



Lesson 18: Sampling Variability and the Effect of Sample Size

Size

Student Outcomes

- Students use data from a random sample to estimate a population mean.
- Students know that increasing the sample size decreases the sampling variability of the sample mean.

Lesson Notes

In this lesson, students will need their work from the previous lesson. In particular, they will need the population tables, the random-digit table, and the dot plot of student sample mean values. Remind students that the goal of the investigation in the previous lesson was to answer the statistical question, “What is the typical time spent at the gym for people in a population of gym members?” Remind students that a statistical question is a question that can be answered by collecting data *and* anticipating variability in the data collected. (Students were introduced to statistical questions in Grade 6.) The question is revisited in this lesson, where students will see the connection between sampling variability and the size of the sample.

Classwork

Example 1 (6 minutes): Sampling Variability

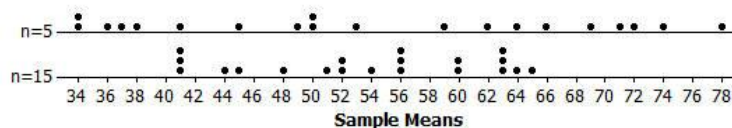
This example reminds students of the concept of sampling variability, and introduces the idea that you want sampling variability to be small.

Let the students think about the questions posed Example 1. In this lesson, students will begin to understand the connection between sampling variability and sample size. For this example, however, have students simply focus on the variability of the two dot plots and the statistical question they are answering.

Discuss the questions posed as a class. Pay careful attention to terminology.

Example 1: Sampling Variability

The previous lesson investigated the statistical question, “What is the typical time spent at the gym?” by selecting random samples from the population of gym members. Two different dot plots of sample means calculated from random samples from the population are displayed below. The first dot plot represents the means of samples with each sample having 5 data points. The second dot plot represents the means of samples with each sample having data points.



Based on the first dot plot, Jill answered the statistical question by indicating the mean time people spent at the gym was between _____ and _____ minutes. She decided that a time approximately in the middle of that interval would be her estimate of the mean time the _____ people spent at the gym. She estimated _____ minutes. Scott answered the question using the second dot plot. He indicated that the mean time people spent at the gym was between _____ and _____ minutes. He also selected a time of _____ minutes to answer the question.

- Describe the differences in the two dot plots.

The first dot plot has a more variability in the sample means than the second dot plot. The sampling variability is greater for the first dot plot.

- Which dot plot do you feel more confident in using to answer the statistical question? Explain your answer.

Possible response: The second dot plot gives me more confidence because the sample means do not differ as much from one another. They are more tightly clustered, so I think I have a better idea of where the population mean is located.

- In general, do you want sampling variability to be large or small? Explain.

The larger the sampling variability, the more that the value of a sample statistic will vary from one sample to another and the further you can expect a sample statistic value to be from the population characteristic. You want the value of the sample statistic to be close to the population characteristic. So, you want sampling variability to be small.

Exercises 1–3 (8 minutes)

Before students begin the exercises, ask them what would minimize sampling variability. Write down some of their suggestions and discuss them. A possible response might be that they would take lots of random samples. Indicate that taking more random samples would not necessarily reduce the sampling variability. If students suggest that increasing the number of observations in a sample would reduce the sampling variability, indicate that this will be covered in the next set of exercises.

Exercises 1–3

In the previous lesson, you saw a population of _____ times spent at the gym. You will now select a random sample of size _____ from that population. You will then calculate the sample mean.

- Start by selecting a three-digit number from the table of random digits. Place the random-digit table in front of you. Without looking at the page, place the eraser end of your pencil somewhere on the table of random digits. Start using the table of random digits at the digit closest to your eraser. This digit and the following two specify which observation from the population will be the first observation in your sample. Write the value of this observation in the space below. (Discard any three-digit number that is _____ or larger, and use the next three digits from the random-digit table.)

Answers will vary.

- Continue moving to the right in the table of random digits from the point that you reached in part (a). Each three-digit number specifies a value to be selected from the population. Continue in this way until you have selected more values from the population. This will make _____ values altogether. Write the values of all _____ observations in the space below.

Answers will vary.

- Calculate the mean of your _____ sample values. Write the value of your sample mean below. Round your answer to the nearest tenth. (Be sure to show your work.)

Answers will vary.

In the previous lesson, students selected a random sample of size n from the population. Here they use the same process to select a random sample of size n . As in the previous lesson, students may want to use calculators in this exercise. Students should work on this exercise individually, but should help their neighbor if that person is having difficulty remembering the process for selecting a random sample from the given population.

Either students or the teacher should write the sample means on the board.

Exercises 4–6 (12 minutes)

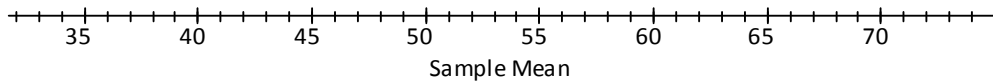
MP.2

Here, all the students' sample means are shown on a dot plot, so that a comparison can be made with the dot plot of sample means from samples of size 5 made in the previous lesson. The fact that the new dot plot (of sample means for samples of size n) shows a smaller spread of values than the dot plot from the previous lesson (of sample means for samples of size 5) is a demonstration of the main point of this lesson: that increasing the sample size decreases the sampling variability of the sample mean. Students are creating a coherent representation of what is happening as the sample size increases. This representation provides students a way to attend to the meaning of what smaller spread of values indicate as the sample size is increased.

Exercises 4–6

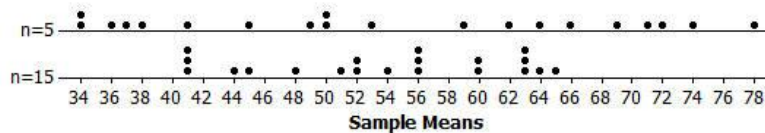
You will now use the sample means from Exercise 3 for the entire class to make a dot plot.

4. Write the sample means for everyone in the class in the space below.
5. Use all the sample means to make a dot plot using the axis given below. (Remember, if you have repeated values or values close to each other, stack the dots one above the other.)



6. In the previous lesson, you drew a dot plot of sample means for samples of size 5. How does the dot plot above (of sample means for samples of size n) compare to the dot plot of sample means for samples of size 5? For which sample size (5 or n) does the sample mean have the greater sampling variability?

The dot plots will vary depending on the results of the random sampling. Dot plots for one set of sample means for random samples of size 5 and for random samples of size n are shown below. The main thing for students to notice is that there is less variability from sample to sample for the larger sample size.



This exercise illustrates the notion that the greater the sample size, the smaller the sampling variability of the sample mean.

If you feel that there is time, and if you think the students will be able to understand the concepts involved, consider asking the question below. (This is a potentially complicated question, so avoid it unless you think students are ready for it.)

- Why do you think the sample mean has less variability when the sample size is large?
 - *When the sample size is small, you could get a sample that contains many large values (making the sample mean large), or many small values (making the sample mean small). But for a larger sample size, it is more likely that the values are equally spread between large and small, and any extreme values are averaged-out with the other values to produce a more central value for the sample mean.*

Exercises 7–8 (5 minutes)

The following exercises ask students to estimate, using the dot plot, a typical degree of accuracy of the sample mean when used as an estimate of the population mean. Use a sample of size n from the population used in this lesson as an example. This is a tricky concept, so you could ask them to work in pairs, or if you think students will have difficulty, you could present this exercise as an example.

The exercise starts with the reminder that in practice you only take *one* sample. Multiple samples are taken in these lessons in order to demonstrate (and analyze) the sampling variability of the sample mean. This is a point that is worth repeating when the opportunity arises.

Exercises 7–8

7. Remember that in practice you only take one sample. Suppose that a statistician plans to take a random sample of size n from the population of times spent at the gym and will use the sample mean as an estimate of the population mean. Based on the dot plot of sample means that your class collected from the population, approximately how far can the statistician expect the sample mean to be from the population mean? (The actual population mean is μ minutes.)

Answers will vary according to the degree of variability that appears in the dot plot and a student’s estimate of an average distance from the population mean. Allow students to use an approximation of μ minutes for the population mean. In the example above, the n samples could be used to estimate the mean distance of the sample means to the population mean of μ minutes.

Sample Mean	Distance from Population Mean

Sample Mean	Distance from Population Mean

The sum of the distances from the mean in the above example is Σ . The mean of these distances, or the expected distance of a sample mean from the population mean, is μ minutes.

8. How would your answer in Exercise 7 compare to the equivalent mean of the distances for a sample of size n ?

Response: My answer for Exercise 7 is smaller than the expected distance for the samples of size n . For samples of size n , several dots are farther from the mean of μ minutes. The mean of the distances for samples of size n would be a larger.

Exercises 9–11 (8 minutes)

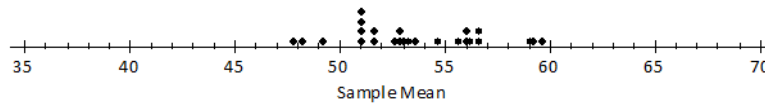
This exercise may be omitted if you are short of time. Here, students are asked to imagine what the dot plot of sample means would look like for samples of size $n = 10$.

Exercises 9–11

Suppose everyone in your class selected a random sample of size $n = 10$ from the population of times spent at the gym.

9. What do you think the dot plot of the class’s sample means would look like? Make a sketch using the axis below.

The student’s sketch should show dots that have less spread than those in the dot plot for samples of size $n = 5$. For example, the student’s dot plot might look like this:



10. Suppose that a statistician plans to estimate the population mean using a sample of size $n = 10$. According to your sketch, approximately how far can the statistician expect the sample mean to be from the population mean?

We were told in Exercise 7 that the population mean is 50. If you calculate the mean of the distances from the population mean (in the same way you did in Exercise 7), the average or expected distance of a sample mean from the population mean is approximately 1.5 minutes for a dot plot similar to the one above. This estimate is made by approximating the average distance of each dot from 50 or 55. Note: If necessary, you can make a chart similar to what was suggested in Exercise 7. Using the dot plot, direct students to estimate the distance of each dot from 50 (a rounding to the nearest whole number is adequate for this question), add up the distances and divide by the number of dots.

11. Suppose you have a choice of using a sample of size $n = 5$, $n = 10$, or $n = 20$. Which of the three makes the sampling variability of the sample mean the smallest? Why would you choose the sample size that makes the sampling variability of the sample mean as small as possible?

Choosing a sample size of $n = 20$ will make the sampling variability of the sample mean the smallest, which is preferable because the sample mean is then more likely to be close to the population mean than it would be for the smaller sample sizes.

Closing (2 minutes)

This question would be particularly beneficial if Exercises 9–11 have been omitted.

- When using a sample mean to estimate a population mean, why do we prefer a larger sample?
 - *A larger sample is preferable because then the sample mean has smaller sampling variability, which means that the sample mean is likely to be closer to the true value of the population mean than it would be for a smaller sample.*

Lesson Summary

Suppose that you plan to take a random sample from a population. You will use the value of a sample statistic to estimate the value of a population characteristic. You want the sampling variability of the sample statistic to be small so that the sample statistic is likely to be close to the value of the population characteristic.

When you increase the sample size, the sampling variability of the sample mean is decreased.

Exit Ticket (5–8 minutes)



Name _____

Date _____

Lesson 18: Sampling Variability and the Effect of Sample Size

Exit Ticket

Suppose that you wanted to estimate the mean time per evening spent doing homework for students at your school. You decide to do this by taking a random sample of students from your school. You will calculate the mean time spent doing homework for your sample. You will then use your sample mean as an estimate of the population mean.

1. The sample mean has “sampling variability.” Explain what this means.
2. When you are using a sample statistic to estimate a population characteristic, do you want the sampling variability of the sample statistic to be large or small? Explain why.
3. Think about your estimate of the mean time spent doing homework for students at your school. Given a choice of using a sample of size or a sample of size , which should you choose? Explain your answer.



Exit Ticket Sample Solutions

Suppose that you wanted to estimate the mean time per evening spent doing homework for students at your school. You decide to do this by taking a random sample of students from your school. You will calculate the mean time spent doing homework for your sample. You will then use your sample mean as an estimate of the population mean.

1. The sample mean has “sampling variability.” Explain what this means.

There are many different possible samples of students at my school, and the value of the sample mean varies from sample to sample.

2. When you are using a sample statistic to estimate a population characteristic, do you want the sampling variability of the sample statistic to be large or small? Explain why.

You want the sampling variability of the sample statistic to be small because then you can expect the value of your sample statistic to be close to the value of the population characteristic that you are estimating.

3. Think about your estimate of the mean time spent doing homework for students at your school. Given a choice of using a sample of size 10 or a sample of size 20, which should you choose? Explain your answer.

I would use a sample of size 20, because then the sampling variability of the sample mean would be smaller than it would be for a sample of size 10.

Problem Set Sample Solutions

1. The owner of a new coffee shop is keeping track of how much each customer spends (in dollars). One hundred of these amounts are shown in the table below. These amounts will form the *population* for this question.

- a. Place the table of random digits in front of you. Select a starting point without looking at the page. Then, taking two digits at a time, select a random sample of size 10 from the population above. Write the values in the space below. (For example, suppose you start at the third digit of row four of the random digit table. Taking two digits gives you 12. In the population above, go to the row labeled “12” and move across to the column labeled “12”. This observation is \$12.00, and that will be the first observation in your sample. Then continue in the random-digit table from the point you reached.)

For example, (starting in the random-digit table at the 8th digit in row 4)

Calculate the mean for your sample, showing your work. Round your answer to the nearest thousandth.

b. Using the same approach as in part (a), select a random sample of size _____ from the population.

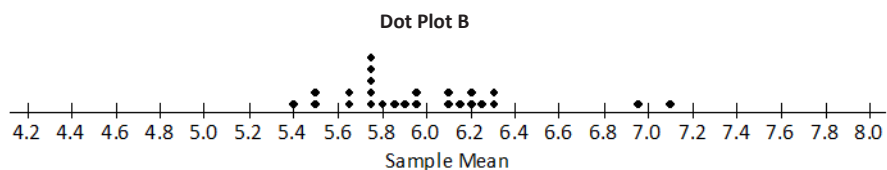
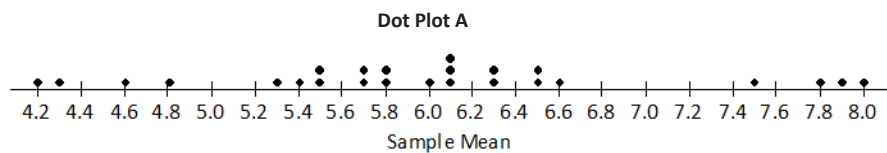
For example, (continuing in the random-digit table from the point reached in Part (a))

Calculate the mean for your sample of size _____. Round your answer to the nearest thousandth.

c. Which of your sample means is likely to be the better estimate of the population mean? Explain your answer in terms of sampling variability.

The sample mean from the sample of size _____ is likely to be the better estimate, since larger samples result in smaller sampling variability of the sample mean.

2. Two dot plots are shown below. One of the dot plots shows the values of some sample means from random samples of size _____ from the population given in Problem 1. The other dot plot shows the values of some sample means from random samples of size _____ from the population given in Problem 1.



Which dot plot is for sample means from samples of size _____, and which dot plot is for sample means from samples of size _____? Explain your reasoning.

The sample means from samples of size _____ are shown in Dot Plot _____.

The variability is greater than in dot plot _____.

The sample means from samples of size _____ are shown in Dot Plot _____.

The variability is smaller compared to dot plot _____, which implies that the sample size was greater.

3. You are going to use a random sample to estimate the mean travel time for getting to school for all the students in your grade. You will select a random sample of students from your grade. Explain why you would like the sampling variability of the sample mean to be *small*.

I would like the sampling variability of the sample mean to be small because then it is likely that my sample mean will be close to the mean time for all students at the school.