Lesson 27: Ruling Out Chance

Student Outcomes

- Given data from a statistical experiment with two treatments, students create a randomization distribution.
- Students use a randomization distribution to determine if there is a significant difference between two treatments.

Lesson Notes

MP.4 In this lesson, students will perform the five steps of a randomization test to assess if the tomato data provide evidence that the nutrient was effective. All of the steps have been practiced or indirectly discussed in the previous two lessons; however, now everything will be detailed and performed together in context.

In the second half of the lesson, students will begin the important task of developing their own randomization distributions—first manually, then using technology. The use of technology is strongly encouraged to assist in the steps of a randomization test, specifically a web-based randomization testing applet/calculator available at http://www.rossmanchance.com/applets/AnovaShuffle.htm.

This lesson may require slightly more time than the two previous lessons, but it is important that students have a solid understanding of these steps (and some familiarity with the technology) prior to moving on to the capstone experiences in the subsequent lessons.

Classwork

Opening (5 minutes)

Before the start of the lesson, ask students to recall what they have learned from earlier lessons, either in writing or by speaking to a neighbor, and then share with the class.

- To recap some highlights of the past few lessons:
  - *You have learned how to carry out repeated random assignments to develop a randomization distribution for the difference in means of two groups.*
  - *You have also established a method for determining if a specific difference value in this distribution is considered extreme or significant.*
  - *When a specific difference is significant, it is evidence that the specific difference did not occur by chance alone and may have, in fact, occurred due to something other than chance; for example, a treatment may have been imposed in one of an experiment's groups.*

Exercises 1–4 (15 minutes): Carrying Out a Randomization Test

Before students begin the exercises, use the following question to encourage discussion about what is needed to carry out a randomization test. Allow for multiple student responses.

- Think back to the work done with the tomato experiment. Statisticians agree there are five important parts to the process of using randomization to reach a conclusion in an experiment. Using what we've learned, see how many you can come up with on your own.

Work through Exercises 1–4 as a class. Pose one question at a time, and allow sufficient time for discussion of each step for carrying out a randomization test.

Exercises 1–4: Carrying Out a Randomization Test

The following are the general steps for carrying out a randomization test to analyze the results of an experiment. The steps are also presented in the context of the "tomato" example of the previous lessons.

***Step 1*—Develop competing claims: No Difference vs. Difference**

One claim corresponds to no difference between the two groups in the experiment. This claim is called the *null hypothesis*.

- **For the tomato example, the null hypothesis is that the nutrient treatment is not effective in increasing tomato weight. This is equivalent to saying that the average weight of treated tomatoes may be the same as the average weight of non-treated (control) tomatoes.**

The competing claim corresponds to a difference between the two groups. This claim could take the form of a "different from," "greater than," or "less than" statement. This claim is called the *alternative hypothesis*.

- **For the tomato example, the alternative hypothesis is that the nutrient treatment is effective in increasing tomato weight. This is equivalent to saying that the average weight of treated tomatoes *is greater than* the average weight of non-treated (control) tomatoes.**

Scaffolding:

- Students may want to create a Frayer diagram and/or rehearse with "null hypothesis" and "alternative hypothesis."
- After the 5 steps are introduced, consider asking students to restate, write, or draw their own representations of them.
- For advanced students, ask them to use the 5 steps to design their own randomized experiment to answer a question of interest.

1. **Previously, the statistic of interest that you used was the difference between the mean weight of the 5 tomatoes in Group A and the mean weight of the 5 tomatoes in Group B. That difference was called "Diff." "Diff" $= \bar{x}_A - \bar{x}_B$. If the treatment tomatoes are represented by Group A and the control tomatoes are represented by Group B, what type of statistically significant values of "Diff" would support the claim that the average weight of treated tomatoes *is greater than* the average weight of non-treated (control) tomatoes: negative values of "Diff," positive values of "Diff," or both? Explain.**

   For the tomato example, since the treatment group is Group A, the "Diff" value of $\bar{x}_A - \bar{x}_B$ is $\bar{x}_{Treatment} - \bar{x}_{Control}$. Since the alternative claim is supported by $\bar{x}_{Treatment} > \bar{x}_{Control}$, we are seeking statistically significant "Diff" values that are positive since if $\bar{x}_{Treatment} > \bar{x}_{Control}$, then $\bar{x}_{Treatment} - \bar{x}_{Control} > 0$. Statistically significant values of "Diff" that are negative in this case would imply that the treatment made the tomatoes smaller on average.

   Note to Instructor: The answer/information above appears in Step 4 of the randomization test of the student material.

***Step 2*—Take measurements from each group, and calculate the value of the "Diff" statistic from the experiment.**

For the tomato example, first measure the weights of the 5 tomatoes from the treatment group (Group A); next, measure the weights of the 5 tomatoes from the control group (Group B); finally, compute "Diff" $= \bar{x}_A - \bar{x}_B$, which will serve as the result from your experiment.

2. Assume that the following represents the two groups of tomatoes from the *actual* experiment. Calculate the value of "Diff" $= \bar{x}_A - \bar{x}_B$. This will serve as the result from your experiment.

These are the same 10 tomatoes used in previous lessons; the identification of which tomatoes are "treatment" vs. "control" is now revealed.

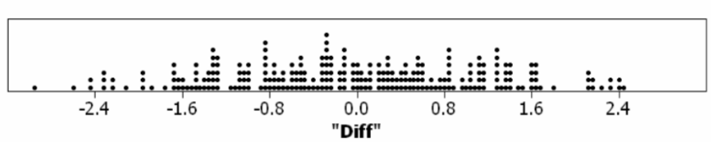
TREATMENT Group A	CONTROL Group B
9.1	7.7
8.4	6.4
8.0	5.2
7.3	4.4
5.9	3.8

"Diff" $= 7.74 - 5.5 = 2.24$

Again, these tomatoes represent the actual result from your experiment. You will now create the randomization distribution by making repeated random assignments of these 10 tomatoes into 2 groups and recording the observed difference in means for each random assignment. This develops a randomization distribution of the many possible difference values that could occur under the assumption that there is no difference between the mean weights of tomatoes that receive the treatment and tomatoes that don't receive the treatment.

***Step 3*–Randomly assign the observations to two groups, and calculate the difference between the group means. Repeat this several times, recording each difference. This will create the randomization distribution for the "Diff" statistic.**

Examples of this technique were presented in a previous lesson. For the tomato example, the randomization distribution has already been presented in a previous lesson and is shown again here. The dots are placed at increments of 0.04 ounces.



***Step 4*—With reference to the randomization distribution (Step 3) and the inequality in your alternative hypothesis (Step 1), compute the probability of getting a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in your experiment (Step 2).**

For the tomato example, since the treatment group is Group A, the "Diff" value of $\bar{x}_A - \bar{x}_B$ is $\bar{x}_{Treatment} - \bar{x}_{Control}$. Since the alternative claim is supported by $\bar{x}_{Treatment} > \bar{x}_{Control}$, you are seeking statistically significant "Diff" values that are positive since if $\bar{x}_{Treatment} > \bar{x}_{Control}$, then $\bar{x}_{Treatment} - \bar{x}_{Control} > 0$.

Statistically significant values of "Diff" that are negative, in this case, would imply that the treatment made the tomatoes smaller on average.

3. Using your calculation from Exercise 2, determine the probability of getting a "Diff" value as extreme as or more extreme than the "Diff" value you obtained for this experiment (in Step 2).

6 out of $250 = 0.024$ *or* 2.4%

***Step 5*—Make a conclusion in context based on the probability calculation (Step 4).**

If there is a small probability of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in your experiment, then the "Diff" value from the experiment is unusual and not typical of chance behavior. Your experiment's results probably did not happen by chance, and the results probably occurred because of a statistically significant difference in the two groups.

- **In the tomato experiment, if you think there is a statistically significant difference in the two groups, you have evidence that the treatment may in fact be yielding heavier tomatoes on average.**

If there is not a small probability of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in your experiment, then your "Diff" value from the experiment is NOT considered unusual and could be typical of chance behavior. The experiment's results may have just happened by chance and not because of a statistically significant difference in the two groups.

- **In the tomato experiment, if you don't think that there is a statistically significant difference in the two groups, then you do not have evidence that the treatment results in larger tomatoes on average.**

In some cases, a specific cut-off value called a *significance level* might be employed to assist in determining how small this probability must be in order to consider results statistically significant.

4. Based on your probability calculation in Exercise 3, do the data from the tomato experiment support the claim that the treatment yields heavier tomatoes on average? Explain.

The data from the tomato experiment support the claim that the treatment yields heavier tomatoes on average. The probability of obtaining a "Diff" value as extreme as or more extreme than 2.24 is very small if the treatment is ineffective. This means that there is a statistically significant difference between the average weights of the treatment and control tomatoes with the treatment tomatoes being heavier on average.

Exercises 5–10 (10 minutes): Developing the Randomization Distribution

MP.5

In this set of exercises, students will begin the important task of developing their own randomization distributions—first manually, then using technology. Students' work will vary, so there will be no specific single answers for certain tasks in this teacher material.

For Exercises 5–7 under ***Manually Generated***, it is strongly recommended that the performing of random assignments (and computing of the "Diff" statistic) be distributed among the class rather than having any one individual perform too many of these randomizations by hand. While the task may be repetitive, it is important for students to grasp what a randomization distribution is, how it is developed, and how technology can greatly assist.

Exercises 5–10: Developing the Randomization Distribution

Although you are familiar with how a randomization distribution is created in the tomato example, the randomization distribution was provided for you. In this exercise, you will develop two randomization distributions based on the same group of 10 tomatoes. One distribution will be developed by hand and will contain the results of at least 250 random assignments. The second distribution will be developed using technology and will contain the results of at least 250 random assignments. Once the two distributions have been developed, you will be asked to compare the distributions.

Manually Generated

Your instructor will provide you with specific guidance regarding how many random assignments you need to carry out. Ultimately, your class should generate at least 250 random assignments, compute the "Diff" value for each, and record these 250 or more "Diff" values on a class or individual dot plot.

5. **To begin, write the 10 tomato weights on 10 equally sized slips of paper – one weight on each slip. Place the slips in a container and shake the container well. Remove 5 slips and assign those 5 tomatoes as Group A. The remaining tomatoes will serve as Group B.**

   For the manually generated randomization distribution, it states that students can use 10 equally sized pieces of paper to be selected from a container. However, manipulatives such as small chips or checkers would be fine as well.

6. **Calculate the mean weight for Group A and the mean weight for Group B. Then, calculate "Diff" = $\bar{x}_A - \bar{x}_B$ for this random assignment.**

   See above.

7. **Record your "Diff" value and add this value to the dot plot. Repeat as needed per your instructor's request until a manually generated randomization distribution of at least 250 differences has been achieved.**

   (Note: This distribution will most likely be slightly different from the "tomato" randomization distribution given earlier in this lesson.)

   Students' work will vary, but, overall, the manually generated distribution for the "Diff" statistic in this tomato example should not differ too much from the distribution presented previously in this and other lessons.

As stated earlier, when it is time to move on to Exercises 8–10, students' computers (or at least one station in the classroom) will require internet access in order to use the "Anova Shuffle" applet. Instructor practice with the applet beforehand is recommended. Specific instructions, comments, and recommendations regarding its use appear in the student lesson material.

This computer generated distribution will most likely be slightly different from both the randomization distribution that appeared earlier in this lesson and the randomization distribution that was manually generated.

Computer Generated

At this stage, you will be encouraged to use a web-based randomization testing applet/calculator to perform the steps above. The applet is located at http://www.rossmanchance.com/applets/AnovaShuffle.htm. To supplement the instructions below, a screenshot of the applet appears as the final page of this lesson.

Upon reaching the applet, do the following:

- **Press the "Clear" button to clear the data under "Sample Data."**
- **Enter the tomato data exactly as shown below. When finished, press the "Use Data" button.**

Group	Ounces
Treatment	9.1
Treatment	8.4
Treatment	8
Control	7.7
Treatment	7.3
Control	6.4
Treatment	5.9
Control	5.2
Control	4.4
Control	3.8

Once the data are entered, notice that dot plots of the two groups appear. Also, the "statistic" window below the data now says "difference in means," and an "Observed Diff" value of 2.24 is computed for the experiment's data (just as you computed in Exercise 2).

By design, the applet will determine the difference of means based on the first group name it encounters in the data set – specifically, it will use the first group name it encounters as the first value in the "difference of means" calculation. In other words, to compute the difference in means as $\overline{x}_{Treatment} - \overline{x}_{Control}$, a "Treatment" observation needs to appear prior to any "Control" observations in the data set as entered.

- **Select the check box next to "Show Shuffle Options" and a dot plot template will appear.**
- **Enter "250" in the box next to "Number of Shuffles," and press the "Shuffle Responses" button. A randomization distribution based on 250 randomizations (in the form of a histogram) is created.**

This distribution will most likely be *slightly* different from both the "tomato" randomization distribution that appeared earlier in this lesson and the randomization distribution that was manually generated in Exercise 7.

8. Write a few comments comparing the manually generated distribution and the computer generated distribution. Specifically, did they appear to have roughly the same shape, center, and spread?

Students should compare and contrast the shape, center, and spread characteristics of the manually generated and computer generated distributions. There should be a great deal of similarity, but outliers and slightly different clustering patterns may be present.

The applet also allows you to compute probabilities. For this case:

- **Under "Count Samples," select "Greater Than." Then, in the box next to "Greater Than," enter "2.2399."**

Since the applet computes the count value as *strictly* "greater than" and not "greater than or equal to," in order to obtain the probability of obtaining a value as extreme as or more extreme than the "Observed Diff" value of 2.24, you will need to enter a value just slightly below 2.24 to ensure that "Diff" observations of 2.24 are included in the count.

- **Select the "Count" button. The probability of obtaining a "Diff" value of 2.24 or more in this distribution will be computed for you.**

  The applet displays the randomization distribution in the form of a histogram, and it shades in red *all* histogram classes that contain *any* difference values that meet your "Count Samples" criteria. Due to the grouping and binning of the classes, some of the red shaded classes (bars) may also contain difference values that do not fit your "Count Samples" criteria. Just keep in mind that the "Count" value stated in red below the histogram will be exact; the red shading in the histogram may be approximate.

9. How did the probability of obtaining a "Diff" value of 2.24 or more using your computer generated distribution compare with the probability of obtaining a "Diff" value of 2.24 or more using your manually generated distribution?

It is expected that the event of finding a "Diff" value of 2.24 or more will have a similarly rare probability of occurrence in both the manually generated and the computer generated distributions.

10. Would you come to the same conclusion regarding the experiment using either the computer generated or manually generated distribution? Explain. Is this the same conclusion you came to using the distribution shown earlier in this lesson back in "Step 3"?

Based on the assumptions regarding the similarity of the distributions as described in Exercise 9 above, it is expected that students will come to the same conclusion (that the treatment is effective) from both the manually generated and the computer generated distributions as was seen in Exercise 4.

Closing (5 minutes)

Realizing that this lesson may require more time than others, if time permits, consider posing the following questions; allow a few student responses for each, or consider investigation/research outside of class.

- Although the idea of randomization testing has been around since the 1930's, it was not used much in practice until the later part of the 20th Century. Why might that be?
  - *Sample response: The computing power necessary to develop a randomization distribution based on numerous random assignments–particularly from data sets with far more observations than those seen in these lessons– was just not as easily or cheaply available in the early 20th Century. With technology today, we can develop a randomization distribution from a large data set and from a very high number of random assignments in seconds.*
- In the previous lesson's Exit Ticket problems, a pain reliever was tested. Subjects who did not receive the pain reliever still received a pill—but the pill had no medicine. Why would this step be taken?
  - *Sample response: Many experiments have shown that human beings tend to exhibit a positive result such as an improvement in pain or a loss of weight even when they are **unknowingly** taking medicine that has no real medicine or benefit, known as a "**placebo**." Since some people in the experiment might exhibit this "placebo effect," rather than compare the treatment group to a group that receives nothing at all, it is important to establish a baseline of comparison that accounts for this power of suggestion in the subjects' minds; thus, the control group is given a placebo without their knowledge.*

Scaffolding:

Some students may not be familiar with the word “placebo.” Have students look up the definition and put it in their own words to ensure they understand the meaning.

Ask students to summarize the main ideas of the lesson in writing or with a neighbor. Use this as an opportunity to informally assess comprehension of the lesson. The Lesson Summary below offers some important ideas that should be included.

Lesson Summary

The following are the general steps for carrying out a randomization test to analyze the results of an experiment.

Step 1*—*Develop competing claims: No Difference vs. Difference

***Develop the null hypothesis*: This claim is that there is no difference between the two groups in the experiment.**

***Develop the alternative hypothesis*: The competing claim is that there is a difference between the two groups. This difference could take the form of a "different from," "greater than," or "less than" statement depending on the purpose of the experiment and the claim being assessed.**

Step 2*—*Take measurements from each group, and calculate the value of the "diff" statistic from the experiment.

This is the observed "Diff" value from the experiment.

Step 3*—*Randomly assign the observations to two groups, and calculate the difference between the group means. Repeat this several times, recording each difference.

This will create the *randomization distribution* for the "Diff" statistic under the assumption that there is no statistically significant difference between the two groups.

Step 4*—*With reference to the randomization distribution (from Step 3) and the inequality in your alternative hypothesis (from Step 1), compute the probability of getting a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in your experiment (from Step 2).

Step 5*—*Make a conclusion in context based on the probability calculation (from Step 4).

Small probability: If the "Diff" value from the experiment is unusual and not typical of chance behavior, your experiment's results probably did not happen by chance. The results probably occurred because of a statistically significant difference in the two groups.

Not a small probability: If the "Diff" value from the experiment is NOT considered unusual and could be typical of chance behavior, your experiment's results may have just happened by chance and NOT because of a statistically significant difference in the two groups.

***Note:* The use of technology is strongly encouraged to assist in Steps 2 – 4.**

Exit Ticket (10 minutes)

Name ______________________________ Date ____________________

Lesson 27: Ruling Out Chance

Exit Ticket

In the Exit Ticket of a previous lesson, an experiment with 20 subjects investigating a new pain reliever was introduced. The subjects were asked to communicate their level of pain on a scale of 0 to 10 where "0" means "no pain" and "10" means "worst pain." Due to the structure of the scale, a patient would desire a lower value on this scale after treatment for pain. The value "ChangeinScore" was recorded as the subject's pain score after treatment minus the subject's pain score before treatment. Since the expectation is that the treatment would lower a patient's pain score, you would desire a ***negative*** value for "ChangeinScore." For example, a "ChangeinScore" value of -2 would mean that the patient was in less pain (for example, now at a "6," formerly at an "8").

In the experiment, the null hypothesis would be that the treatment had no effect. The average change in pain score for the treatment group would be the same as the average change in pain score for the placebo (control) group.

1. The alternative hypothesis would be that the treatment was effective. Using this context, which mathematical relationship below would match this alternative hypothesis? *Choose one*:

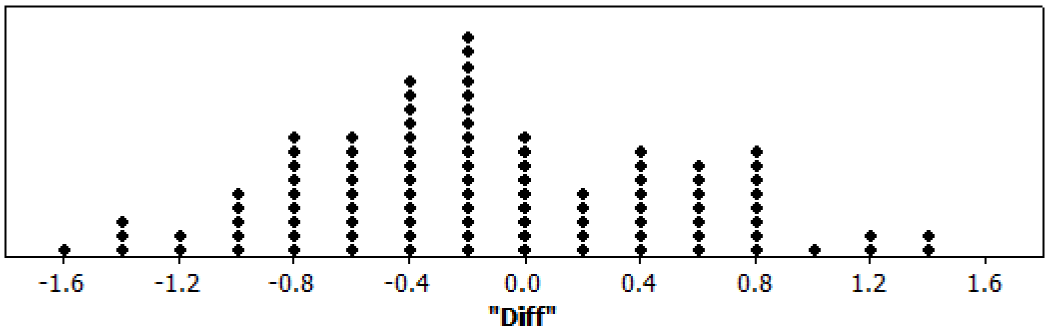
   a. The average change in pain score (the average "ChangeinScore") for the treatment group would be less than the average change in pain score for the placebo group (supported by $\bar{x}_{Treatment} < \bar{x}_{Control}$, or $\bar{x}_{Treatment} - \bar{x}_{Control} < 0$).

   b. The average change in pain score (the average "ChangeinScore") for the treatment group would be greater than the average change in pain score for the placebo group (supported by $\bar{x}_{Treatment} > \bar{x}_{Control}$, or $\bar{x}_{Treatment} - \bar{x}_{Control} > 0$).

2. Imagine that the 20 "ChangeinScore" observations below represent the change in pain levels of the 20 subjects (chronic pain sufferers) who participated in the clinical experiment. The 10 individuals in Group A (the treatment group) received a new medicine for their pain while the 10 individuals in Group B received the pill with no medicine (*placebo*). Assume for now that the 20 individuals have similar initial pain levels and medical conditions. Calculate the value of "Diff" = $\bar{x}_A - \bar{x}_B = \bar{x}_{Treatment} - \bar{x}_{Control}$. This is the result from the experiment.

Group	ChangeinScore
A	0
A	0
A	-1
A	-1
A	-2
A	-2
A	-3
A	-3
A	-3
A	-4
B	0
B	0
B	0
B	0
B	0
B	0
B	-1
B	-1
B	-1
B	-2

3. Below is a randomization distribution of the value "Diff" ($\bar{x}_A - \bar{x}_B$) based on 100 random assignments of these 20 observations into two groups of 10 (shown in a previous lesson).



With reference to the randomization distribution above and the inequality in your alternative hypothesis, compute the probability of getting a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in the experiment.

4. Based on your probability value from Problem 3 and the randomization distribution above, *choose one* of the following conclusions:

   a. Due to the small chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment, we believe that the observed difference did not happen by chance alone, and we support the claim that the treatment is effective.

   b. Because the chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment is not small, it is possible that the observed difference may have happened by chance alone, and we cannot support the claim that the treatment is effective.

Exit Ticket Sample Solutions

In the Exit Ticket of a previous lesson, an experiment with 20 subjects investigating a new pain reliever was introduced. The subjects were asked to communicate their level of pain on a scale of 0 to 10 where "0" means "no pain" and "10" means "worst pain." Due to the structure of the scale, a patient would desire a lower value on this scale after treatment for pain. The value "ChangeinScore" was recorded as the subject's pain score after treatment minus the subject's pain score before treatment. Since the expectation is that the treatment would lower a patient's pain score, you would desire a *negative* value for "ChangeinScore." For example, a "ChangeinScore" value of -2 would mean that the patient was in less pain (for example, now at a "6," formerly at an "8").

In the experiment, the null hypothesis would be that the treatment had no effect. The average change in pain score for the treatment group would be the same as the average change in pain score for the placebo (control) group.

1. **The alternative hypothesis would be that the treatment was effective. Using this context, which mathematical relationship below would match this alternative hypothesis? *Choose one*:**

   a. **The average change in pain score (the average "ChangeinScore") for the treatment group would be less than the average change in pain score for the placebo group (supported by $\bar{x}_{Treatment} < \bar{x}_{Control}$, or $\bar{x}_{Treatment} - \bar{x}_{Control} < 0$).**

   b. **The average change in pain score (the average "ChangeinScore") for the treatment group would be greater than the average change in pain score for the placebo group (supported by $\bar{x}_{Treatment} > \bar{x}_{Control}$, or $\bar{x}_{Treatment} - \bar{x}_{Control} > 0$).**

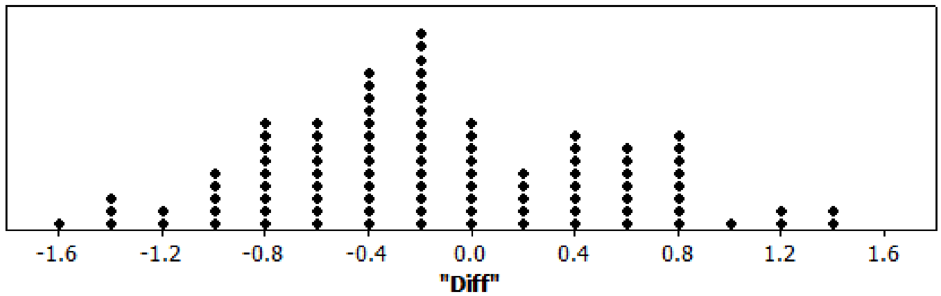
   Choice (a): The average change in pain score (the average "ChangeinScore") for the treatment group would be less than the average change in pain score for the placebo group.

2. **Imagine that the 20 "ChangeinScore" observations below represent the change in pain levels of the 20 subjects (chronic pain sufferers) who participated in the clinical experiment. The 10 individuals in Group A (the treatment group) received a new medicine for their pain while the 10 individuals in Group B received the pill with no medicine (*placebo*). Assume for now that the 20 individuals have similar initial pain levels and medical conditions. Calculate the value of "Diff" = $\bar{x}_A - \bar{x}_B = \bar{x}_{Treatment} - \bar{x}_{Control}$. This is the result from the experiment.**

   "Diff" = -1.4 (Group A mean = -1.9; Group B mean = -0.5)

Group	ChangeinScore
A	0
A	0
A	-1
A	-1
A	-2
A	-2
A	-3
A	-3
A	-3
A	-4
B	0
B	0
B	0
B	0
B	0
B	0
B	-1
B	-1
B	-1
B	-2

3. **Below is a randomization distribution of the value "Diff" ($\bar{x}_A - \bar{x}_B$) based on 100 random assignments of these 20 observations into two groups of 10 (shown in a previous lesson).**



With reference to the randomization distribution above and the inequality in your alternative hypothesis, compute the probability of getting a "Diff" value as extreme as or more extreme than the "Diff" value you obtained in the experiment.

4% (4 out of 100)

4. **Based on your probability value from Problem 3 and the randomization distribution above, *choose one* of the following conclusions:**

   a. **Due to the small chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment, we believe that the observed difference did not happen by chance alone, and we support the claim that the treatment is effective.**

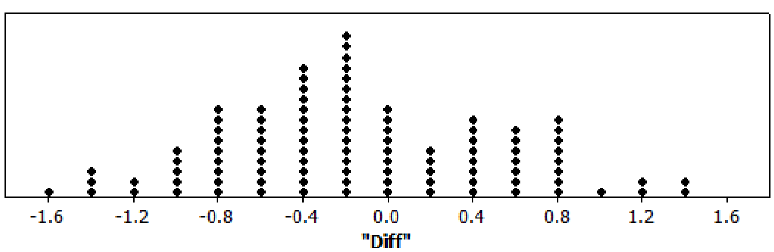
   b. **Because the chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment is not small, it is possible that the observed difference may have happened by chance alone, and we cannot support the claim that the treatment is effective.**

   Choice (a): Due to the small chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment, we believe that the observed difference did not happen by chance alone, and we support the claim that the treatment is effective.

Problem Set Sample Solutions

1. **Using the 20 observations that appear in Exit Ticket Question 2 for the "changes in pain" scores of 20 individuals, use the "Anova Shuffle" applet to develop a randomization distribution of the value "Diff" ($\bar{x}_A - \bar{x}_B$) based on 100 random assignments of these 20 observations into two groups of 10. Enter the data exactly as shown in Question 2. Describe similarities and differences between this new randomization distribution and the distribution shown in Exit Ticket Question 3.**

Group	ChangeinScore
A	0
A	0
A	-1
A	-1
A	-2
A	-2
A	-3
A	-3
A	-3
A	-4
B	0
B	0
B	0
B	0
B	0
B	0
B	-1
B	-1
B	-1
B	-2



Answers will vary, but the distribution should not differ too much from the distribution presented previously in this lesson (Exit Ticket Question 3) and other lessons.

2. **In a previous lesson, the burn times of 6 candles were presented. It is believed that candles from Group A will burn longer on average than candles from Group B. The data from the experiment (now shown with group identifiers) are provided below.**

Group	Burntime
A	18
A	12
B	9
A	6
B	3
B	0

Perform a randomization test of this claim. Carry out all 5 steps, and use the "Anova Shuffle" applet to perform Steps 2 to 4. Enter the data exactly as presented above, and in Step 3, develop the randomization distribution based on 200 random assignments.

Step 1—Null Hypothesis: Candles from Group A will burn for the same amount of time on average as candles from Group B. (No difference in average burn time)

Alternative Hypothesis: Candles from Group A will burn longer on average than candles from Group B.

Step 2—"Diff" = $12 - 4 = 8$ minutes.

Step 3—Randomization Distribution of "Diff" developed by student.

Step 4—Compute probability of "Diff" greater than or equal to 8 (value should be close to 10%).

Step 5—While there is no specific criteria stated in the question for what is a "small probability," students should consider probability values from previous work in determining "small" vs. "not small." Again, student values will vary. The important point is that the students' conclusion should be consistent with the probability value and the students' assessment of that value as follows:

If students deem the probability to be "small," then they should state a conclusion based on a statistically significant result. More specifically, due to the small chance of obtaining a "Diff" value as extreme as or more extreme than the "Diff" value obtained in the experiment, it is believed that the observed difference did not happen by chance alone, and we support the claim that the Group A candles burn longer on average.

If students deem the probability to be NOT "small," then they should state a conclusion based on a result that is NOT statistically significant. More specifically, it is believed that the observed difference may have occurred by chance, and we do NOT have evidence to support the claim that the Group A candles burn longer on average.

Appendix: Screenshot of Applet