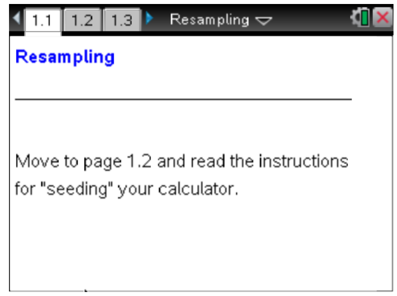
Resampling
Student Activity

Name ______________

Class ______________

Open the TI-Nspire document *Resampling.tns.*

All hypothesis tests require information about the sampling distribution of the sample statistic being used—comparison of the actual sample to the sampling distribution is the key step. Traditional tests require the mean, standard deviation, and shape of the sampling distribution to be known. This required information about the sampling distribution comes from mathematical theory, relying heavily on assumptions of an underlying normal distribution. However, it is possible to carry out valid hypothesis tests even when there is no information about the underlying distributions.



Move to pages 1.2 and 1.3.

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

A class of 18 students decided to test the relative effectiveness of two brands of mosquito repellant. In particular, their research question was “Is Brand B more effective than Brand A in preventing mosquito bites?” Note that this creates a hypothesis test in which the alternative hypothesis is one-sided, favoring Brand B. Each member of the class tossed a coin to decide which brand to test, applied the appropriate repellant according to the manufacturer’s instructions, then joined the rest of the class at the sports stadium for the evening’s athletic event. Times until the first mosquito bite were recorded (in minutes).

1. Do you think that all the students who test brand A will get exactly the same protection times? If “yes,” explain why. If “no,” identify possible sources of variability.

2. If the two brands protect, on average, equally effectively, do you think that the student sample data will show identical mean times before the first bite? If “yes,” explain why. If “no,” tell what could cause the difference.

Move to page 2.1.

3. Page 2.1 displays the data obtained from the students. Based only on the comparative dotplots, do you believe that the average protection times for the two brands really differ, or could these results be the result of chance variation? Explain your reasoning.

©2011 Texas Instruments Incorporated

4. a. The table below provides space to write the values of the time in minutes before the first bite for each of the 18 students. Hover over each point in the dotplots on Page 2.1 to obtain the numerical values of all the data. Record each value in the top row below, and record whether it was brand A or brand B in the second row.

Protection time in minutes

   b. Click the arrow on Page 2.1 to display the mean protection times of the two brands as vertical lines on the plot. Click on each line to see its numerical value. Record these values.

   c. Use the dotplot or your table of values from part a to determine the median protection time of each brand. Record your results.

Every hypothesis test compares data from an observed sample or treatment allocation to a sampling distribution of what would happen if the null hypothesis were true.
Suppose the null hypothesis is ‘no essential difference in the effectiveness of the two repellants, at least as far as protection time is concerned.’ We need to obtain a sampling distribution of what can happen under these conditions.

5. a. If the null hypothesis is true, explain why it would be reasonable to think of the two sets of protection times as really being just one data set.

   b. If the null hypothesis is true, what would explain any apparent differences between the two plots in the students’ sample data?

If the two brands are essentially alike (which is our null hypothesis), then the only mechanism leading to any particular protection time’s being associated with a particular brand is the coin toss that the class did to decide which repellant to use. That is, those 18 times would still have been the data, but they could have been “reassigned” as far as A and B are concerned, had the coin tosses come out differently. So, looking back at your table in 4a, you could use coin tosses to assign brands A and B to those same numbers, and you would have another possible outcome of the class’s experiment.

From the point of view of a hypothesis test, the question now is whether the observed data the 18 students did get is “typical” or “unusual” when compared to all the other sample data they could have obtained if the null hypothesis were true.

Move to page 3.1.

The arrow on the left side of Page 3.1 performs the equivalent of 18 coin tosses and assigns the corresponding As and Bs to the same numbers you listed in 4a. Thus, each new click gives another possible set of data.

6. a. Click the arrow once, and then click on each vertical line to see the mean of each of the two sets. Record the difference between means, $\bar{x}_B - \bar{x}_A$. Do the new comparative dotplots indicate that the average protection times for the two brands differ? Explain your reasoning.

   b. Click the arrow a few more times. Record the difference of means, $\bar{x}_B - \bar{x}_A$, for each new rearrangement. Decide whether each new pair of comparative dotplots indicates that the repellants’ average protection times differ by brand. Explain your reasoning.

   c. Think about how the new reassignments are being made. How is the major assumption behind these reassignments related to the null hypothesis?

The logic of hypothesis testing requires that we compare some statistic from the observed data to the sampling distribution of that same statistic obtained when the null hypothesis is known to be true. If the observed statistic differs from what is typical in the sampling distribution for a true null hypothesis, then either the sample was unlucky (unusual just by chance) or the null hypothesis is not actually true.

Let the test statistic be the difference of sample means, $\bar{x}_B - \bar{x}_A$, but make no assumptions about the shape of the population distribution. If the two populations are essentially alike (the null hypothesis), then the simulation you carried out for Question 6 produces exactly the samples that lead to the necessary sampling distribution.

©2011 Texas Instruments Incorporated

Move to page 3.2.

Page 3.2 displays a dotplot of the differences $\bar{x}_B - \bar{x}_A$ for the samples you examined on Page 3.1.

Hover over the points on the plot to verify that they agree with what you recorded in Question 6.

7. a. Click the "draw" arrow at the top left of Page 3.2 to continue generating values in the simulated sampling distribution of $\bar{x}_B - \bar{x}_A$ until you have 100 simulated samples. (Note the counter in the top panel.) Remember, this simulation uses a process that guarantees the null hypothesis is true—the only difference between As and Bs is random assignment. Describe the simulated sampling distribution of differences of means, including noting the range of values that seem "typical."

   b. Click the "show diff" arrow at the top center of Page 3.2 to display the difference of means from the actual class data, both as a vertical line in the plot and as a numerical value above the plot. Does the value from the class seem unusual when compared to the possible values that make up the sampling distribution for a true null hypothesis? Explain.

   c. Based on your comparison of the class data to the simulated sampling distribution, state your conclusion for the test of our null hypothesis.

8. One advantage of this method of completing a hypothesis test is that it can be applied to statistics other than means. For example, you observed the medians of the two data sets in 4c. Describe how you could conduct a hypothesis test of these data using medians.

Move to page 4.1.

9. Click the arrow a few times. Record the differences of medians, $\tilde{x}_B - \tilde{x}_A$, for each new rearrangement. Decide whether each new pair of comparative dotplots indicates that the populations' average times before the first bite for the two brands differ. Explain your reasoning.

©2011 Texas Instruments Incorporated

Move to page 4.2.

10. a. Click the arrow at the top left of Page 4.2 to continue generating values in the simulated sampling distribution of $\tilde{x}_B - \tilde{x}_A$ until you have 100 simulated samples. Describe the simulated sampling distribution of differences of medians.

    b. Click the arrow at the top center of Page 4.2 to display (both as a vertical line and a numerical value) the difference of medians from the actual class data. Does the value from the class seem unusual when compared to the possible values that make up the sampling distribution for a true null hypothesis? Explain.

    c. Based on your comparison of the class's observed sample median to the simulated sampling distribution, state your conclusion for the test of the null hypothesis.