Math Objectives

- Students will describe the meaning of bias in estimation.
- Students will recognize that the mean of a random sample is an unbiased estimator for the population mean.
- Students will recognize that estimates of the population mean using expert judgment can be biased.
- Students will recognize that the mean of a sample obtained using non-random sampling methods can be a biased estimator of the population mean.
- Look for and make use of structure (CCSS Mathematical Practices).

Vocabulary

- bias
- convenience sample
- mean
- random sample

About the Lesson

- This lesson involves approximating the mean of a variable.
- As a result, students will:
  - View a collection of rectangles and create an informal estimate of their mean area.
  - Use the mean of a convenience sample as an estimate of the mean area of the entire collection.
  - Use the mean of a random sample as an estimate of the mean area of the entire collection.
  - Compare the collective results of the three estimation methods across the class to deduce that random samples provide estimates that, on average, are closer to the true population mean than the other two methods.

TI-Nspire™ Navigator™ System

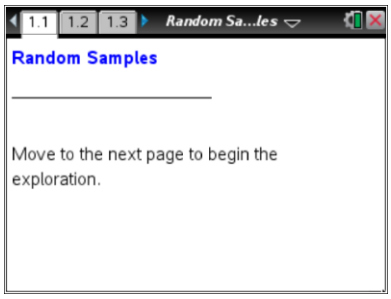
- Use Class Capture to compare students' means for each method (educated guess, judgment sample, and simple random sample).

Prerequisite Knowledge

- dot plots
- mean

Related Lessons

- Stratified Sampling



TI-Nspire™ Technology Skills:

- Download a TI-Nspire document
- Open a document
- Move between pages

Tech Tips:

- Make sure the font size on your TI-Nspire handhelds is set to Medium.

Lesson Files:

Student Activity

- Random_Samples_Student.pdf
- Random_Samples_Student.doc
- Random_Samples_Appendix.doc

TI-Nspire document

- Random_Samples.tns

Visit www.mathnspired.com for lesson updates and tech tip videos.

Discussion Points and Possible Answers

Teacher Tip: This activity requires collecting data from the entire class and adding it to the .tns file. After the completion of Question 4, select the aggregation approach that you prefer, depending on your class:
(i) After Question 4, have students call out data for you to enter into the file, and then send the completed data file to all students, or
(ii) As students call out data, have each student enter the data into the file for themselves.

Teacher Tip: Hand out the Random Samples Appendix to each student face down. Remind students not to look at it until instructed to do so.

You will be given a handout, face down, that contains the images of exactly 100 rectangles of various sizes.

Do not look at the handout until you are instructed to do so.

Your task throughout this activity will be to obtain an estimate for the mean of the areas of all 100 rectangles.

One method for estimating the mean area is simply to look at the whole set of rectangles and make an educated guess based on what you see.

1. When you are told, turn the handout face up. You will have 10 seconds to examine the rectangles and arrive at a reasonable estimate for the mean of the areas of the 100 rectangles. When time is called at the end of 10 seconds, turn your handout face down again. Write your estimate below.

   Sample Answers: Typical responses will probably be 10 or higher, but will vary from student to student.

Another way to estimate the mean area of all the rectangles might be to take a sample of the actual rectangles on the sheet and find the mean area for the sample.

Tech Tip: In Question 2, if students are uncomfortable computing the mean of their sample by hand, you might want to allow them to go to the **Home Screen > Scratchpad > Calculator** application on their handheld to complete their calculations.

2. When you are told, turn the handout face up again, and visually select exactly five rectangles you think would be a representative sample. Note the ID number and the area of each rectangle you select. Record your data below. Find the mean area of your sample, keeping any digits past the decimal in your mean. Turn your handout face down again.

ID	Area	Mean Area of Sample
_____	_____	
_____	_____	_____
_____	_____	
_____	_____	
_____	_____	

Sample Answers: Typical responses might be around 10, but will vary from student to student.

> **Teacher Tip:** If you have a small class, or you just want a larger set of data for discussion later, you might want to repeat Question 2. It is important that the sample size remain 5 and that no "computations" be done that might influence student opinions of the rectangles they select in their samples.

Open the TI-Nspire™ document *Random_Samples.tns.*

Move to page 1.2.

Tip: This activity involves generating a number of random samples from a population. In order to avoid having your results be identical to those for another student in the room, it is necessary to "seed" the random number generator. Read the instructions on Page 1.2 for seeding your random number generator, and then carry out that seeding on Page 1.3.

The sample you selected in Question 2 was what the media usually call an "unscientific" sample. Statisticians would call it a **judgment sample**, since you just selected five rectangles you thought were typical. Another kind of sample that statisticians use is called a **simple random sample**.

> **Tech Tip:** In Question 3, if any students obtain one or more repeated ID numbers in their list of five, simply have them press enter again to generate another list.

3. You will use the Scratchpad on your handheld to select five random rectangle ID numbers, as follows. Open a Scratchpad Calculator page. Type the command **randint(0,99,5),** and press enter. You should see five ID numbers as output. Turn the handout face up again, locate the five rectangles whose ID numbers were generated, and record their IDs and areas below. Find the mean area of your sample, keeping any digits past the decimal in your mean.

ID	Area	Mean Area of Sample
_____	_____	
_____	_____	_____
_____	_____	
_____	_____	
_____	_____	

Sample Answers: Typical responses will usually be lower than before, but will vary from student to student.

You have now used three different methods to estimate the mean area of the set of rectangles. In statistics, a method is called **unbiased** if its long-term behavior, on average, is correct. More precisely, if the mean of the estimates produced by the method is actually equal to the value that the method is trying to estimate, then that method is unbiased.

In contrast, the mean of estimates from a **biased** estimation method would differ from the true value being estimated. Thus the estimates from a biased method will be systematically too high or too low. In general, unbiased estimators are usually more desirable.

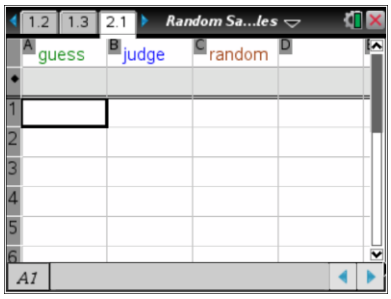
4. Which, if any, of the three estimation methods—educated guess, judgment sample, or random sample—do you think might be unbiased? Why?

Sample Answers: Random sample. The random process can't be thrown off by what is visible—what a person "sees"—so large and small rectangles have just as good a chance for being included.

One way to determine which of the three estimation methods you used to estimate the mean area of the rectangles is best is to compare their results from everyone in your class. Your teacher will help you collect the estimates for the three methods from your classmates.

Teacher Tip: Following completion of Question 4, use the method of your choice as outlined at the beginning to collect class data into Page 2.1 of the tns file and distribute the tns file to each student.

Move to page 2.1.

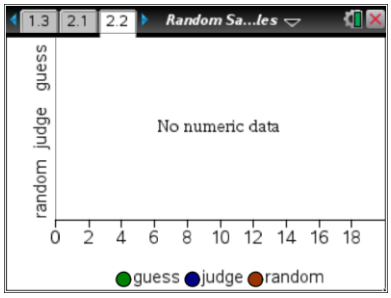


5. Examine your class's estimated mean areas from the three methods, compiled in the respective spreadsheet columns on Page 2.1. Based on looking at just the numbers, does it seem that any particular method is more consistent than the others?

   Sample Answers: Answers will vary depending on student impressions and the actual consistency within the class.

   Teacher Tip: Remind students that consistency is good, but if an estimation method is consistent but biased, then it is not particularly useful.

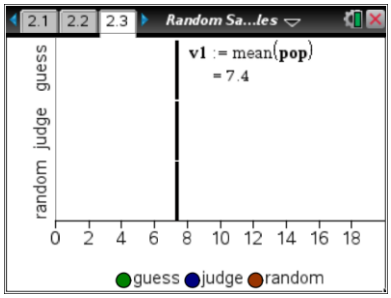
Move to page 2.2.



6. Page 2.2 displays three dot plots with your class's results from the three different methods of estimating the mean area of the rectangles. Compare the three distributions with respect to center, spread, and shape. Which estimation method seems best?

   Sample Answers: Answers will vary depending on the class. However, students should include some mention of center, spread, and shape in each description. Students should realize that it is probably impossible to tell which method is "best" yet since the actual mean area is still not known.

Move to page 2.3.



7. Page 2.3 shows the same dot plots you saw on Page 2.2, but now a vertical line has been added to the plot.
   a. Click on the vertical line. What information does it convey?

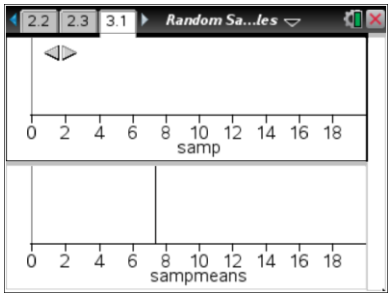
      Answer: It's the mean area that we have been trying to estimate.

   b. Based on this new information, which estimation method seems best? Explain.

Sample Answers: The estimates based on random samples come closest to being unbiased—the center of that distribution is pretty close to the plotted line. Both of the other methods seem to estimate too high. The variability seems to be greater in the educated guess and judgment samples.

Teacher Tip: It is worthwhile to discuss why the results of this comparison are as they are. In particular, point out that it is much easier to overlook the small (area = 1, say) rectangles than it is to overlook the larger ones. Thus, there is a strong visual bias toward larger areas in those methods that rely on the human eye.

Move to page 3.1.



8. Page 3.1 permits you to take many random samples of size five.
   a. Click the right arrow to select one random sample. Describe the contents of each panel on the screen, and interpret them in terms of the areas of the rectangles.

      Answer: The top panel shows the five areas in the sample, together with a vertical line at the mean area of that sample. The bottom panel shows a dot corresponding to the sample mean in the top panel, together with a vertical line at 7.4, the true mean of the population of areas.

   b. Continue to select more samples until a clear pattern emerges in the lower panel. Describe the important features of the distribution, and explain what it means with respect to estimation and bias.

      Sample Answers: The distribution becomes more unimodal and mound-shaped, centered pretty close to the vertical line at 7.4. The fact that the center appears to be the same as the population mean supports the idea that the mean of a random sample seems to be an unbiased estimator for the population mean.

Wrap Up

Upon completion of the lesson, the teacher should ensure that students are able to understand:

- "Unbiased" in statistics means that a particular process produces the "right answer" on average.
- "Biased" in statistics means that a particular process produces results that are systematically too high or too low, yielding a mean that is not correct.
- Estimating a population parameter using only personal judgment (an educated guess) is likely to be biased.
- Estimating a population parameter using a sample selected using only personal judgment is likely to be biased.
- Estimating a population parameter using a sample that is selected randomly is unbiased.

Assessment

1. The statistics classes in two schools, one urban and one rural, think that the average height of their respective students might be different and want to compare the heights of the male students in the two schools. Comment on their plans described below, giving one advantage and one flaw of each plan.

   Urban school: The students will collect information by observing students in the main hallway during passing time and choosing forty boys that seem to represent the span of heights, making sure to choose some short boys and some tall ones.

   Rural school: The students will collect information by randomly generating 40 student numbers on their calculators and obtaining the heights of students having those student ID numbers, being sure to look only at male students.

   Sample Answers: The urban school's plan is very quick and convenient, but its results can be influenced by the appearance of the boys in the halls. The rural school's plan will be time-consuming to carry out but is unbiased.

2. Sari was worried that the method used in the rural school might not give them a sample that had both tall and short boys. What would you say to her?

   Sample Answers: Her objection is valid, but if the sample size is reasonably large it is very unlikely that such a "one-sided" sample would occur.