Building Concepts: Why Random Sampling? TEACHER NOTES

Lesson Overview

In this TI-Nspire lesson students focus on what it means to take a "random" sample and how a random sample is an unbiased method of identifying a sample. Students eyeball a population, select what they think will be a random sample and calculate a sample statistic for their sample. They contrast the distribution of these "judgment" sample means with a distribution of sample means for samples when random numbers are used to select the elements in the sample.

Random sampling is a better method of collecting data than using an "expert" opinion.

Learning Goals

1. Understand that random sampling as a method of collecting information from a population is likely to produce a sample that is representative of the population;
2. recognize that a sample based on "expert" judgment is typically not random and often produces estimates for a population mean that are not representative of the population;
3. recognize that the means of random samples of the same size will cluster around the population mean.

Prerequisite Knowledge

Why Random Sampling? is the seventeenth lesson in a series of lessons that explore the concepts of statistics and probability. This lesson builds on the concepts of the previous lessons. Prior to working on this lesson students should have completed *Probability and Simulation* and *Law of Large Numbers*. Students should understand:

- how to use simulation to estimate probabilities;
- how to compare the experimental probabilities with the theoretical probability;
- how to identify the set of possible outcomes from a chance event.

Vocabulary

- **random sample:** sample in which every possible combination of sample size *n* from the population has the same chance of being selected.
- **sample mean:** the value obtained by dividing the sum of a set of quantities by the number of quantities in the set
- **population:** the entire group of objects or individuals considered for a survey

Lesson Pacing

This lesson should take 50–90 minutes to complete with students, though you may choose to extend, as needed.

Lesson Materials

- Compatible TI Technologies:

  TI-Nspire CX Handhelds, TI-Nspire Apps for iPad®, TI-Nspire Software

- Why Random Sampling_Student.pdf
- Why Random Sampling _Student.doc
- Why Random Sampling.tns
- Why Random Sampling_Teacher Notes
- To download the TI-Nspire activity (TNS file) and Student Activity sheet, go to http://education.ti.com/go/buildingconcepts.

Class Instruction Key

The following question types are included throughout the lesson to assist you in guiding students in their exploration of the concept:

Class Discussion: Use these questions to help students communicate their understanding of the lesson. Encourage students to refer to the TNS activity as they explain their reasoning. Have students listen to your instructions. Look for student answers to reflect an understanding of the concept. Listen for opportunities to address understanding or misconceptions in student answers.

Student Activity: Have students break into small groups and work together to find answers to the student activity questions. Observe students as they work and guide them in addressing the learning goals of each lesson. Have students record their answers on their student activity sheet. Once students have finished, have groups discuss and/or present their findings. The student activity sheet can also be completed as a larger group activity, depending on the technology available in the classroom.

Deeper Dive: These questions are provided for additional student practice and to facilitate a deeper understanding and exploration of the content. Encourage students to explain what they are doing and to share their reasoning.

Mathematical Background

A subset of a population is called a *sample* and collecting information from a sample is called *sampling*. Relating a sample to the population is valid only if the sample data are representative of that larger group. Random sampling is the key to collecting good data because a random sample is typically representative of the population from which it was drawn. Random sampling is intended to reduce the difference between a sample and population from which the sample was taken. When random selection is used, differences between samples will be due to chance. Understanding this chance variation is what leads to the predictability of results.

In statistics, a method is called *unbiased* if its long-term behavior, on average, produces a result that 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 in some regular way from the true value being estimated. Thus, the estimates from a biased method will be systematically too high or too low. Improper or biased sample selection tends to systematically favor certain outcomes and can produce misleading results and erroneous conclusions. Random sampling is a way to remove bias in sample selection, and tends to produce samples that are representative of the population, which is emphasized in this lesson.

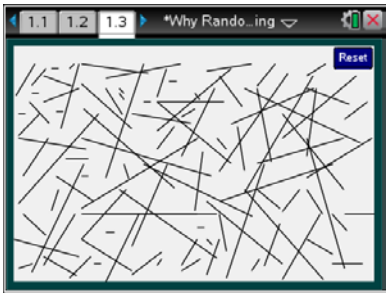
Note that all of the Class Discussion questions need to be completed by all of the students as the concept of a random sample is developed over the entire set of questions.

Part 1, Page 1.3

Focus: Students will look at a population, select a "representative" sample from that population, and calculate a sample statistic for their sample.

Page 1.3 contains 100 segments of various lengths measured in millimeters. Selecting the segment displays the length of the segment. Students can select up to five segments (i.e., a sample of five).

Reset resets the page to the original screen.



TI-Nspire Technology Tips

tab cycles through the segments beginning at random locations.

enter selects segment highlighted using tab and displays length.

ctrl del resets the page to the original screen.

Class Discussion

The following set of questions has students find a mean length for the segments on page 1.3 in three ways: making a guess at a typical length, selecting a sample by eyeballing what they think is typical, and using a sample that has been generated randomly.

Have students…	Look for/Listen for…
Look at the segments on page 1.3.	
• ***Find a segment you think is typical of the length of the segments on the page. (Just an educated guess) Select the segment and record its length in millimeters.***	Answers will vary.
• ***Identify what you think are the longest and shortest segments in the set of segments.***	Answer: The longest segments seem to be 18 mm (segments 40, 52, 13, for example). The shortest segments seem to be 1 mm (segments 81, 31, 5, for example).
Select five segments you think will be typical of the segments on the page.	
• ***Record the lengths, given in millimeters, of these segments.***	Answers will vary. One sample gave segment 38 with a length of 16 mm, segment 79 with a length of 5 mm; segment 81 with a length of 1 mm, segment 46 with a length of 9 mm, segment 53 with a length of 3 mm.

Class Discussion (continued)

- ***Compute the mean length of the segments in your sample and report it to your teacher. Record all of the class mean lengths on the dot plot on the bottom of page 3.4.***

  Answers will vary. The mean length of the sample above is $\frac{34}{5} = 6.8$ mm

> **Teacher Tip:** After students have completed the question above, collect the measures from the class's "judgment" samples. These measurements can be entered onto the bottom graph on page 3.4. The points you enter can be moved to the student's reported mean using the arrow keys or dragging the point on the axis. You may want to enter the values as students report them, have a student enter all the data into the graph so that everyone can see the resulting plot, or have each student enter the class results so they all have the same plot (this can be more time consuming). Discuss the distribution that results, but do not do anything else on that page as you will return to it at the end of the lesson.

Part 1, Page 2.2

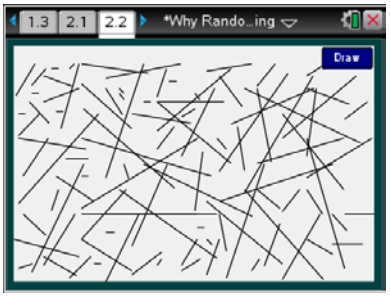
Focus: Students will find the mean of the sample of their population.

On page 2.2, random samples of 5 segments may be drawn.

Draw displays a random sample of five numbers from the numbers 1 to 100.

Measure displays the length in millimeters of each segment associated with the random numbers determined by the sample.

Reset returns to the original screen.



TI-Nspire Technology Tips

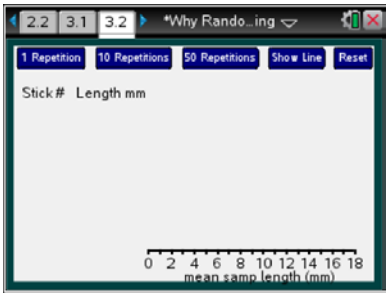
enter cycles through **Draw**, **Measure**, and **Reset**.

Class Discussion

Have students…	**Look for/Listen for…**
***On page 2.2, select* Draw**.	
• ***What segment numbers were in your sample?***	Answers may vary. 3, 47, 35, 54, 48
• ***Select* Measure *and determine the mean length of the sample of segments.***	Answers will vary. $4+10+4+3+6 = \frac{27}{5} = 5.4$ mm
• ***Compare your mean to the mean of your judgment sample.***	Answers may vary. The judgment sample for the examples above had slightly higher mean (6.8 mm as compared to 5.4 mm).

Part 1, Page 3.2

Focus: Students will continue their investigations on the mean of their sample population.



On page 3.2 students can select a random sample of segments, see the lengths, the mean and a dot plot of the mean. Selecting any dot will display the random sample associated with the sample mean represented by that dot.

1 Repetition generates a single sample of 5 segments and displays the lengths, the mean, and a dot on the number line for the mean.

10 Repetitions generates enough samples of 5 segments to get to the next multiple of 10 and displays a dot plot of the mean lengths for each of the samples.

50 Repetitions generates enough samples of 5 segments to get to the next multiple of 50 and displays a dot plot of the mean lengths for each of the 50 samples.

Show Line displays the mean length of the entire set of segments.

TI-Nspire Technology Tips

tab cycles through the samples or buttons for sampling and Show Line.

Up/Down arrow keys direct tab.

enter activates the highlighted button.

Student Activity Questions—Activity 1

1. **A *random sample* of a certain size is a sample in which every set in the population of that size has the same chance of being selected. On page 3.2, *1 Repetition* will generate random samples of 5 segments.**

   a. **Select *1 Repetition* and describe the random sample of segments.**

   Answers will vary. One sample produced segments of lengths 17, 5, 16, 12 and 4 for a mean length of 10.8 mm.

   b. **Select *10 Repetitions*. Highlight one of the points in the dot plot. Explain what that point represents.**

   Answers will vary. A mean of 3.2 mm came from a sample of lengths 4, 3, 4, 1, 4

   c. **Describe the dot plot and explain what it represents.**

   Answers will vary. The plot has ten dots spread from 3.2 mm to 11.6 mm. Each dot represents the mean length in millimeters of one of the samples of five segments that was randomly drawn from all 100 segments.

Student Activity Questions—Activity 1 (continued)

2. Select *10 Repetitions*, look at the dot plot and select another *10 repetitions*.

a. Continue until you have about 100 samples. Describe how the dot plot changed as the number of samples increased from 10 to 100.

Answer: As the number of repetitions increases, the dot plot becomes more and more mound shaped and symmetric around 7.5 mm. The spread (range) of the sample means is 11.8, from 2.6 mm to 14.4 mm.

b. Based on the distribution of sample mean lengths you generated, make a conjecture about what you think the mean length of all of the segments will be. Explain your reasoning.

Answers will vary. The mean length of the all of the segments will probably be the center of the distribution of the sample means, around 7.5 mm or so.

c. Select *Show Line*. How close was your conjecture and how does it fit into the distribution of the sample means?

Answers will vary. The estimate of the mean length for all the segments was around 7.5 mm, which is close to the actual mean length of 7.2 mm. The distribution piles up around that value.

d. Select *50 Repetitions* until you have a distribution of 500 sample mean lengths. How does the distribution change?

Answer: The range is 12.8, from 1.6 to 14.4 mm. The distribution of the sample mean lengths is getting more symmetric around the mean, 7.2 mm.

3. Reset and generate another distribution of sample mean lengths. How does this distribution of sample means compare to the distribution from 2d?

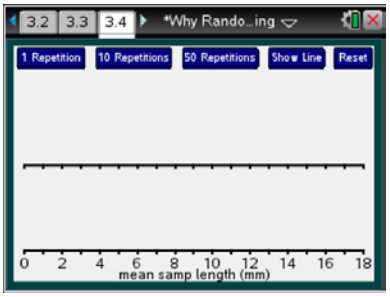
Answers will vary. For example, the distribution is almost the same. This one goes from 1.6 mm to 14.6 mm and is not quite as symmetric as the first one.

Teacher Tip: Be sure that students recognize that, in statistics a method produces a representative sample if, over the long run (i.e., many, many samples), on average, the estimate is correct. More precisely, if the mean of the estimates produced by the method approaches the population value that the method is trying to estimate, then that method produces samples that are representative of the population. In contrast, the mean of estimates from a method that is not representative will be systematically too high or too low.

Part 2, Page 3.4

Focus: Students will investigate ways to find a sample that is most representative of the population.

On page 3.4, students can enter data on the bottom axis by selecting a number on the axis. The points can be moved using the arrow keys or by dragging the dot. The top axis displays the distribution of mean lengths from random samples of size five that was generated on page 3.2. A new distribution of the mean lengths from the random samples can be produced on page 3.4, as well.



TI-Nspire Technology Tips

tab cycles through the student data or buttons for sampling and show line.

Up/Down arrow keys direct tab

enter activates highlighted button or enters a new student data point at 9.

Student Activity Questions—Activity 2

1. a. **Which, if any, of the three estimation methods—educated guess, judgment sample, or random sample—do you think is more likely to give a sample that is most representative of the population? Why?**

   Answers will vary. In a guess and judgment sample, you are deciding what seems typical yourself and in a random sample, you are not entering into the decision at all. So random sampling seems most likely to produce a sample that is representative of the population.

   b. **Reset page 3.2. Generate a distribution of 30 sample means based on 30 random samples. Return to page 3.4. The means calculated from the judgment samples you made as a class are graphed on the bottom axis. The top axis shows the distribution of the 30 mean lengths from the random samples you generated on page 3.2. How do the distributions compare?**

   Answers will vary. The distribution of the mean lengths from the judgment samples has a wider spread, and the center is usually greater than the center of the distribution of the mean lengths calculated from the random samples.

   c. **Which sampling method seems to give results that are more representative of the population? Explain your reasoning.**

   Answers will vary. The mean lengths from the random samples should be most representative of the population since individual judgment is removed from the decision.

Student Activity Questions—Activity 2 (continued)

> **Teacher Tip:** It is worthwhile to discuss that in the judgment sample, it is much easier to overlook the small segments (length 1 for example) than it is to overlook the larger segments. There is a strong visual bias toward larger segments in methods that rely on the human eye.

2. **Saundra claimed that the mean of a random sample seems to be a good estimator for a population mean. Do you agree or disagree with Saundra? Use the TNS activity to support your reasoning.**

   Answers will vary. When you generate many random samples, the distribution of the mean lengths becomes more unimodal and mound-shaped, centered pretty close to the vertical line at 7.2. 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 a pretty fair estimator for the population mean.

3. **Which of the following methods for collecting information about the average grade point of seventh graders in a school do you think are biased? Explain your reasoning.**

   a. **Collect information by observing students in the main hallway during passing time and choose forty students that seem to represent different kinds of students, making sure to choose some boys and some girls.**

      Answers will vary.

   b. **Number the second hour seventh grade classes in school, then randomly choose two of the classes and find the average grade points for the students in those classes.**

      Answers will vary.

Deeper Dive — Pages 3.2 and 3.4

- ***Cal claims he can estimate the average area of the counties or regions in a state by looking at a map where the different regions are colored in different colors and selecting a judgmental sample of the regions by sight. What would you say to Cal?***

  Answers will vary. The method is likely to give a sample that is not representative of the population because some colors will attract your eyes more than others and so those colors are more likely to be chosen.

Deeper Dive — Pages 3.2 and 3.4 (continued)

- ***Timon worried that one random sample might not give a very good estimate for a population mean. What would you say to Timon?***

  Answers will vary. In one repetition, the random sample gave a mean of 6.2 mm. In another random sample, the mean length was 9 mm. Both of these were off, but in the long run, the estimates from the random samples will be better than our estimates based on eyeballing the population and making our own judgments. When the class did the judgment samples, one sample mean length was 4.5 mm and another was 10.2 mm. Both of these were farther off than the estimates of the mean length using the random samples.

- ***Write a summary of the activity, describing what you did and what you learned. What was the biggest idea you took away about sampling?***

  Answers will vary.

Sample Assessment Items

After completing the lesson, students should be able to answer the following types of questions. If students understand the concepts involved in the lesson, they should be able to answer the following questions without using the TNS activity.

1. Which of the following methods is likely to produce a sample that is representative of the students in your school?

   a. Choosing the students in your first hour class.

   b. Have the principal select students she feels would be typical of the students in the school.

   c. Choosing the students in the hallway before your first hour class.

   d. Putting each student's ID number in a bag, shaking the bag to mix the numbers and selecting 30 numbers from the bag.

   Answer: d. Putting each student's ID number in a bag, shaking the bag to mix the numbers and selecting 30 numbers from the bag.

2. Which of the following is a true statement about a random sample of students from your class. A random sample is

   a. choosing whoever comes to mind as you think about students in the class.

   b. choosing two boys and two girls, one of each who gets good grades and one of each who gets lower grades.

   c. a sample where every student in the class has the same chance to be selected

   d. a sample selected by the teacher based on student grades

   Answer: c. a sample where every student in the class has the same chance to be selected

3. The mean hand span of seventh graders is 18.2 centimeters. Which of the following would be surprising?

   a. The mean hand span from a random sample of 10 seventh grade students was 27 centimeters.

   b. The mean hand span from a random sample of 10 seventh grade students was 19 centimeters.

   c. The mean hand span from a sample of 10 seventh grade students was 27 centimeters.

   d. One sample of 10 seventh grade students had a mean hand span of 18.6 centimeters, and a second sample of seventh grade students had a mean hand span of 17.9 centimeters.

   Answer: a. The mean hand span from a random sample of 10 seventh grade students was 27 centimeters.

Student Activity Solutions

In these activities you will use random samples to find the sample mean of a population. After completing the activities, discuss and/or present your findings to the rest of the class.

Activity 1 [Page 3.2]

1. A *random sample* of a certain size is a sample in which every set in the population of that size has the same chance of being selected. On page 3.2, **1 Repetition** will generate random samples of 5 segments.

   a. Select **1 Repetition** and describe the random sample of segments.

   Answers will vary. One sample produced segments of lengths 17, 5, 16, 12 and 4 for a mean length of 10.8 mm.

   b. Select **10 Repetitions**. Highlight one of the points in the dot plot. Explain what that point represents.

   Answers will vary. A mean of 3.2 mm came from a sample of lengths 4, 3, 4, 1, 4

   c. Describe the dot plot and explain what it represents.

   Answers will vary. The plot has ten dots spread from 3.2 mm to 11.6 mm. Each dot represents the mean length in mm of one of the samples of five segments that was randomly drawn from all 100 segments.

2. Select **10 Repetitions**, look at the dot plot and select another **10 repetitions**.

   a. Continue until you have about 100 samples. Describe how the dot plot changed as the number of samples increased from 10 to 100.

   Answer: As the number of repetitions increases, the dot plot becomes more and more mound shaped and symmetric around 7.5 mm. The spread (range) of the sample means is 11.8, from 2.6 mm to 14.4 mm.

   b. Based on the distribution of sample mean lengths you generated, make a conjecture about what you think the mean length of all of the segments will be. Explain your reasoning.

   Answers will vary. The mean length of the all of the segments will probably be the center of the distribution of the sample means, around 7.5 mm or so.

   c. Select Show Line. How close was your conjecture and how does it fit into the distribution of the sample means?

   Answers will vary: The estimate of the mean length for all the segments was around 7.5 mm, which is close to the actual mean length of 7.2 mm. The distribution piles up around that value.

   d. Select **50 Repetitions** until you have a distribution of 500 sample mean lengths. How does the distribution change?

   Answer: The range is 12.8, from 1.6 to 14.4 mm. The distribution of the sample mean lengths is getting more symmetric around the mean, 7.2 mm.

3. Reset and generate another distribution of sample mean lengths. How does this distribution of sample means compare to the distribution from 2d?

Answers will vary: For example, the distribution is almost the same. This one goes from 1.6 to 14.6 and is not quite as symmetric as the first one.

Activity 2 [Page 3.4]

1. a. Which, if any, of the three estimation methods—educated guess, judgment sample, or random sample—do you think is more likely to give a sample that is most representative of the population? Why?

Answers will vary. In a guess, you are just looking and kind of sensing what seems typical. In a judgment sample, you are deciding what seems typical yourself and in a random sample, you are not entering into the decision at all- the random sample is done for you. So random sampling seems most likely to produce a sample that is representative of the population.

b. Reset page 3.2. Generate a distribution of 30 sample means based on 30 random samples. Return to page 3.4. The means calculated from the judgment samples you made as a class are graphed on the bottom axis. The top axis shows the distribution of the 30 mean lengths from the random samples you generated on page 3.2. How do the distributions compare?

Answers will vary. The distribution of the mean lengths from the judgment samples has a wider spread, and the center is greater than the center of the distribution of the mean lengths calculated from the random samples.

c. Which sampling method seems to give results that are more representative of the population? Explain your reasoning.

Answers will vary. The mean lengths from the judgment samples are pretty consistently greater than the mean lengths from the random samples, which is what happens when the method of collecting data does not give a sample that is representative of the population.

2. Saundra claimed that the mean of a random sample seems to be a good estimator for a population mean. Do you agree or disagree with Saundra? Use the TNS activity to support your reasoning.

Answers will vary. When you generate many random samples, the distribution of the mean lengths becomes more unimodal and mound-shaped, centered pretty close to the vertical line at 7.2. 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 a pretty fair estimator for the population mean.

3. Which of the following methods for collecting information about the average grade point of seventh graders in a school do you think are biased? Explain your reasoning.

   a. Collect information by observing students in the main hallway during passing time and choose forty students that seem to represent different kinds of students, making sure to choose some boys and some girls.

      Answers will vary.

   b. Number the second hour seventh grade classes in school, then randomly choose two of the classes and find the average grade points for the students in those classes.

      Answers will vary.