



Lesson 26: 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

In the previous lesson, students investigated examples of the random assignment of 10 tomatoes to two groups of 5 each. In each case, they calculated "Diff" $= \bar{x}_A - \bar{x}_B$, the difference between the mean weight of the 5 tomatoes in Group A and the mean weight of the 5 tomatoes in Group B. The "Diff" values varied.

This lesson asks students to consider if certain "Diff" values may be unusual or extreme. To do this, students need to develop a sense of the center, spread, and shape of the distribution of possible "Diff" values. Developing a distribution of **ALL** possible values based on this random assignment approach would be very difficult—particularly if there were a greater number of tomatoes involved. Thus, repeated simulation is employed to develop something called a *randomization distribution* to adequately approximate the true probability distribution of "Diff." The randomization distributions are pre-developed when presented in this lesson; in the next lesson, students will create their own randomization distributions.

Classwork

Opening Exercise (3 minutes)

Recall the scenario that students worked through in the previous lesson. Let students answer the questions independently and share their responses with a neighbor. Then, discuss as a class the following:

- Even a distribution that is centered at 0 could have some variability. This means it is important to examine the entire distribution of the "Diff" statistic to establish just how unusual or extreme a specific "Diff" value is. To do this, you need a sense of the center, spread, and shape of the distribution in order to determine if a specific value is unusual or extreme.
- Developing a distribution of ALL possible differences based on this random assignment approach could be very difficult—particularly if there were a greater number of tomatoes involved. Fortunately, you can use simulation to develop something called a *randomization distribution* to adequately approximate the true probability distribution of "Diff."

Scaffolding:

Use the following as a concrete example that illustrates the point made in this opening:

If the treatment was not effective, which of the following mean weights would you expect?

A	B
4.2	4.0
2.9	2.3
3.0	3.4
2.0	1.2
5.0	4.0

This highlights the difficulty in quantifying what is a "meaningful" difference in means and illustrates the point of the lesson.

Opening Exercise

Previously, you considered the random assignment of 10 tomatoes into two distinct groups of 5 tomatoes each called Group A and Group B. With each random assignment, you calculated “Diff” = $\bar{x}_A - \bar{x}_B$, the difference between the mean weight of the 5 tomatoes in Group A and the mean weight of the 5 tomatoes in Group B.

- a. Summarize in writing what you learned in the last lesson. Share your thoughts with a neighbor.

In the last lesson, when the single group of observations was randomly divided into two groups, the means of these two groups differed by chance. In some cases, the difference in the means of these two groups was very small (or “0”), but in other cases, this difference was larger. However, in order to determine which differences were typical and ordinary vs. unusual and rare, a sense of the center, spread, and shape of the distribution of possible differences is needed.

- b. Recall that 5 of these 10 tomatoes are from plants that received a nutrient treatment in the hope of growing bigger tomatoes. But what if the treatment was *not* effective? What difference would you expect to find between the group means?

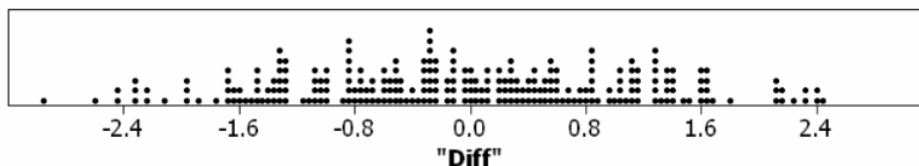
I would expect there to be no difference between the mean of Group A and the mean of Group B when performing these randomization assignments; in other words, I would expect a value of “Diff” equal to 0.

Exercises 1–2 (7 minutes): The Distribution of “Diff” and Why “0” is Important

Read through the beginning of the exercise and verify that students understand the information displayed in the dot plot. Let students work with a partner on Exercises 1 and 2. Then, confirm answers as a class.

Exercises 1–2: The Distribution of “Diff” and Why “0” is Important

In the previous lesson, 3 instances of the tomato randomization were considered. Imagine that the random assignment was conducted an additional 247 times, and 250 “Diff” values were computed from these 250 random assignments. The results are shown graphically below in a dot plot where each dot represents the “Diff” value that results from a random assignment:



This dot plot will serve as your *randomization distribution* for the “Diff” statistic in this tomato randomization example. The dots are placed at increments of 0.04 ounces.

1. Given the distribution picture above, what is the *approximate* value of the median and mean of the distribution? Specifically, do you think this distribution is centered near a value that implies “No Difference” between Group A and Group B?

The “Diff” value that implies “No Difference” between Group A and Group B would be “0.” From visual inspection, the approximate value of the median and mean of the distribution appears to be near “0,” based on the near symmetry and the center of the distribution. (Note: The actual mean in this case is -0.053 ounces, and the median is -0.08 ounces—both just slightly below “0.”)

2. Given the distribution pictured above and based on the simulation results, determine the approximate probability of obtaining a "Diff" value in the cases described in (a), (b), and (c).

a. of 1.64 ounces or more

17 out of 250 are 1.64 or more.

$$\frac{17}{250} = 0.068 \text{ or } 6.8\%$$

b. of -0.80 ounces or less

69 out of 250 are -0.80 or less.

$$\frac{69}{250} = 0.276 \text{ or } 27.6\%$$

c. within 0.80 ounces of 0 ounces

121 out of 250 are between -0.80 and 0.80.

$$\frac{121}{250} = 0.484 \text{ or } 48.4\%$$

d. How do you think these probabilities could be useful to people that are designing experiments?

The probabilities could be used to help determine if the differences occurred by chance or not.

Exercises 3–5 (15 minutes): Statistically Significant "Diff" Values

Determining how unusual or extreme a "Diff" value is will allow students to then consider if their experiment's results are statistically significant. The reasoning is as follows.

- 5 of these 10 tomatoes are from plants that received a nutrient treatment in the hope of growing bigger tomatoes, and the other 5 received no such treatment. If the treatment was *not* effective, then one would generally expect there to be no difference between the mean of Group A and the mean of Group B when performing these randomization assignments; in other words, we would expect a value of "Diff" equal to 0.
- However, as seen in the previous lesson, the value of "Diff" varies due to chance behavior.
- After establishing a sense of the full distribution of "Diff," if the observed difference from an experiment is "extreme" (far from "0") and not typical of chance behavior, it may be considered "statistically significant" and possibly not the result of chance behavior.
- If the difference is not the result of chance behavior, then maybe the difference didn't just happen by chance alone.
- If the difference didn't just happen by chance alone, maybe the difference observed in the experiment is caused by the treatment in question, which, in this case, is the nutrient.

Keep in mind that by saying "statistically significant" in this case, we are saying that the observed difference between two groups is not likely due to chance.

In Exercises 3–5 and 6–8, students are asked to speculate if certain values of the "Diff" statistic are statistically significant. Specifically, students are told that the "Diff" statistic value should be considered statistically significant if there is a **low** probability of obtaining a result as extreme as or more extreme than the value in question. However, a "cutoff value" as to what is considered a "low" probability has been deliberately omitted. Ideally, students should give some thought as to what probability values might be associated with "unusual" events. In real-world situations, the "cutoff" probability value used for determining statistical significance (called a *significance level*) varies from situation to situation based on context; however, many introductory statistics references use a value of 0.05, or 5%, as a benchmark.

Exercises 3–5: Statistically Significant 'Diff' Values

In the context of a randomization distribution that is based upon the assumption that there is no real difference between the groups, consider a "Diff" value of X to be "statistically significant" if there is a low probability of obtaining a result that is as extreme as or more extreme than X .

3. Using that definition and your work above, would you consider any of the "Diff" values below to be statistically significant? Explain.

a. 1.64 ounces

possibly statistically significant; an event with a 6.8% probability of occurring is not a very frequent occurrence.

b. -0.80 ounces

not statistically significant; an event with a 27.6% probability of occurring is a fairly common occurrence.

c. Values within 0.80 ounces of 0 ounces

Values within 0.80 ounces of "0 ounces"—not statistically significant; these values are not very far from "0," and they are fairly common. Also, given the symmetry, if -0.80 is not considered statistically significant (in part (b), above), then values that are closer to "0" would also not be considered statistically significant.

4. In the previous lessons, you obtained "Diff" values of 0.28 ounces, 2.44 ounces, and 0 ounces for 3 different tomato randomizations. Would you consider any of those values to be "statistically significant" for this distribution? Explain.

The values of 0 and 0.28 ounces would not be statistically significant based on the work above and the fact that neither value is very far from "0" in the distribution. However, the value of "2.44" would be statistically significant because it is very far from "0" (maximum observation), and there is only a 1 in 250 chance (0.004 or 0.4% chance) of obtaining a value that extreme in this distribution.

5. Recalling that "Diff" is the mean weight of the 5 Group A tomatoes minus the mean weight of the 5 Group B tomatoes, how would you explain the meaning of a "Diff" value of 1.64 ounces in this case?

The 5 tomatoes of Group A have a mean weight that is 1.64 ounces higher than the mean weight of Group B's 5 tomatoes.

Exercises 6–8 (10 minutes): The Implication of Statistically Significant "Diff" Values

Read through the exercise as a class. You may want to work through one or two of the “Diff” values in Exercise 6 as a class. Let students continue to work with their partner on the exercises. Then, confirm answers as a class.

Exercises 6–8: The Implication of Statistically Significant “Diff” Values

Keep in mind that for reasons mentioned earlier, the randomization distribution above is demonstrating what is likely to happen *by chance alone* if the treatment was not effective. As stated in the previous lesson, you can use this randomization distribution to assess whether or not the *actual* difference in means *obtained from your experiment* (the difference between the mean weight of the 5 actual control group tomatoes and the mean weight of the 5 actual treatment group tomatoes) is consistent with usual chance behavior. The logic is as follows:

- If the observed difference is “extreme” and not typical of chance behavior, it may be considered “statistically significant” and possibly not the result of chance behavior.
- If the difference is not the result of chance behavior, then maybe the difference did not just happen by chance alone.
- If the difference did not just happen by chance alone, maybe the difference you observed is caused by the treatment in question, which, in this case, is the nutrient. In the context of our example, a statistically significant “Diff” value provides evidence that the nutrient treatment did in fact yield heavier tomatoes on average.

6. For reasons that will be explained in the next lesson, for your tomato example, “Diff” values that are *positive* and statistically significant will be considered as good evidence that your nutrient treatment did in fact yield heavier tomatoes on average. Again, using the randomization distribution shown earlier in the lesson, which (if any) of the following “Diff” values would you consider to be statistically significant and lead you to think that the nutrient treatment did, in fact, yield heavier tomatoes on average? Explain for each case.

Diff = 0.4, Diff = 0.8, Diff = 1.2, Diff = 1.6, Diff = 2.0, Diff = 2.4

Diff = 0.4: not statistically significant; not very far from “0,” 91 of 250 values (36.4%) are greater than or equal to 0.4.

Diff = 0.8: not statistically significant; not very far from “0,” 64 of 250 values (25.6%) are greater than or equal to 0.8.

Diff = 1.2: not statistically significant; not too far from “0,” 38 of 250 values (15.2%) are greater than or equal to 1.2.

Diff = 1.6: possibly statistically significant; somewhat far from “0,” 21 of 250 values (8.4%) are greater than or equal to 1.6.

Diff = 2.0: statistically significant; very far from “0,” only 11 of 250 values (4.4%) are greater than or equal to 2.0.

Diff = 2.4: statistically significant; very far from “0,” only 3 of 250 values (1.2%) are greater than or equal to 2.4.

7. In the first random assignment in the previous lesson, you obtained a "Diff" value of 0.28 ounces. Earlier in this lesson, you were asked to consider if this might be a "statistically significant" value. Given the distribution shown in this lesson, if you had obtained a "Diff" value of 0.28 ounces *in your experiment* and the 5 Group A tomatoes had been the "treatment" tomatoes that received the nutrient, would you say that the "Diff" value was extreme enough to support a conclusion that the nutrient treatment yielded heavier tomatoes on average? Or, do you think such a "Diff" value may just occur by chance when the treatment is ineffective? Explain.

I would say that the "Diff" value of 0.28 was NOT extreme enough to support a conclusion that the nutrient treatment yielded heavier tomatoes on average. Such a "Diff" value may just occur by chance in this case. See earlier work in Exercise 4. Also, referencing the question above, 0.28 is even closer to "0" than other "not statistically significant" values.

8. In the second random assignment in the previous lesson, you obtained a "Diff" value of 2.44 ounces. Earlier in this lesson, you were asked to consider if this might be a "statistically significant" value. Given the distribution shown in this lesson, if you had obtained a "Diff" value of 2.44 ounces *in your experiment* and the 5 Group A tomatoes had been the "treatment" tomatoes that received the nutrient, would you say that the "Diff" value was extreme enough to support a conclusion that the nutrient treatment yielded heavier tomatoes on average? Or do you think such a "Diff" value may just occur by chance when the treatment is ineffective? Explain.

I would say that the "Diff" value of 2.44 was extreme enough to support a conclusion that the nutrient treatment yielded heavier tomatoes on average. This most likely did NOT just occur by chance. See earlier work in Exercise 4. Also, referencing the question above, 2.44 is even farther away from "0" than other statistically significant values.

MP.2

Closing (5 minutes)

- If you were about to prepare for a debate, a court case, or some other situation where you were challenging an existing claim such as "the treatment is not effective" or "the process is not harming anyone," what kind of probability value would you be looking for in your results before you would claim "statistical significance" and feel comfortable challenging the existing claim? What characteristics of the situation might influence your decision?
 - *Sample response: In order to be confident in your challenge, your probability value needs to be fairly small. In the tomato example, consider that you are saying "If the nutrient was not effective, the chances of our obtaining a value this extreme in our experiment is ____; therefore, we think the nutrient must be working." You are acknowledging that your experiment's results could still occur a certain percent of the time even if the nutrient really didn't work! If you were trying to convince an audience, constituency, or jury, you probably wouldn't want to say, "There's a 1 in 3 (33%) chance of obtaining a value this extreme." Rather, you'd want to say that the chances are "1 in 20 (5%)," or "1 in 100 (1%)," or "1 in 1,000 (0.1%)."* Considerations include the amount of money and time required for sampling, the severity of the claim that is being examined, and the risks of falsely rejecting the claim of "no difference" when it was true all along, and so on.
- Calculate and interpret "Diff" = the mean of Group A minus the mean of Group B, in the following instance. Group A: 10 similar homes with insulated windows have an average monthly electric bill of \$123. Group B: 10 similar homes with non-insulated windows have an average monthly electric bill of \$157.
 - *Sample response: The homes with insulated windows have an average monthly electric bill that is \$34 less than the homes with non-insulated windows.*

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

In the previous lesson, the concept of randomly separating 10 tomatoes into 2 groups and comparing the means of each group was introduced. The randomization distribution of the difference in means that is created from multiple occurrences of these random assignments demonstrates what is likely to happen *by chance alone* if the nutrient treatment is not effective. When the results of your tomato growth experiment are compared to that distribution, you can then determine if the tomato growth experiment's results were typical of chance behavior.

If the results appear typical of chance behavior and near the center of the distribution (that is, not relatively very far from a "Diff" of 0), then there is little evidence that the treatment was effective. However, if it appears that the experiment's results are not typical of chance behavior, then, maybe, the difference you are observing didn't just happen by chance alone. It may indicate a statistically significant difference between the treatment group and the control group, and the source of that difference might be (in this case) the nutrient treatment.

Exit Ticket (5 minutes)

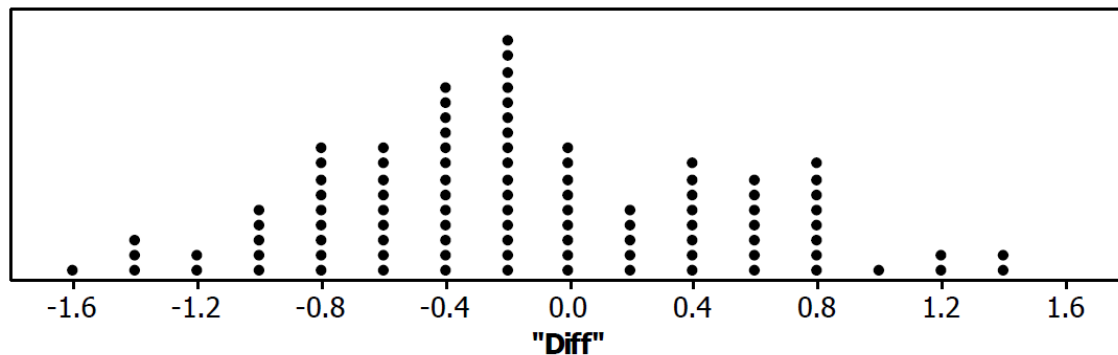
Name _____

Date _____

Lesson 26: Ruling Out Chance

Exit Ticket

Medical patients who are in physical pain are often asked to communicate their level of pain on a scale of 0 to 10 where "0" means "no pain" and "10" means "worst pain." (Note: Sometimes a visual device with "pain faces" is used to accommodate the reporting of the pain score.) Due to the structure of the scale, a patient would desire a lower value on this scale after treatment for pain.



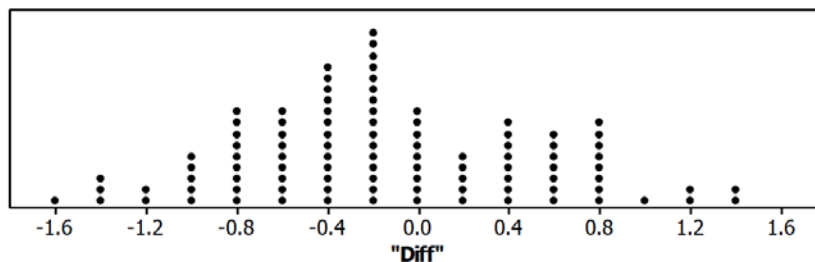
Imagine that 20 subjects participate in a clinical experiment and that a variable of "ChangeinScore" is recorded for each subject as the subject's pain score after treatment minus the subject's pain score before treatment. Since the expectation is that the treatment would lower a patient's pain score, 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").

Although the 20 "ChangeinScore" values for the 20 patients are not shown here, 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 (Group A and Group B).

Exit Ticket Sample Solutions

Note: A graphic of the Wong-Baker FACES Pain Rating Scale (as seen in many physicians' offices) or a similar visual reference may assist students with the Exit Ticket questions. One can be found at <http://pain.about.com/od/testingdiagnosis/ig/pain-scales/Wong-Baker.htm>

Medical patients who are in physical pain are often asked to communicate their level of pain on a scale of 0 to 10 where "0" means "no pain" and "10" means "worst pain." (Note: Sometimes a visual device with "pain faces" is used to accommodate the reporting of the pain score.) Due to the structure of the scale, a patient would desire a lower value on this scale after treatment for pain.



Imagine that 20 subjects participate in a clinical experiment and that a variable of "ChangeinScore" is recorded for each subject 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").

Although the 20 "ChangeinScore" values for the 20 patients are not shown here, 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 (Group A and Group B).

- From the distribution above, what is the probability of obtaining a "Diff" score of -1 or less?

$$\frac{11}{100} = 0.11 \text{ or } 11\%$$

- With regard to this distribution, would you consider a "Diff" value of -0.4 to be statistically significant? Explain.

No. The value is not far from "0," and 42 of the 100 values in the distribution are at or below -0.4 .

- With regard to how "Diff" is calculated, if Group A represented a group of patients in your experiment who received a new pain relief treatment, and Group B received a pill with no medicine (called a *placebo*), how would you interpret a "Diff" value of -1.4 pain scale units in context?

The group that received the pain relief treatment had an average reduction in pain that was 1.4 units better (lower) than the group that received the placebo.

- Given the distribution above, if you obtained such a value of "Diff" (-1.4) from your experiment, would you consider that to be significant evidence of the new treatment being effective on average in relieving pain? Explain.

Yes. -1.4 is far from "0," and the probability of obtaining a "Diff" value of -1.4 or less is only 4%. The value provides evidence that the new treatment may be effective in relieving pain.

Problem Set Sample Solutions

In each of the 3 cases below, calculate the "Diff" value as directed, and write a sentence explaining what the "Diff" value means in context. Write the sentence for a general audience.

1. Group A: 8 dieters lost an average of 8 pounds.

Group B: 8 non-dieters lost an average of 2 pounds over the same time period.

Calculate and interpret "Diff" = the mean of Group A minus the mean of Group B.

"Diff" = -6. The 8 dieters lost an average of 6 pounds more than the 8 non-dieters.

2. Group A: 11 students were on average 0.4 seconds faster in their 100 meter run times after following a new training regimen.

Group B: 11 students were on average 0.2 seconds slower in their 100 meter run times after not following any new training regimens.

Calculate and interpret "Diff" = the mean of Group A minus the mean of Group B.

"Diff" = -0.6 (from $-0.4 - 0.2$ since Group A is 0.4 seconds faster). The 11 students following the new training regimen were on average 0.6 seconds faster in their 100 meter run times than the 11 students not following any new training regimens.

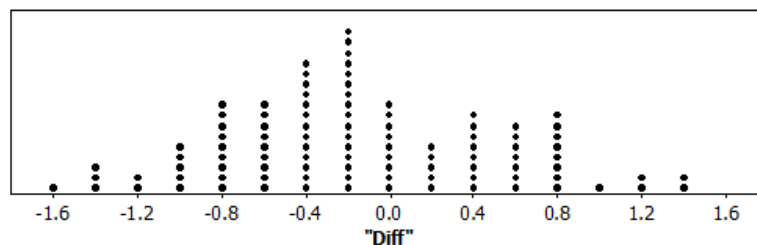
3. Group A: 20 squash that have been grown in an irrigated field have an average weight of 1.3 pounds.

Group B: 20 squash that have been grown in a non-irrigated field have an average weight of 1.2 pounds.

Calculate and interpret "Diff" = the mean of Group A minus the mean of Group B.

"Diff" = 0.1. The 20 squash grown in an irrigated field have an average weight that is 0.1 pounds higher than the 20 squash grown in a non-irrigated field.

4. Using the randomization distribution shown in the Exit Ticket, what is the probability of obtaining a "Diff" value of -0.6 or less?



$$\frac{29}{100} = 0.29 \text{ or } 29\%$$

5. Would a "Diff" value of -0.6 or less be considered a "statistically significant difference"? Why or why not?

No. -0.6 is not far from "0," and the probability of obtaining a "Diff" value of -0.6 or less is 29%.

6. Using the randomization distribution shown in the Exit Ticket, what is the probability of obtaining a "Diff" value of -1.2 or less?

The probability of obtaining a "Diff" value of -1.2 or less is 6 out of 100, or 6%.

7. Would a "Diff" value of -1.2 or less be considered a "statistically significant difference"? Why or why not?

Possibly statistically significant. -1.2 is far from "0," and the probability of obtaining a "Diff" value of -1.2 or less is only 6%. The value provides evidence that the new treatment may be effective in relieving pain.