Lesson 18: Sampling Variability in the Sample Mean

Student Outcomes

- Students understand the term “sampling variability” in the context of estimating a population mean.
- Students understand that the standard deviation of the sampling distribution of the sample mean offers insight into the accuracy of the sample mean as an estimate of the population mean.

Lesson Notes

MP.4 This is the first of two lessons to build on the concept of sampling variability in the sample mean first developed in Grade 7 (Module 5, Lessons 17–19). Students use simulation to approximate the sampling distribution of the sample mean and explore how the simulated sampling distribution provides information about the anticipated estimation error when using a sample mean to estimate a population mean. Students learn how simulating samples gives us information about how sample means will vary.

Each student or small group of students should have a bag with slips of paper numbered one to 100. (Or, as an alternative, you may have them generate the random numbers using technology.) Prepare a number line on a wall or board that goes from 1–5 with each unit divided into tenths. Students should have post-it notes to post the mean segment lengths in their random samples on the number line, so leave enough space for the post-it notes.

In Exercises 4 and 5, students will need to share the values of the means for their individual samples. You might have them report their means while the others record them (or enter them if they are using technology). To facilitate the process, consider giving students a copy of the values from the table below used in the simulated sampling distribution of means suggested as possible answers to Exercise 3 and Exercise 5. (Or, you may wish to do this as a whole class activity with one student entering the values to avoid errors in entering the data.)

Length of Segments

2	3	3	2	4	8	2	5	2	2
5	5	4	3	1	2	3	2	2	3
1	1	2	4	1	1	4	5	4	3
2	1	1	1	2	2	3	1	8	4
4	2	1	1	3	5	1	1	4	2
1	3	7	3	3	3	1	2	3	3
1	2	4	3	1	1	7	3	1	7
2	2	3	2	3	2	2	1	2	1
3	1	8	4	2	2	1	1	5	3
1	4	1	2	5	3	3	3	5	4

Be sure students label their graphs completely in their answers to the questions. Understanding what they are graphing is an important part of understanding the concepts involved in this exercise.

Classwork

Exercises 1–7 (40 minutes): Random Segments

Provide each student or small group with a copy of the worksheet that is located at the end of this lesson.

Exercises 1–7: Random Segments

The worksheet contains 100 segments of different lengths. The length of a segment is the number of rectangles spanned on the grid. For example, segment 2 has length 5.

1. **Briefly review the sheet and estimate of the mean length of the segments. Will your estimate be close to the actual mean? Why or why not?**

   Answers will vary. Some may estimate 5, others as low as 2.

   Sample response: I estimate the mean length is 4. I believe my estimate will be close to the actual mean because it appears as if a large number of the segments are around 4, and if I average the longer and shorter lengths, the average is also around 4.

2. **Look at the sheet. With which of the statements below would you agree? Explain your reasoning.**

   The mean length of the segments is:

   a. close to 1
   b. close to 8
   c. around 5
   d. between 2 and 5

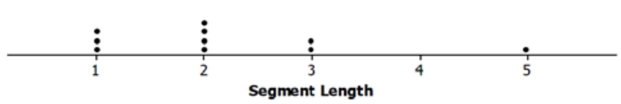
   Possible answers: The only choice that makes sense to me is (d), between 2 and 5. The smallest segment length was 1, so it does not make sense that the mean would be the smallest. The largest segment length was 8, so again, it would not make sense to have the mean be 8 or even 5 because there are a lot of segments of lengths 1 and 2.

Some estimates for the mean are not reasonable because they are lengths of the longest segments, which would not account for lengths that are shorter. (This would also be true for the shortest segment lengths.) While an interval estimate might make sense, it is still hard to know for sure whether that interval is a good estimate. A better way to get a good estimate is to use random samples.

3. **Follow your teacher's directions to select ten random numbers between 1 and 100. For each random number, start at the upper left cell with a segment value of 2, and count down and to the right the number of cells based on the random number selected. The number in the cell represents the length of a randomly selected segment.**

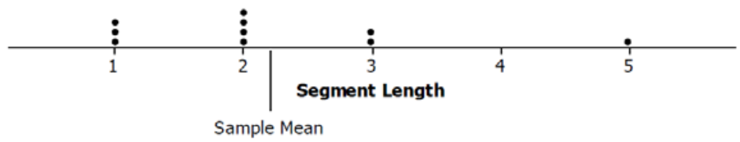
   a. **On a number line, graph the lengths of the corresponding segments on the worksheet.**

   Possible answer: For the random numbers $\{8, 23, 35, 74, 40, 75, 9, 50, 54, 64\}$ the sample lengths would be $\{2, 2, 1, 1, 2, 1, 3, 5, 2, 3\}$. The graph would look like:



b. **Find the mean and standard deviation of the lengths of the segments in your sample. Mark the mean length on your graph from part (a).**

Possible answer: The mean sample length is 2.2 units, and the standard deviation is 1.23 units.

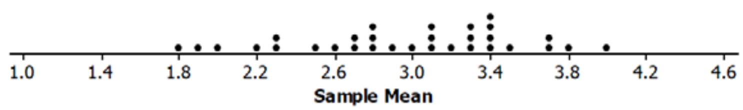


4. **Your sample provides some information about the mean length of the segments in one random sample of size 10, but that sample is only one among all the different possible random samples. Let's look at other random samples and see how the means from those samples compare to the mean segment length from your random sample.**

Record the mean segment length for your random sample on a post-it note, and post the note in the appropriate place on the number line your teacher set up.

Sample response (based on a class with 31 students):

Simulated sampling distribution of mean segment lengths for samples of size 10



a. **Jonah looked at the plot and said, "Wow, our means really varied." What do you think he meant?**

Possible answer: Many of the samples had mean lengths that were different. Some were the same, but most were not.

b. **Describe the simulated sampling distribution of mean segment lengths for samples of size 10.**

Possible answer: The maximum mean segment length in our samples of size 10 was 4.0, and the minimum was 1.7. The sample means seemed to center around 3, with most of the segments from about 2.5 to 3.5 units long.

c. **How did your first estimate (from Exercise 1) compare to your sample mean from the random sample? How did it compare to the means in the simulated distribution of the sample means from the class?**

Possible answer: My estimate was way off. I thought the mean length would be larger, like maybe 4.5 units. My sample mean was only 2.2 units long, which was smaller than all but three of the other sample means.

5. Collect the values of the sample means from the class.

   a. Find the mean and standard deviation of the simulated distribution of the sample means.

      Sample response (based on the 31 sample means used to produce the answer to Exercise 3): The mean of the simulated distribution of sample means is 2.97, and the standard deviation is 0.57.

   b. Interpret the standard deviation of the simulated sampling distribution in terms of the length of the segments.

      Sample response: A typical distance of a sample mean from the center of the sampling distribution is about 0.57.

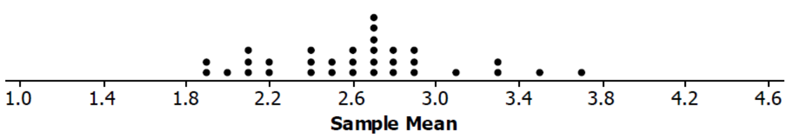
   c. What do you observe about the values of the means in the simulated sampling distribution that are within two standard deviations from the mean of the sampling distribution?

      Sample response: All of the sample means were within two standard deviations from the mean of the sampling distribution, from 1.53 to 4.11.

6. Generate another set of ten random numbers, find the corresponding lengths on the sheet, and calculate the mean length for your sample. Put a post-it note with your sample mean on the second number line. Then answer the following questions:

   Sample response (based on a class with 31 students):

   Second simulated sampling distribution of mean segment lengths for 33 samples of size 10



   a. Find the mean and standard deviation of the simulated distribution of the sample means.

      Sample response: The mean is 2.63 units, and the standard deviation is 0.44 units.

   b. Interpret the standard deviation of the simulated sampling distribution in terms of the length of the segments.

      Sample response: A typical distance of a sample mean from the center of the sampling distribution is about 0.44.

   c. What do you observe about the values of the means in the simulated sampling distribution that are within two standard deviations from the mean of the sampling distribution?

      Sample response: Only one or two sample means were not within two standard deviations from the mean of the sampling distribution, 1.75 to 3.51.

MP.2

7. Suppose that we know the actual mean of all the segment lengths is 2.78 units.

a. Describe how the population mean relates to the two simulated distributions of sample means.

Possible answer: The simulated sampling distributions of sample means were both centered around values very close to the population mean.

b. Tonya was concerned that neither of the simulated distributions of sample means had a value around 5, but some of the segments on the worksheet were 5 units long and some were as big as 8 units long. What would you say to Tonya?

Possible answer: The simulated sampling distribution was of the means of the samples, not of individual segment lengths. It could be possible to have a mean length of 5 from a sample of ten segment lengths, but it is not very likely.

Closing (2–3 minutes)

- Why is the concept of *random* samples important in exploring how a simulated sampling distribution provides information about the anticipated estimation error when using a sample mean to estimate a population mean?
    - *Possible answer: If the samples are not random, the distribution of the sample means might not have centers close to the population mean, and the standard deviations of different sampling distributions might not tell the same story about the distributions of the sample means.*
- What is the difference between a "sample mean" and a "distribution of sample means"? (You may use the segment lengths in your answer.)
    - *Possible answer: A sample mean is the mean of the values of the segment lengths in a given sample. A distribution of sample means is the distribution of all sorts of sample means calculated from many different samples.*

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

Lesson Summary

In this lesson you drew a sample from a population and found the mean of that sample.

- **Drawing many samples of the same size from the same population and finding the mean of each of those samples allows you to build a simulated sampling distribution of the sample means for the samples you generated.**
- **The mean of the simulated sampling distribution of sample means is close to the population mean.**
- **In the two examples of simulated distributions of sample means we generated, most of the sample means seemed to fall within two standard deviations of the mean of the simulated distribution of sample means.**

Exit Ticket (3 minutes)

Name ____ Date ____________________

Lesson 18: Sampling Variability in the Sample Mean

Exit Ticket

Describe what a “simulated distribution of sample means” is and what the “standard deviation of the distribution” indicates. You may want to refer to the segment lengths in your answer.

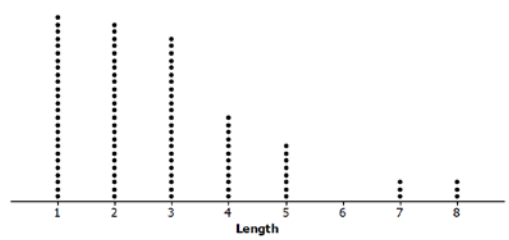
Exit Ticket Sample Solutions

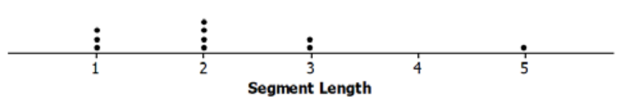
Describe what a "simulated distribution of sample means" is and what the "standard deviation of the distribution" indicates. You may want to refer to the segment lengths in your answer.

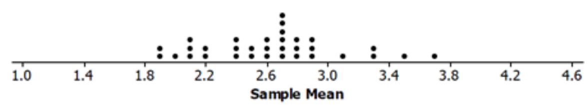
Possible answer: You draw samples of a given size from a population (here it was the 100 segment lengths), find the mean segment length of each sample, and plot the sample mean lengths. The resulting distribution of the sample means from those random samples is the simulated distribution of sample means. The standard deviation of that distribution gives you an idea of how the sample means vary from sample to sample.

Problem Set Sample Solutions

1. **The three distributions below relate to the population of all of the random segment lengths and to samples drawn from that population. The eight phrases below could be used to describe a whole graph or a value on the graph. Identify where on the appropriate graph the phrases could be placed. (For example, segment of length 2 could be placed by any of the values in the column for 2 on the plot labeled "Length.")**



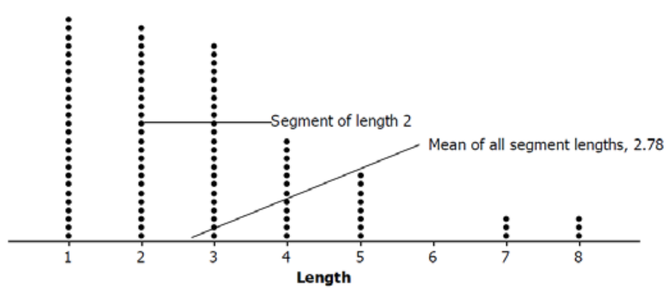




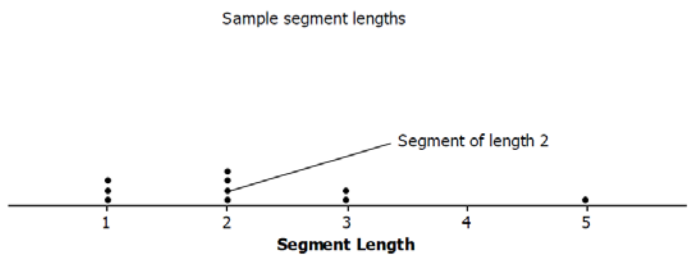
a. Random sample of size 10 of segment lengths.
b. Segment of length 2
c. Sample mean segment length of 2
d. Mean of sampling distribution, 2.6
e. Simulated distribution of sample means
f. Sample segment lengths
g. Population of segment lengths
h. Mean of all segment lengths, 2.78

Possible answers: (Note that segment of length 2 could be used in more than one place and on more than one graph.)

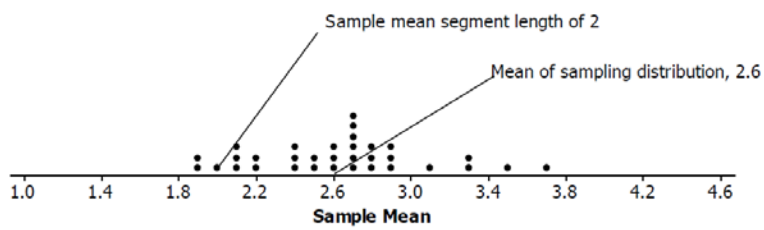
Population of segment lengths



Random sample of size 10 of segment lengths



Simulated distribution of sample means



2. The following segment lengths were selected in four different random samples of size 10.

Note: If students work in small groups, you might indicate that each group member work on a different sample.

Lengths Sample A	Lengths Sample B	Lengths Sample C	Lengths Sample D
1	1	1	2
2	3	5	2
1	1	1	7
5	2	3	2
3	1	4	5
1	5	2	2
2	3	2	3
2	4	4	5
3	3	3	5
1	3	4	4

a. **Find the mean segment length of each sample.**

Sample A mean is 2.1 (1.29); Sample B mean is 2.6 (1.35); Sample C mean is 2.9 (1.37); Sample D mean is 3.7 (1.77).

The standard deviation of each sample is noted in parentheses as reference for part (b).

b. **Find the mean and standard deviation of the four sample means.**

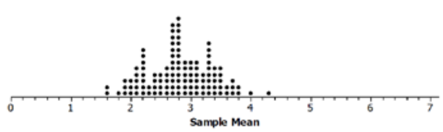
The mean of the sample means is 2.825, and the standard deviation of the sample means is 0.67.

c. **Interpret your answer to part (b) in terms of the variability in the sampling process.**

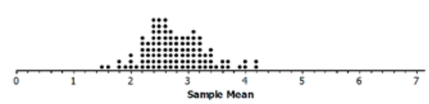
Possible answer: A typical distance of a sample mean from the mean of the four samples (2.825) is 0.67.

3. **Two simulated sampling distributions of the mean segment lengths from random samples of size 10 are displayed below.**

Distribution A



Distribution B



a. **Compare the two distributions with respect to shape, center, and spread.**

Possible answer: Both distributions are mound shaped with the center a bit below 3, about 2.8. The maximum mean segment length in both is about 4.2 units, and the minimum is about 1.5 or 1.6. Most of the sample means in both distributions are between about 2 and 4.

b. **Distribution A has a mean of 2.82, and Distribution B has a mean of 2.77. How do these means compare to the population mean of 2.78?**

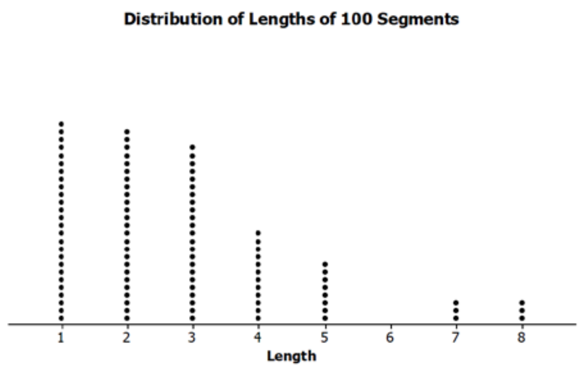
Possible answer: The mean segment lengths of the two simulated distributions of sample means are very close to the actual mean segment length.

c. **Both Distribution A and Distribution B have a standard deviation 0.54. Make a statement about the distribution of sample means that makes use of this standard deviation.**

Answers will vary. Possible answers include:

- *Almost all of the sample means in Distribution A are within two standard deviations of the mean of the sample means, 1.74 to 3.90. The same is true for the sample means in Distribution B; the sample means are almost all from 1.69 to 3.85.*
- *A typical distance of a sample mean from the center of the sampling distribution is 0.54.*

4. **The population distribution of all the segment lengths is shown in the dot plot below. How does the population distribution compare to the two simulated sampling distributions of the sample means in Problem 3?**



Possible answer: The distribution of all of the lengths is skewed right. Most of the lengths were 1, 2, or 3 units. The simulated sampling distributions of the sample means were both mound shaped, with the centers about the same, and not like the shape of the population.

Exercises 1–7: Random Segments