|  |
| --- |
|  |

Lesson 19: 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 conveys information about the anticipated accuracy of the sample mean as an estimate of the population mean.

Lesson Notes

This is the second of two lessons building 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 for random samples from a population. They also explore how the simulated sampling distribution provides information about the anticipated estimation error when using a sample mean to estimate a population mean and how sample size affects the distribution of the sample mean.

Classwork

This lesson uses simulation to approximate the sampling distribution of the sample mean for random samples from a population, explores how the simulated sampling distribution provides insight into the anticipated estimation error when using a sample mean to estimate a population mean, and covers how sample size affects the distribution of the sample mean.

Exercises 1–6 (35 minutes): SAT scores

In this lesson, you may want to give students the population data and have them use technology to take random samples (without replacement) from the population. (You could copy and paste the table into a spreadsheet and send it to students.) A typical set of commands to generate a random sample might be: randsamp(SAT\_scores, ), where SAT\_scores is the name of the list containing the scores and is the sample size. The random sample should refresh by clicking on the command line or by using a command such as Control R.

Note that the sample answers for the simulated distributions typically display the means from about 50 random samples. If it is possible to generate many more samples quickly with technology, students should do so. The basic characteristics of a distribution of sample means (center, spread, mound shaped) do not change much as more samples are added to the first or so—which is suggested by contrasting Exercise 3 parts (a) and (b)—and students should experience this themselves by generating their own distributions with many samples.

Part (a) of both Exercises 2 and 3 are important to discuss as they highlight the difference in the distribution of the values in the sample (the scores) and the distribution of the sample means (the mean of the scores) in a sample.

Have students share the simulated sampling distributions they generate for Exercise 3. You might use screen capture/sharing software or have students walk around the room with post-it notes looking at each student’s handheld or computer screen and recording what they see about the distributions. Recognizing that all of the simulated distributions have essentially the same characteristics is a key factor in understanding why it is possible to make general statements about how random samples behave in relation to the population.

**MP.2**

Exercises 1–6: SAT scores

1. SAT test scores vary a lot. The table displays the scores for students in one New York school district for a given year.

Table 1: SAT scores for district students

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

* 1. Looking at the table above, how would you describe the population of SAT scores?

Sample response: It is hard to tell. I see some numbers as low as and others in the s. You cannot tell much from just looking at the individual numbers.

* 1. Jason used technology to draw a random sample of size from all of the scores and found a sample mean of . What does this value represent in terms of the graph below?

Random sample from District SAT scores

Sample response: This represents the average SAT score for the sample and indicates where the scores in the dot plot are centered. If you computed the mean of the values for the SAT scores in the sample (i.e., one student had an SAT score about , two had scores a bit over , one student scored about , and so on), you would find the mean SAT scores for those students to be .

1. If you were to take many different random samples of from this population, describe what you think the sampling distribution of these sample means would look like.

Sample response: Maybe centered on a value close to . I am not sure of the spread though, possibly from to like the sample in the example.

1. Everyone in Jason’s class drew several random samples of size and found the mean SAT score. The plot below displays the distribution of the mean SAT scores for their samples.

Random sample from District SAT scores

* 1. How does the simulated sampling distribution compare to your conjecture in Exercise 2? Explain any differences.

Possible answer: I estimated the mean SAT score to be a bit higher than it seems to be from the simulated distribution, but my estimate was not that far off. I thought the spread would range from to like in Jason’s sample, but that was not a good estimate. I was thinking of the individual SAT scores in the sample and mixing them up with the means of samples of scores, which is why my estimate was off.

* 1. Use technology to generate many more samples of size , and plot the means of those samples. Describe the shape of the simulated distribution of sample mean SAT scores.

**MP.5**

Answers will vary. The following displays an example of a simulated distribution of sample means from random samples from District SAT scores

The distribution is normal. The mean of the sample mean SAT scores appears to be around , and the sample mean scores range from a little over to about .



* 1. How did the simulated distribution using more samples compare to the one you generated in Exercise 3?

Sample response: The maxima and minima were nearly the same in both of the distributions, to and a little over to . The mean SAT scores of the simulated distributions of sample means in both seemed to be about .

* 1. What are the mean and standard deviation of the simulated distribution of the sample mean SAT scores you found in part (b)? (Use technology and your simulated distribution of the sample means to find the values.)

Sample response: The mean SAT score of the simulated distribution of sample means is . The standard deviation is .

* 1. Write a sentence describing the distribution of sample means that uses the mean and standard deviation you calculated in part (d).

Sample response: Almost all of the SAT scores are within two standard deviations from the mean, from to .

1. Reflect on some of the simulated sampling distributions you have considered in previous lessons.
	1. Make a conjecture about how you think the size of the sample might affect the distribution of the sample SAT means.

Sample response: the larger the sample size, the smaller the spread of the distribution of sample means.

* 1. To test the conjecture, investigate the following sample sizes: and as well as the simulated distribution of sample means from Exercise 3. Divide the sample sizes among your group members, and use technology to simulate sampling distributions of mean SAT scores for samples of the different sizes. Find the mean and standard deviation of each simulated sampling distribution.

**MP.5**

The following represent simulated distributions of sample means of District SAT scores for different size random samples:

|  |  |
| --- | --- |
| mean ; standard deviation is  | mean ; standard deviation is  |

|  |  |
| --- | --- |
| mean ; standard deviation is  | mean ; standard deviation is  |

|  |  |
| --- | --- |
| mean ; standard deviation is  |   |

* 1. How does the sample size seem to affect the simulated distributions of the sample SAT mean scores? Include the simulated distribution from part (b) of Exercise 3 in your response. Why do you think this is true?

Sample response: As the sample size increases, the spread decreases. The standard deviation went from for a sample of size to about for a sample of size . The means of the sampling distributions of mean SAT scores varied from for the distribution for samples of size to for samples of size . I would expect that a bigger sample would be more likely to look a lot like the population, and so bigger samples wouldn’t tend to be as different from one another as smaller samples. Because of this, the sample means wouldn’t differ as much from sample to sample for bigger samples.

* 1. For each of the sample sizes, consider how the standard deviation seems to be related to the range of the sample means in the simulated distributions of the sample SAT means you found in Exercise 4.

Sample response: In each case, nearly all of the sample means are within two standard deviations of the mean or are a normal distribution.

* 1. How do your answers to part (a) compare to the answers from other groups?

Possible answer: Everyone had simulated sampling distributions that looked fairly alike and were centered in about the same place. The sample means for each sampling distribution were typically within two standard deviations of the mean of the simulated sampling distributions. The distributions were normal.

* 1. Make a graph of the distribution of the population consisting of the SAT scores for all of the students.

Possible response below of a distribution of the SAT scores for district students

* 1. Find the mean of the distribution of SAT scores. How does it compare to the mean of the sampling distributions you have been simulating?

Possible answer: The mean SAT score for the students in the district is or . This is close to the means of the sampling distributions, even for fairly small samples.

Closing (5minutes)

* Why do we call the sampling distributions we generated “*simulated* sampling distribution of sample means” rather than the “sampling distribution of sample means”?
	+ *The sampling distribution of sample means is the distribution of all the possible sample means for a sample of a given size (i.e., every possible combination of the population values for that size). The simulated sampling distribution is a subset of the actual sampling distribution that, because it is random, approximates the actual sampling distribution.*
* If you had a simulated distribution of the mean SAT scores for random samples of size , how do you think the distribution would compare to the distribution you found for samples of size ?
	+ *Sample response: The mean would be somewhere around , and the standard deviation would be smaller, so the values (the sample means) would be closer together.*
* Ask students to summarize the main ideas of the lesson in writing or with a neighbor. Use this opportunity to informally assess comprehension of the lesson. The Lesson Summary below offers some important ideas that should be included.

Lesson Summary

For a given sample you can find the sample mean.

* There is variability in the sample mean. The value of the sample mean varies from one random sample to another.
* A graph of the distribution of sample means from many different random samples is a simulated sampling distribution.
* Sample means from random samples tend to cluster around the value of the population mean. That is, the simulated sampling distribution of the sample mean will be centered close to the value of the population mean.
* The variability in the sample mean decreases as the sample size increases.
* Most sample means are within two standard deviations of the mean of the simulated sampling distribution.

Exit Ticket (5 minutes)

Name Date

Lesson 19: Sampling Variability in the Sample Mean

Exit Ticket

1. Describe the difference between a population distribution, a sample distribution, and a simulated sampling distribution and make clear how they are different.
2. Use the standard deviation and mean of the sampling distribution to describe an interval that includes most of the sample means.

Exit Ticket Sample Solutions

1. Describe the difference between a population distribution, a sample distribution, and a simulated sampling distribution and make clear how they are different.

Possible answer: The distribution of the elements in a population (the SAT scores for students in a district) is a population distribution; the distribution of the elements in a random sample from that population (a subset of a given size chosen at random from the SAT scores) is a sample distribution; a simulated distribution of sample means for many random samples of a given size chosen from the population (the means of different random samples of the same size of students’ SAT scores) is a simulated sampling distribution.

Some students might also suggest that the meaning of sampling distribution of all samples is the samples of a given size selected from a population. This would be the distribution of the means of every possible sample that might be chosen.

1. Use the standard deviation and mean of the sampling distribution to describe an interval that includes most of the sample means.

Sample response: Typically, most of the means of the different random samples of the same size chosen from a population will be within two standard deviations of the mean or the Mean standard deviations.

Problem Set Sample Solutions

1. Which of the following will have the smallest standard deviation? Explain your reasoning.

A sampling distribution of sample means for samples of size:

a. b. c.

Possible answer: The largest sample size, , will have the smallest standard deviation because as the sample size increases, the variability in the sample mean decreases.

1. In light of the distributions of sample means you have investigated in the lesson, comment on the statements below for random samples of size chosen from the District SAT scores.
	1. Josh claimed he took a random sample of size and had a sample mean score of .

Possible answer: A mean score of seems very unlikely. None of the samples we have investigated had a sample mean score that low.

* 1. Sarfina stated she took a random sample of size and had a sample mean of .

Possible answer: This seems plausible for the simulated distributions of sample mean scores; was high but still some of the random samples had mean scores greater than .

* 1. Ana announced that it would be pretty rare for the mean SAT score in a random sample to be more than three standard deviations from the mean SAT score of .

Possible answer: Given that the sample means in nearly all of the simulated distributions of the sample means were usually within two standard deviations from the mean, Ana is correct. It could happen, but it would not be usual.

1. Refer to your answers for Exercise 4, and then comment on each of the following:
	1. A random sample of size produced a mean SAT score of .

Sample response: A mean score of was less than any of the sample means in the simulated sampling distribution of sample means for samples of size , so this seems unlikely.

* 1. A random sample of size produced a mean SAT score of .

Sample response: A mean score of was within two standard deviations of the mean for random samples of size , so it could have come from one of the samples.

* 1. For what sample sizes was a sample mean SAT score of plausible? Explain your thinking.

Sample response: A sample mean of occurred in the simulated sampling distributions for samples of size but not at all in the simulated distributions for samples of size and. So it seems like was a plausible outcome for samples of size and .

1. Explain the difference between the sample mean and the mean of the sampling distribution.

Possible answer: Each sample of SAT scores had a mean SAT score, which is the sample mean. Then all of those sample means formed a distribution of sample means, and we found the mean of that set, the mean of the sampling distribution of the sample mean – the mean of the means of the different samples.