Lesson 21: Why Worry about Sampling Variability?

Student Outcomes

- Students understand that a *meaningful* difference between two sample means is one that is greater than would have been expected due to just sampling variability.

Lesson Notes

This lesson is the first of three lessons in which students are asked to compare the means of two populations. In this lesson, students will first consider random samples from two populations that have the same mean. Sampling variability must be considered when deciding if there is *evidence* that the two population means are actually the same or different, and this idea is developed throughout the exercises. Students then repeat the process by examining two populations that have different means. The investigation is presented in stages. Questions are posed in the stages that help students think about the sampling variability in different sample means. This lesson may take more than one class period.

Classwork

Preparation: Organize the class in small groups. Before class, prepare three bags for each group, labeled , , and . Each bag contains numbers (written on -inch square pieces of cardstock, round counters, or precut foam squares, etc.) Bag and Bag both consist of the population of numbers as follows: five 's, five 's, five 's, five 's … five 's. Bag consists of the population of numbers: five 's, five 's, five 's, . . . , five 's. Templates that can be used to produce the numbers for the bags are included at the end of this lesson.

There are three bags, Bag , Bag , and Bag , with numbers in each bag. You and your classmates will investigate the population mean (the mean of all numbers) in each bag. Each set of numbers has the same range. However, the population means of each set may or may not be the same. We will see who can uncover the mystery of the bags!

Exercises 1–5 (8–10 minutes)

Before students begin this activity, ask: "How would you go about determining the mean of the numbers in each bag?" Anticipate students would indicate they would need to know all numbers from each bag, and then use the numbers to determine the means. Explain that although knowing all the numbers would allow them to calculate the means, this activity introduces a way to estimate the means using random samples. This activity suggests how to estimate a mean when it is not possible or practical to know the values of the entire population.

The instructions for the investigation should be read aloud to the class. While each group needs only one bag of each population, each student is expected to draw a sample from each bag. Warn students that *before* they draw a new sample from a bag, *all* numbers must be returned to the bag. Students should be allowed minutes to take samples of ten numbers from Bag , Bag , and Bag . Charts are provided to aid students in this process.

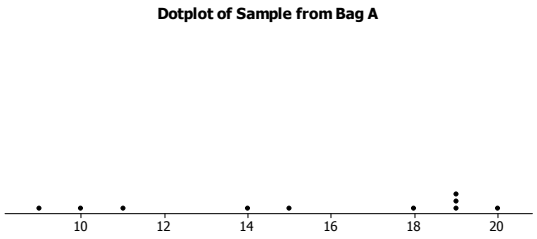
Exercises 1–5

1. To begin your investigation, start by selecting a random sample of ten numbers from Bag . Remember to mix the numbers in the bag first. Then, select one number from the bag. Do not put it back into the bag. Write the number in the chart below. Continue selecting one number at a time until you have selected ten numbers. Mix up the numbers in the bag between each selection.

Selection										
Bag										

   a. Create a dot plot of your sample of ten numbers. Use a dot to represent each number in the sample.

   Dot plot will vary based on the sample selected. One possible answer is shown here.



   b. Do you think the mean of all the numbers in Bag might be ? Why or why not?

   Anticipate that students will indicate a mean of a sample from Bag is greater than . Responses depend on the students' samples and the resulting dot plots. In most cases, the dots will center around a value that is greater than because the mean of the population is greater than .

   c. Based on the dot plot, what would you estimate the mean of the numbers in Bag to be? How did you make your estimate?

   Answers will vary depending on students' samples. Anticipate that most students' estimates will correspond to roughly where the dots in the dot plot center. The population mean here is , so answers around or would be expected.

   d. Do you think your sample mean will be close to the population mean? Why or why not?

   Students could answer "Yes," "No," or "I don't know." The goal of this question is to get students to think about the difference between a sample mean and the population mean.

   e. Is your sample mean the same as your neighbors' sample means? Why or why not?

   No, when selecting a sample at random, different students will get different sets of numbers. This is sampling variability.

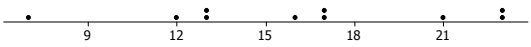
2. Repeat the process by selecting a random sample of ten numbers from Bag .

Selection										
Bag										

a. Create a dot plot of your sample of ten numbers. Use a dot to represent each of the numbers in the sample.

Dot plot will vary based on the sample selected. One possible answer is shown here.

Dotplot of Sample from Bag B



b. Based on your dot plot, do you think the mean of the numbers in Bag is the same or different than the mean of the numbers in Bag ? Explain your thinking.

Answer will vary as students will compare their center of the dot plot of the sample from Bag to the center of the dot plot of the sample from Bag . The centers will probably not be exactly the same; however, anticipate centers that are close to each other.

3. Repeat the process once more by selecting a random sample of ten numbers from Bag .

Selection										
Bag										

a. Create a dot plot of your sample of ten numbers. Use a dot to represent each of the numbers in the sample.

Dot plot will vary based on the sample selected. One possible answer is shown here.

Dotplot of Sample from Bag C

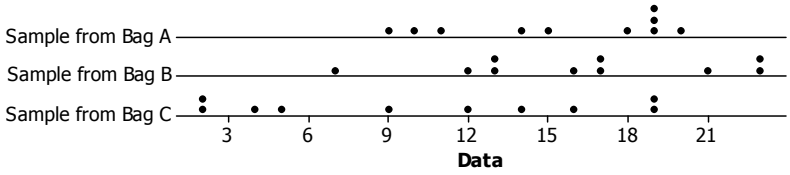
b. Based on your dot plot, do you think the mean of the numbers in Bag is the same or different than the mean of the numbers in Bag ? Explain your thinking.

Anticipate that students will indicate that the center of the dot plot of the sample from Bag is less than the center of the dot plot of the sample from Bag . Because the population mean for Bag is less than the population mean for Bag , the center of the dot plot will usually be less for the sample from Bag .

4. Are your dot plots of the three bags the same as the dot plots of other students in your class? Why or why not?

Dot plots will vary. Because different students generally get different samples when they select a sample from the bags, the dot plots will vary from student to student.

Below are dot plots of all three samples drawn on the same scale. Notice how the samples from Bags and tend to center at approximately , while the sample from Bag centers around .



Next, students will calculate the means for each of the three samples. As students finish their calculation of the means, lead a discussion about why a student's sample means might differ from his or her neighbors' sample means. Point out that what they are observing is *sampling variability* – the chance variability that happens from one sample to another when repeated samples are taken from the same population.

5. Calculate the mean of the numbers for each of the samples from Bag , Bag, , and Bag .

	Mean of the sample of numbers
Bag	
Bag	
Bag	

Answers will vary. For the samples shown in the dot plots above, the sample means are: Bag A , Bag , Bag .

a. Are the sample means you calculated the same as the sample means of other members of your class? Why or why not?

No, when selecting a sample at random, you get different sets of numbers (again, sampling variability).

b. How do your sample means for Bag and for Bag compare?

Students might answer that the mean for the sample from Bag is larger, smaller, or equal to the mean for the sample from Bag , depending on their samples. For the example, given above, the sample mean for Bag is smaller than the sample mean for Bag .

c. Calculate the difference of sample mean for Bag minus sample mean for Bag *B* . Based on this difference, can you be sure which bag has the larger population mean? Why or why not?

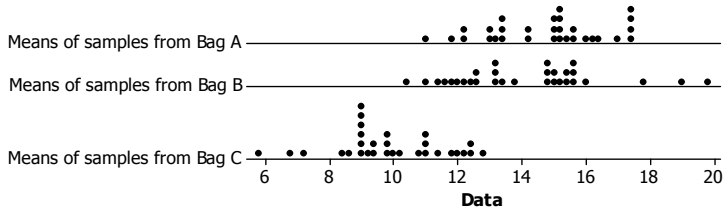
No, it is possible that you could get a sample mean that is larger than the population mean of Bag , and then get a sample mean that is smaller than the population mean of Bag , or vice versa.

MP.

Students begin to examine their samples and the samples of other students using the means and the dot plots. They begin to make conjectures about the populations based on their own samples and the samples of other students.

Creating dot plots of the means that each student calculated for each bag provides another way for students to formulate conjectures about the means of each bag. Have each student plot his or her sample mean for the sample from Bag on a class dot plot. Label this dot plot "Means of samples from Bag ." Then, construct class dot plots for the means from Bags and . Anticipate that the dot plots for the means for Bags and will have similar distributions regarding center and spread. Anticipate that the dot plot for Bag will have similar spread to the dot plots of Bags or ; however, the center will be less than the center for Bags and .

For a class of students, the following dot plots were obtained. The dot plots for your class won't be exactly like these, but they will probably be similar.



Exercises 6–10 (8–10 minutes)

Continue the discussion with students as they work through the next set of exercises. Allow students to individually provide an answer to these questions. After a few minutes, organize a class discussion of each question.

Exercises 6–10

6. **Based on the class dot plots of the sample means, do you think the mean of the numbers in Bag and the mean of the numbers in Bag are different? Do you think the mean of the numbers in Bag and the mean of the numbers in Bag are different? Explain your answers.**

   At this stage of the lesson, students may suspect that Bags and are similar, and also that Bag is different than the other two. This question sets the stage for the rest of this lesson.

7. **Based on the difference between sample mean of Bag and the sample mean of Bag , , that you calculated in Exercise 5, do you think that the two populations (Bags and) have different means or do you think that the two population means might be the same?**

   Answers will vary as the difference of the means will be based on each student's samples. Anticipate answers that indicate the difference in the sample means that the population means might be the same for differences that are close to . (Students learn later in this lesson that the populations of Bags and are the same, so most students will see differences that are not too far from .)

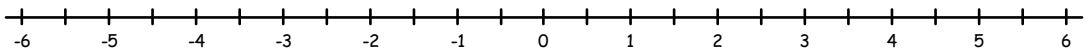
8. **Based on this difference, can you be sure which bag has the larger population mean? Why or why not?**

   No, it is possible that you could get a sample mean that is larger than the population mean of Bag and then get a sample mean that is smaller than the population mean of Bag , or vice versa.

9. **Is your difference in sample means the same as your neighbors' differences? Why or why not?**

   No. As the samples will vary due to sampling variability, so will the means of each sample.

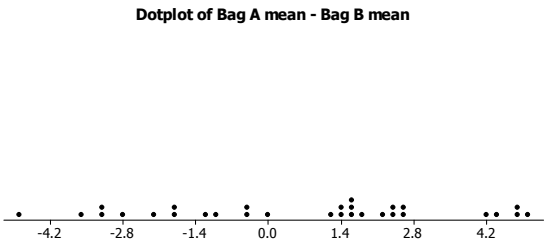
Next have the students plot their differences in sample means on a dot plot. A number line similar to the following may be drawn on the board or chart paper. If necessary, make the number line longer to include all of the differences.



Sample Mean A - Sample Mean B

10. **Plot your difference of the means, , on a class dot plot. Describe the distribution of differences plotted on the graph. Remember to discuss center and spread.**

Provide each student an opportunity to place his or her difference of the means on this class dot plot. The distribution of the differences is expected to cluster around . One example for a class of students is shown here.



Exercises 11–13 (5 minutes)

Reveal to students that the population of numbers in Bag is identical to the population of numbers in Bag .

Use Exercises 11–13 for a class discussion of why the differences in sample means are not all equal to even though the bags are the same. Point out that sampling variability will result in different means (and thus differences that are not all). Students should also think about why the differences in sample means may be relatively far from . (For example, why are some differences or more?) In Exercise 12, point out that one reason for this is that a student could have randomly selected a sample from Bag in which the calculated sample mean was higher than the population mean of Bag . Then the same student could have randomly selected a sample from Bag in which the calculated sample mean was lower than the population mean of Bag . The difference in these sample means would then have a larger value. If there is an example provided by one of the students to illustrate this type of difference, use it to show that some larger differences in sample means are possible.

Exercises 11–13

11. **Why are the differences in the sample means of Bag and Bag not always ?**

Discuss this question in the context of sampling variability. Students should comment on the fact that in order for the difference in sample means to be , the sample means must be the same value. This would rarely happen when selecting random samples.

12. **Does the class dot plot contain differences that were relatively far away from ? If yes, why do you think this happened?**

After students have a chance to respond to this question, discuss the points made above about why some larger differences might occur. For the dot plot given as an example above, some differences were as large as .

13. **Suppose you will take a sample from a new bag. How big would the difference in the sample mean for Bag and the sample mean for the new bag have to be before you would be convinced that the population mean for the new bag is different than the population mean of Bag ? Use the class dot plot of the differences in sample means for Bags and (which have equal population means) to help you answer this question.**

Students should recognize that the difference would need to be relatively far away from . They may give answers like "a difference of (or larger)" or something similar. Remind students that the differences noted in the class dot plot are a result of sampling from bags that have the same numbers in them. As a result, you would expect students to suggest values that are greater than the values in the class dot plot.

The differences in the class dot plot occur because of sampling variability—the chance variability from one sample to another. In Exercise 13, you were asked about how great the difference in sample means would need to be before you have convincing evidence that one population mean is larger than another population mean. A "*meaningful*" difference between two sample means is one that is unlikely to have occurred by chance if the population means are equal. In other words, the difference is one that is greater than would have been expected just due to sampling variability.

How large the difference in sample means needs to be in order to say that it is unlikely to have occurred by chance depends on the context in which the data were collected, the sample size, and how much variability there is in the population, so it is not possible to give a general rule (such as "bigger than ") that you can always use. In the lessons that follow, you will explore one method for determining if a difference in sample means is *meaningful* that depends on expressing the difference in sample means in terms of a measure of how much variability there is in the population.

Exercise 14–16 (5–7 minutes)

Students examine similar questions with Bag . After students develop their answers to the exercises, discuss their responses as a class.

Exercises 14–17

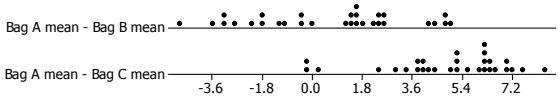
14. Calculate the sample mean of Bag minus the sample mean of Bag , .

Answers will vary as the samples collected by students will vary. Students might suspect, however, that you are setting them up for a discussion about populations that have different means. As a result, ask students what they think their difference is indicating about the populations of the two bags. For several students, this difference is larger than the difference they received for Bags and , and might suggest that the means of the bags are different. Not all students, however, will have differences that are noticeably different from what they obtained for bags and , and as a result, they will indicate that the bags could have the same or similar distribution of numbers.

MP. Students begin to reason quantitatively the meaning of the differences of the sample means. Encourage students to discuss what it means if the difference is positive. For those cases where the difference might have resulted in a negative difference, encourage students to indicate what that indicates about the two populations.

15. Plot your difference on a class dot plot.

Have each student or group of students place their differences on a class dot plot similar to what was developed for the dot plot of the difference of means in Bags and . Place the dot plots next to each other so that students can compare the centers and spread of each distribution. One example based on a class of students is shown here. Notice that the differences for Bag Bag center around , while the differences for Bag Bag do not center around .



16. How do the centers of the class dot plots for and compare?

Students should recognize that the center of the second dot plot is shifted over to the right. Thus, it is not centered at , rather it is centered over a value that is larger than .

Exercise 17 (5 minutes)

17. **Each bag has a population mean that is either or . State what you think the population mean is for each bag. Explain your choice for each bag.**

The population mean is for bags and and for bag . Students indicate their selections based on the class dot plots and the sample means they calculated in the exercises.

Closing (5 minutes)

Ask students to describe how this lesson involved *sampling variability***.** Also, ask students to indicate what the comparison of the dot plots indicated about the population of the numbers in the bags.

Discuss the following Lesson Summary with students.

Lesson Summary

- **Remember to think about sampling variability—the chance variability from sample to sample.**
- **Beware of making decisions based just on the fact that two sample means are not equal.**
- **Consider the distribution of the difference in sample means when making a decision.**

Exit Ticket (3–5 minutes)

Name _____ Date_____

Lesson 21: Why Worry about Sampling Variability?

Exit Ticket

How is a *meaningful* difference in sample means different from a *non-meaningful* difference in sample means? You may use what you saw in the dot plots of this lesson to help you answer this question.

Exit Ticket Sample Solutions

How is a *meaningful* difference in sample means different from a *non-meaningful* difference in sample means? You may use what you saw in the dot plots of this lesson to help you answer this question.

A meaningful difference in sample means is one that is not likely to have occurred by just chance if there is no difference in the population means. A meaningful difference in sample means would be one that is very far from (or not likely to happen if the population means are equal). A non-meaningful difference in sample means would be one that is relatively close to , which indicates the population means are equal.

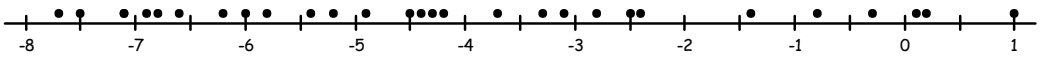
Note that how big this difference needs to be in order to be declared "meaningful" depends on the context, the sample size, and on the variability in the populations. This is explored in the next lesson.

Problem Set Sample Solutions

Below are three dot plots. Each dot plot represents the differences in sample means for random samples selected from two populations (Bag and Bag). For each distribution, the differences were found by subtracting the sample means of Bag from the sample means of Bag (sample mean sample mean).

1. Does the graph below indicate that the population mean of is larger than the population mean of ? Why or why not?

No, since most of the differences are negative, it appears that the population mean of Bag is smaller than the population mean of Bag .



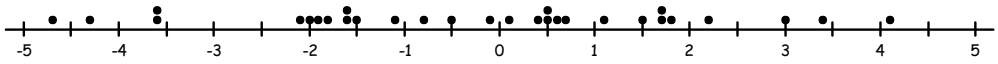
Sample Mean A - Sample Mean B

2. Use the graph above to estimate the difference in the population means .

About , this is about the middle of the graph.

3. Does the graph below indicate that the population mean of is larger than the population mean of ? Why or why not?

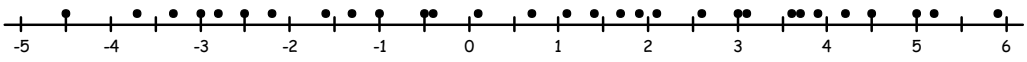
No, the dots are all centered around , meaning that the population means of Bag and Bag might be equal.



Sample Mean A - Sample Mean B

4. Does the graph below indicate that the population mean of is larger than the population mean of ? Why or why not?

Yes, the dots are near . There is a small difference in the population means, but it is so small that it is difficult to detect. (Some students may answer this "No, the dots appear centered around ". Problem 6 should cause students to rethink this answer.)



Sample Mean A - Sample Mean B

5. In the above graph, how many differences are greater than ? How many differences are less than ? What might this tell you?

There are dots greater than , and dots less than . It tells me that there are more positive differences, which may mean that the population mean for Bag is bigger than the population mean for Bag .

6. In Problem 4, the population mean for Bag is really larger than the population mean for Bag . Why is it possible to still get so many negative differences in the graph?

It is possible to get so many negative values because the population mean of Bag may only be a little bigger than the population mean of Bag .

Template for Bags A and B

Template for Bag C