



## Lesson 17: Sampling Variability

### Student Outcomes

- Students use data from a random sample to estimate a population mean.
- Students understand the term “sampling variability” in the context of estimating a population mean.

### Lesson Notes

Consider starting this lesson with a review of the activities from the last lesson. The number of total number of words for books in a random sample of children’s books was used to estimate the mean total number of words for the population of children’s books. A random sample of prices from a population of grocery prices was used to estimate the proportion of grocery prices that would round up to the nearest dollar. This lesson and the lessons that follow continue to develop the idea of using data from a random sample to estimate a population mean or a population proportion. Sampling variability in sample means and sample proportions is explored in order to further student understanding of this important concept. In later grades, students learn how to take sampling variability into account when generalizing from a random sample to the population from which it was selected. In this lesson, students build on their developing understanding of sampling variability from the previous lessons by constructing a dot plot of the means of different random samples selected from the same population.

In previous lessons, students have selected random samples by selecting items from a bag or by using a random-number generator from a website or graphing calculator. This lesson adds another tool that can be used in selecting random samples, namely a random-number table. A random-number table is just a table of digits from 0 to 9 that are arranged in a random order. A random-number table and directions for using it to select a random sample are provided as part of the lesson. Students may ask how the numbers in a random-number table are generated. Today, computer programs are used to generate random number tables, but the idea behind these tables is relatively simple. The digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 are placed in a bag or jar (with each digit written on a slip of paper or a chip). The digits are thoroughly mixed. A slip or chip is picked. The number on the slip or chip is recorded, and then returned to the bag or jar. This process is repeated, generating a random list of 0’s to 9’s.

## Classwork

## Example 1 (5 minutes): What is a Sample Mean?

## Example 1: Estimating a Population Mean

The owners of a gym have been keeping track of how long each person spends at the gym. Eight hundred of these times (in minutes) are shown in the population tables located at the end of the lesson. These times will form the population that you will investigate in this lesson.

Look at the values in the population. Can you find the longest time spent in the gym in the population? Can you find the shortest?

On average, roughly how long do you think people spend at the gym? In other words, by just looking at the numbers in the two tables, make an estimate of the population mean.

You could find the population mean by typing all numbers into a calculator or a computer, adding them up, and dividing by . This would be extremely time-consuming, and usually it is not possible to measure every value in a population.

Instead of doing a calculation using every value in the population, we will use a random sample to find the mean of the sample. The sample mean will then be used as an estimate of the population mean.

In this example, students see a population of times spent by people at a gym. The times are displayed in the population tables at the end of the lesson. Convey to students that it is not practical to use all times to calculate the mean.

It is important to make students aware how unusual it would be to know all of the population values. Indeed, if you do have all the population values, then there is little point in taking a sample because you could calculate the population mean (using a computer if necessary).

If there is time, ask the students to:

- Think of a population where you do not have all the population values.
  - *Virtually any population will suffice in answer to this question; for example, the heights of all the adults in the town where the school is located.*
- Think of another population where you *do* have all the population values.
  - *This is less common, but one example might be the salaries of all the employees of a company. All of the employee salaries will be stored in the company's computer.*

Discuss the questions in the example. Provide students time to think about and respond to the questions. The following responses provide a summary of the important points students should see as they work through the exercise.

- Look at the values in the population. Can you find the longest time spent in the gym in the population? Can you find the shortest?
  - *The longest time is min., and the shortest time is min. Students will provide other estimates, but direct them to these times.*
- On average, roughly how long do you think people spend at the gym? In other words, by just looking at the numbers in the two tables, make an estimate of the *population mean*.
  - *It seems from the numbers in the tables that the mean time spent at the gym is around min.*

**Example 2 (5–7 minutes): Selecting a Sample Using a Table of Random Digits**

Carefully discuss the sample directions outlined in this section. The goal is to recognize that each of the population values has an equal chance in being selected into the sample. Introduce the random-number table and as students work through this example, make sure that they understand how the table is used to identify individuals to be included in the sample. Once a sample is selected, the sample mean is calculated for this sample, and students are encouraged to think about how different samples would result in different values of the sample mean (this being an example of sampling variability).

**Example 2: Selecting a Sample Using a Table of Random Digits**

The table of random digits provided with this lesson will be used to select items from a population to produce a random sample from the population. The list of digits are determined by a computer program that simulates a random selection of the digits , , , , , , , , or . Imagine that each of these digits is written on a slip of paper and placed in a bag. After thoroughly mixing the bag, one slip is drawn and its digit is recorded in this list of random digits. The slip is then returned to the bag and another slip is selected. The digit on this slip is recorded, and then returned to the bag. The process is repeated over and over. The resulting list of digits is called a random-number table.

Encourage students to think about the following question and to possibly share their suggestions prior to discussing the subsequent steps with all students.

**How could you use a table of random digits to take a random sample?**

**Step 1:** Place the table of random digits in front of you. Without looking at the page, place the eraser end of your pencil somewhere on the table. Start using the table of random digits at the number closest to where your eraser touched the paper. This digit and the following two specify which observation from the population tables will be the first observation in your sample.

For example, suppose the eraser end of your pencil lands on the twelfth number in row of the random digit table. This number is and the two following numbers are and . This means that the first observation in your sample is observation number from the population. Find observation number in the population table. Do this by going to Row and moving across to the column heading “ ” This observation is , so the first observation in your sample is .

If the number from the random-number table is any number or greater, you will ignore this number and use the next three digits in the table.

Point out that four more selections are to be made to complete the random sample of five. Emphasize that the next four observations are determined by continuing on from the starting position in the random number table. The following directions and observations are important for students to understand:

**Step 2:** Continue using the table of random digits from the point you reached, and select the other four observations in your sample like you did above.

For example, continuing on from the position in the example given in Step 1:

- The next number from the random-digit table is , and observation is
- The next number from the random-digit table is , and observation is .
- The next number from the random-digit table is , and there is no observation . So we ignore these three digits.
- The next number from the random-digit table is , and there is no observation . So we ignore these three digits.
- The next number from the random-digit table is , and observation is .
- The next number from the random-digit table is , and observation is .

**Exercises 1–4 (8–10 minutes)**

Students should work independently on Exercises 1–4. Then discuss and confirm as a class.

**Exercises 1–4**

Initially, you will select just five values from the population to form your sample. This is a very small sample size, but it is a good place to start to understand the ideas of this lesson.

The values found in Example 2 are used to illustrate the answers to the following questions, but each student should select their own random sample of five.

1. Use the table of random number to select five values from the population of times. What are the five observations in your sample?

*Make sure students write down these values and understand how they were selected. In the example given, the following observations were selected: min., min., min., min., and min.*

2. For the sample that you selected, calculate the sample mean.

*For the given example, the sample mean is \_\_\_\_\_ min.*

3. You selected a random sample and calculated the sample mean in order to estimate the population mean. Do you think that the mean of these five observations is exactly correct for the population mean? Could the population mean be greater than the number you calculated? Could the population mean be less than the number you calculated?

*Make sure that students see that the value of their sample mean ( \_\_\_\_\_ minutes in the given example, but students will have different sample means) is not likely to be exactly correct for the population mean. The population mean could be greater or less than the value of the sample mean.*

4. In practice, you only take one sample in order to estimate a population characteristic. But, for the purposes of this lesson, suppose you were to take another random sample from the same population of times at the gym. Could the new sample mean be closer to the population mean than the mean of these five observations? Could it be further from the population mean?

*Make sure students understand that if they were to take a new random sample, the new sample mean is unlikely to be equal to the value of the sample mean of their first sample. It could be closer to or further from the population mean.*

**Exercises 5–7 (7 minutes)**

Before students begin the following exercises, put all of their sample means on the board or poster paper. Then, ask the following questions: What do you notice about the sample means? Why do you think this is true?

Anticipate that students will indicate that the sample means posted are different and that several of them are very close to each other. They may also point out how different some of them are from each other. Students should recognize that the samples come from the same population, but that the actual selections can often be different. The different selections result in different sample means. Point out that the sample mean is an example of a sample statistic. The variation of sample statistics from sample to sample is called sampling variability.

Here, each student selects a random sample from the population and calculates the sample mean. Students should work individually, helping their neighbors as needed. The use of a calculator to calculate the sample mean should be encouraged to save time and increase accuracy. Point out that for this exercise, students should determine the mean to the nearest tenth of a minute.

#### Exercises 5–7

As a class, you will now investigate sampling variability by taking several samples from the same population. Each sample will have a different sample mean. This variation provides an example of sampling variability.

5. Place the table of random digits in front of you and, without looking at the page, place the eraser end of your pencil somewhere on the table of random numbers. Start using the table of random digits at the number closest to where your eraser touches the paper. This digit and the following two specify which observation from the population tables will be the first observation in your sample. Write this three-digit number and the corresponding data value from the population in the space below.

*Response: Answers will vary based on the numbers selected.*

6. Continue moving to the right in the table of random digits from the place ended in Exercise 5. Use three digits at a time. Each set of three digits specifies which observation in the population is the next number in your sample. Continue until you have four more observations, and write these four values in the space below.

*Response: Answers will vary based on the numbers selected.*

7. Calculate the mean of the five values that form your sample. Round your answer to the nearest tenth. Show your work and your sample mean in the space below.

*Response: Answers will vary based on the numbers selected.*

#### Exercises 8–11 (11 minutes)

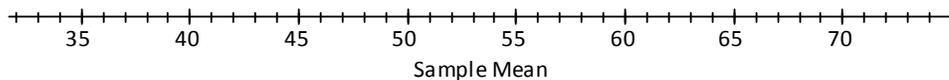
#### Exercises 8–11

You will now use the sample means from Exercise 7 from the entire class to make a dot plot.

8. Write the sample means for everyone in the class in the space below.

*Response: Answers will vary based on collected sample means.*

9. Use all the sample means to make a dot plot using the axis given below. (Remember, if you have repeated or close values, stack the dots one above the other.)



10. What do you see in the dot plot that demonstrates sampling variability?

*Sample response: The dots are spread out indicating that the sample means are not the same. The results indicate what we discussed as sampling variability. (See explanation provided at the end of this lesson.)*

11. Remember that in practice you only take one sample. (In this lesson, many samples were taken in order to demonstrate the concept of sampling variability.) Suppose that a statistician plans to take a random sample of size from the population of times spent at the gym and that he or she will use the sample mean as an estimate of the population mean. Approximately how far can the statistician expect the sample mean to be from the population mean?

*Answers will vary. Allow students to speculate as to what the value of the population mean might be and how far a sample mean would be from that value. The distance students indicate will be based on the dot plot. Students may indicate the sample mean to be exactly equal to the population and that the distance would be . They may also indicate the sample mean to be one of the dots that is a minimum or a maximum of the distribution, and suggest a distance from the minimum or maximum to the center of the distribution. Students should begin to see that the distribution has a center that they suspect is close to the population's mean.*

MP.

Here the sample means for the class are combined in order to form a visual illustration of sampling variability (in the form of a dot plot of the class's sample means. Students are asked to show how the pattern in the dot plot supports the notion of sampling variability. Students will need to refer to the dot plot in their work in the next lesson, so every student will need to complete the dot plot.

Exercise 10 summarizes an important part of this lesson. The fact that the sample mean varies from sample to sample is an example of *sampling variability*. Clearly emphasize this summary.

Exercise 11 covers an important concept that is worth thinking about in advance of the class. When you take a sample and calculate the sample mean, you know that the value of the sample mean is very unlikely to be *exactly* equal to the population mean. But how close can you expect it to be? The dot plot constructed in this lesson provides a way to answering this question (at least for a random sample of size from the population presented in this class.) The population mean is approximately in the *center* of the dots in the dot plot constructed in Exercise 9. Indicate that the distances of the dots from the population mean provides information about how far the statistician can expect the sample mean to be from the population mean.

An important point is also made at the beginning of Exercise 11, which is worth emphasizing whenever the opportunity arises in the remaining lessons. In practice, you only take *one* sample from a population. If you took two samples, then the sensible thing would be to combine them into one larger sample.

We only take multiple samples in this lesson in order to demonstrate the concept of sampling variability.

- Suppose that you are going to take *one* sample from a population. Once you have your sample, you will calculate the sample mean. Explain what is meant by sampling variability in this context.
  - *Before taking your sample, you know that there are many different sets of individuals who could form your sample. Every different possible sample could give a different value for the sample mean. This is known as sampling variability.*

It is important that students be aware that the idea of sampling variability does not apply only to the sample mean. Suppose, for example, an advertising agency plans to take a random sample of people and ask those people whether they have seen a particular commercial. The population is all the people who could be asked the question. The *proportion* of people in the sample who have seen the commercial is a sample statistic, and the value of this sample statistic will vary from sample to sample. This statistic also has sampling variability. Simulations will be developed in later lessons to illustrate sampling variability of proportions calculated using data from random samples of a population. Lesson 20 begins students' investigations with the proportion as the statistic of interest.







Table of Random Digits

--	--

**Closing (4 minutes)**

If there is time, ask students these questions:

- You have learned today about sampling variability. Do you want sampling variability to be large or small?
  - *We want sampling variability to be small.*
- Why do you want sampling variability to be small?
  - *Suppose you are using the sample mean to estimate a population mean. You want the sampling variability of the sample mean to be small, so that you can expect your value of the sample mean to be close to the population mean.*

**Lesson Summary**

A population characteristic is estimated by taking a random sample from the population and calculating the value of a statistic for the sample. For example, a population mean is estimated by selecting a random sample from the population and calculating the sample mean.

The value of the sample statistic (e.g., the sample mean) will vary based on the random sample that is selected. This variation from sample to sample in the values of the sample statistic is called sampling variability.

**Exit Ticket (5–8 minutes)**



Name \_\_\_\_\_

Date \_\_\_\_\_

## Lesson 17: Sampling Variability

### Exit Ticket

Suppose that you want to estimate the mean time per evening students at your school spend doing homework. You will do this using a random sample of \_\_\_\_\_ students.

1. Suppose that you have a list of all the students at your school. The students are numbered \_\_\_\_\_. One way to select the random sample of students is to use the random-digit table from today's class, taking three digits at a time. If you start at the third digit of row \_\_\_\_\_, what is the number of the first student you would include in your sample?
2. Suppose that you have now selected your random sample and you have asked each student how long he or she spends doing homework each evening. How will you use these results to estimate the mean time spent doing homework for *all* students?
3. Explain what is meant by *sampling variability* in this context.

## Exit Ticket Sample Solutions

Suppose that you want to estimate the mean time per evening students at your school spend doing homework. You will do this using a random sample of \_\_\_\_\_ students.

- Suppose that you have a list of all the students at your school. The students are numbered \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_, \_\_\_\_\_. One way to select the random sample of students is to use the random-digit table from today's class, taking three digits at a time. If you start at the third digit of row \_\_\_\_\_, what is the number of the first student you would include in your sample?

*The first student in the sample would be student number \_\_\_\_\_.*

- Suppose that you have now selected your random sample, and you have asked each student how long he or she spends doing homework each evening. How will you use these results to estimate the mean time spent doing homework for *all* students?

*I would calculate the mean time spent doing homework for the students in my sample.*

- Explain what is meant by *sampling variability* in this context.

*Different samples of students would result in different values of the sample mean. This is sampling variability of the sample mean.*

## Problem Set Sample Solutions

- Yousef intends to buy a car. He wishes to estimate the mean fuel efficiency (in miles per gallon) of all cars available at this time. Yousef selects a random sample of \_\_\_\_\_ cars and looks up their fuel efficiencies on the Internet. The results are shown below.

- Yousef will estimate the mean fuel efficiency of all cars by calculating the mean for his sample. Calculate the sample mean, and record your answer below. (Be sure to show your work.)

\_\_\_\_\_

- In practice, you only take one sample to estimate a population characteristic. However, if Yousef were to take another random sample of \_\_\_\_\_ cars from the same population, would he likely get the same value for the sample mean?

*No, it is not likely that Yousef would get the same value for the sample mean.*

- What if Yousef were to take *many* random samples of \_\_\_\_\_ cars? Would the entire sample means be the same?

*No, he could get many different values of the sample mean.*

- Using this example, explain what sampling variability means.

*The fact that the sample mean will vary from sample to sample is an example of sampling variability.*



2. Think about the mean number of siblings (brothers and sisters) for all students at your school.

- a. Approximately what do you think is the value of the mean number of siblings for the population of all students at your school?

*Answers will vary.*

- b. How could you find a better estimate of this population mean?

*I could take a random sample of students, ask each student in my sample how many siblings he or she has and find the mean for my sample.*

- c. Suppose that you have now selected a random sample of students from your school. You have asked all of the students in your sample how many siblings they have. How will you calculate the sample mean?

*I will add up all of the values in the sample and divide by the number of students in the sample.*

- d. If you had taken a different sample, would the sample mean have taken the same value?

*No, a different sample would generally produce a different value of the sample mean. It is possible, but unlikely, that the sample mean for a different sample would have the same mean.*

- e. There are many different samples of students that you could have selected. These samples produce many different possible sample means. What is the phrase used for this concept?

*Sampling variability.*

- f. Does the phrase you gave in part (e) apply only to sample means?

*No, the concept of sampling variability applies to any sample statistic.*