# Lesson 18: Sampling Variability and the Effect of Sample Size 

## Classwork

## 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 800 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 20 samples with each sample having 5 data points. The second dot plot represents the means of 20 samples with each sample having 15 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 34 and 78 minutes. She decided that a time approximately in the middle of that interval would be her estimate of the mean time the 800 people spent at the gym. She estimated 52 minutes. Scott answered the question using the second dot plot. He indicated that the mean time people spent at the gym was between 41 and 65 minutes. He also selected a time of 52 minutes to answer the question.

- Describe the differences in the two dot plots.
- Which dot plot do you feel more confident in using to answer the statistical question? Explain your answer.
- In general, do you want sampling variability to be large or small? Explain.


## Exercises 1-3

In the previous lesson, you saw a population of 800 times spent at the gym. You will now select a random sample of size 15 from that population. You will then calculate the sample mean.

1. 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 800 or larger, and use the next three digits from the random digit table.)
2. Continue moving to the right in the table of random digits from the point that you reached in Exercise 1. Each three-digit number specifies a value to be selected from the population. Continue in this way until you have selected 14 more values from the population. This will make 15 values altogether. Write the values of all 15 observations in the space below.
3. Calculate the mean of your 15 sample values. Write the value of your sample mean below. Round your answer to the nearest tenth. (Be sure to show your work.)

## 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 15) compare to the dot plot of sample means for samples of size 5? For which sample size (5 or 15 ) does the sample mean have the greater sampling variability?

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

## 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 15 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 53.9 minutes.)
8. How would your answer in Exercise 7 compare to the equivalent mean of the distances for a sample of size 5?

## Exercises 9-11

Suppose everyone in your class selected a random sample of size 25 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.

10. Suppose that a statistician plans to estimate the population mean using a sample of size 25 . According to your sketch, approximately how far can the statistician expect the sample mean to be from the population mean?
11. Suppose you have a choice of using a sample of size 5,15 , or 25 . 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?

## 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.

## Problem Set

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.

|  | $\mathbf{0}$ | $\mathbf{1}$ | $\mathbf{2}$ | $\mathbf{3}$ | $\mathbf{4}$ | $\mathbf{5}$ | $\mathbf{6}$ | $\mathbf{7}$ | $\mathbf{8}$ | $\mathbf{9}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\mathbf{0}$ | 6.18 | 4.67 | 4.01 | 4.06 | 3.28 | 4.47 | 4.86 | 4.91 | 3.96 | 6.18 |
| $\mathbf{1}$ | 4.98 | 5.42 | 5.65 | 2.97 | 2.92 | 7.09 | 2.78 | 4.20 | 5.02 | 4.98 |
| $\mathbf{2}$ | 3.12 | 1.89 | 4.19 | 5.12 | 4.38 | 5.34 | 4.22 | 4.27 | 5.25 | 3.12 |
| $\mathbf{3}$ | 3.90 | 4.47 | 4.07 | 4.80 | 6.28 | 5.79 | 6.07 | 7.64 | 6.33 | 3.90 |
| $\mathbf{4}$ | 5.55 | 4.99 | 3.77 | 3.63 | 5.21 | 3.85 | 7.43 | 4.72 | 6.53 | 5.55 |
| $\mathbf{5}$ | 4.55 | 5.38 | 5.83 | 4.10 | 4.42 | 5.63 | 5.57 | 5.32 | 5.32 | 4.55 |
| $\mathbf{6}$ | 4.56 | 7.67 | 6.39 | 4.05 | 4.51 | 5.16 | 5.29 | 6.34 | 3.68 | 4.56 |
| $\mathbf{7}$ | 5.86 | 4.75 | 4.94 | 3.92 | 4.84 | 4.95 | 4.50 | 4.56 | 7.05 | 5.86 |
| $\mathbf{8}$ | 5.00 | 5.47 | 5.00 | 5.70 | 5.71 | 6.19 | 4.41 | 4.29 | 4.34 | 5.00 |
| $\mathbf{9}$ | 5.12 | 5.58 | 6.16 | 6.39 | 5.93 | 3.72 | 5.92 | 4.82 | 6.19 | 5.12 |

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 10 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 19. In the population above, go to the row labeled 1, and move across to the column labeled 9. This observation is 4.98, and that will be the first observation in your sample. Then, continue in the random digit table from the point you reached.)
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 20 from the population.

Calculate the mean for your sample of size 20. 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.
2. Two dot plots are shown below. One of the dot plots shows the values of some sample means from random samples of size 10 from the population given in Problem 1. The other dot plot shows the values of some sample means from random samples of size 20 from the population given in Problem 1.

## Dot Plot A



Which dot plot is for sample means from samples of size 10 , and which dot plot is for sample means from samples of size 20? Explain your reasoning.

The sample means from samples of size 10 are shown in Dot Plot $\qquad$ -.

The sample means from samples of size 20 are shown in Dot Plot $\qquad$ .
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.

