



Lesson 14: Selecting a Sample

Student Outcomes

- Students understand that how a sample is selected is important if the goal is to generalize from the sample to a larger population.
- Students understand that random selection from a population tends to produce samples that are representative of the population.

Lesson Notes

It is important for students to understand how samples are obtained from a population. Students may think that getting a sample is relatively easy. For example, ask students to consider the statistical question, “How many hours of network television does the typical student in our school watch each week?” Also, ask students how they would go about getting responses to that question that would be representative of the school population. Often students will indicate that they will ask their friends or the students in their math class. This type of sample is referred to as a “convenience sample.” Ask the class to think about why convenience samples generally do not result in responses that represent the population. Students should begin to understand that there is something unfair about a convenience sample, and therefore, might not be the best sample to represent the entire population.

This lesson sets the stage for selecting and summarizing random samples. Explain that in the above example, a sample is obtained where not all students are equally likely to be selected into the sample. A sampling method that avoids this is called a random sampling. In this lesson, students will observe how a random sample is representative of the population.

Read through the directions for Exercises 3–11. Exercises 3–11 involve analyzing the words of the poem *Casey at the Bat*. Prepare the materials needed for students to complete the exercises as indicated in the directions. The samples obtained in the exercises are an important part in understanding why random samples are used to learn about population characteristics.

Classwork

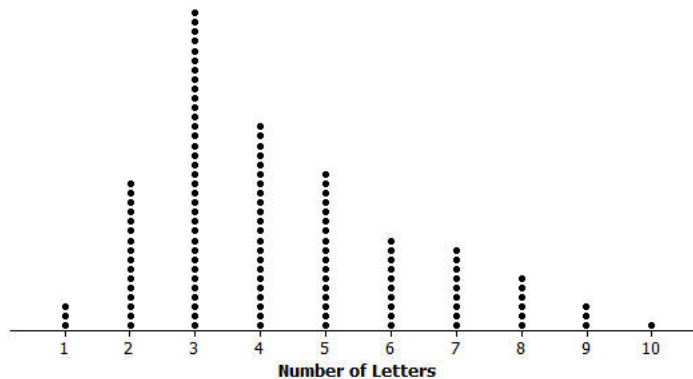
As you learned in Lesson 13, sampling is a central concept in statistics. Examining every element in a population is usually impossible. So research and articles in the media typically refer to a “sample” from a population. In this lesson, you will begin to think about how to choose a sample.

Exercises 1–2 (10 minutes): What is Random?

Students often do not understand what the word “random” means in statistics. Before random samples are developed, explore the ideas of random behavior that are developed in the first two exercises. These exercises indicate that random behavior may not necessarily look like what students expect. Similarly, random samples are generally representative of the population despite not always looking as expected.

(Note: Teachers may also consider organizing the class responses to Exercises 4 and 6 in a back-to-back stem and leaf plot or on two parallel number lines with the same scale to make the difference between a self-selected sample and a random sample more visible. These visuals may also show that the sample distribution of the mean number of letters in the words in the random samples will tend to center around the population mean, 4.2 letters. If you collect the set of all of the means from the different random samples in the class, the distribution of the sample means will usually center around 4.2 letters.) The set of numbers selected by students just picking “representative” words will typically be larger than the population mean because people usually do not recognize how many two and three letter words are in the population. The graph below shows the population distribution of the length of the words in the poem.

Dot Plot of Number of Letters in Words from Casey at the Bat



Each symbol represents up to 5 observations.

Exercises 3–11: Length of Words in the Poem *Casey at the Bat*

3. Suppose you wanted to learn about the lengths of the words in the poem *Casey at the Bat*. You plan to select a sample of eight words from the poem and use these words to answer the following statistical question: On average, how long is a word in the poem? What is the population of interest here?

The population of interest is all of the words in the poem.

4. Look at the poem, *Casey at the Bat* by Ernest Thayer, and select eight words you think are representative of words in the poem. Record the number of letters in each word you selected. Find the mean number of letters in the words you chose.

Answers will vary. Sample response: The words “their, while, thousand, ball, bat, strike, muscles, grow” would have a mean of 5.25 letters.

5. A random sample is a sample in which every possible sample of the same size has an equal chance of being chosen. Do you think the set of words you wrote down was random? Why or why not?

Answers will vary. Sample response: I thought it was random because I tried to use some little words and some long ones.



6. Working with a partner, follow your teacher’s instruction for randomly choosing eight words. Begin with the title of the poem and count a hyphenated word as one word.

a. Record the eight words you randomly selected and find the mean number of letters in those words.

Sample response: We drew group 1, 12 (nine); group 1, 18 (four); group 5, 7 (seemed); group 3, 17 (in); group 23,6 (fraud); group 27, 11 (is); group 27,10 (air); group 16, 16 (close). The mean length of the words was 3.875.

b. Compare the mean of your random sample to the mean you found in Exercise 4. Explain how you found the mean for each sample.

Sample response: The mean of the sample from Exercise 4 is based on the length of eight words I selected. The mean of the sample in this exercise is the mean of eight words randomly selected from the poem. Anticipate that for most students, the mean from the random sample will be greater than the mean for the self-selected sample.

7. As a class, compare the means from Exercise 4 and the means from Exercise 6. Your teacher will provide a chart to compare the means. Record your mean from Exercise 4 and your mean for Exercise 6 on this chart.

Organize the responses in a table posted in the front of class. Have students add their means to the poster. Consider the following example:

<i>Mean from the sample in Exercise 4</i>	<i>Mean from the random sample in Exercise 6</i>
6.3	4.5
4.9	4.1
6.3	4.4

8. Do you think means from Exercise 4 or the means from Exercise 6 are more representative of the mean of all of the words in the poem? Explain your choice.

Sample response: The means in the random sample seem to be similar. As a result, I think the means from the random sample are more representative of the words in the poem.

9. The actual mean of the words in the poem *Casey at the Bat* is 4.2 letters. Based on the fact that the population mean is 4.2 letters, are the means from Exercise 4 or means from Exercise 6 a better representation the mean of the population. Explain your answer.

Sample response: The means from the random samples are similar and are closer to the mean of 4.2. Also, the means from Exercise 4 are generally larger than the mean of the population.

10. How did population mean of 4.2 letters compare to the mean of your random sample from Exercise 6 and to the mean you found in Exercise 4?

Sample response: The mean number of letters in all of the words in the poem is 4.2, which is about four letters per word and the mean of my random sample, 3.875, was also about four letters per word. The mean of my sample in Exercise 6 was about five letters per word.

11. Summarize how you would estimate the mean number of letters in the words of another poem based on what you learned in the above exercises.

Sample response: Students should summarize a process similar to what they did in this lesson. They may simply indicate that they would number each word in the poem. They would make slips of paper from 1 to the number of words in the poem, place the slips of paper in a bag or jar, and select a sample of eight or more slips of paper. Students would then record the number of letters in the words identified by the slips of paper. As in the exercises, the mean of the sample would be used to estimate the mean of all of the words in the poem.

Closing (5 minutes)

Discuss the following items with students. Allow a few student responses for each item.

- Ask students whether or not the sample of words based on picking numbers from the bag for the poem *Casey at the Bat* was a random sample.
 - *Yes, the sample of words was a random sample because each group had $\frac{1}{29}$ chance of being selected and each word in each group had $\frac{1}{20}$ chance of being selected, so every word had the same chance of being selected.*
- Ask students to identify the population, the sample, and the sample statistic involved in the exercises.
 - *Sample response: The population was all of the words in the poem. The sample was the set of eight words we randomly chose. The sample statistic was the mean number of letters in the eight words.*
- Ask students to consider the following situation. You are interested in whether students *at your school* enjoyed the poem *Casey at the Bat*? A survey asking students if they liked the poem was given to a sample consisting of the players on the school's baseball team. Is this sample a random sample? Do you think this sample is representative of the population? Explain.
 - *Sample response: The data from the baseball team would represent a sample. The proportion of students who indicated they like the poem would be a sample statistic. However, selecting the baseball team to complete the survey did not give an equal chance for all students in our school to complete the survey. It is not a random sample of the school. I would not trust the results as there is a chance baseball players would tend to have a different answer to this question than students who do not play baseball.*

Lesson Summary

When choosing a sample, you want the sample to be representative of a population. When you try to select a sample just by yourself, you do not usually do very well, like the words you chose from the poem to find the mean number of letters. One way to help ensure that a sample is representative of the population is to take a random sample, a sample in which every element of the population has an equal chance of being selected. You can take a random sample from a population by numbering the elements in the population, putting the numbers in a bag and shaking the bag to mix the numbers. Then draw numbers out of a bag and use the elements that correspond to the numbers you draw in your sample, as you did to get a sample of the words in the poem.

Exit Ticket (5 minutes)



Name _____

Date _____

Lesson 14: Selecting a Sample

Exit Ticket

Write down three things you learned about taking a sample from the work we have done today.

Exit Ticket Sample Solutions

Write down three things you learned about taking a sample from the work we have done today.

A random sample is one where every element in the set has an equal chance of being selected.

When people just choose a sample they think will be random, it will usually be different from a real random sample.

Random samples are usually similar to the population.

Problem Set Sample Solutions

The Problem Set is intended to reinforce material from the prior lesson and have students think about examples of samples that are random and those that are not.

1. Would any of the following provide a random sample of letters used in text of the book *Harry Potter and the Sorcerer's Stone* by J.K. Rowling? Explain your reasoning.
 - a. Use the first letter of every word of a randomly chosen paragraph

This is not a random sample. Some common letters, like 'u', don't appear very often as the first letter of a word and may tend to be underrepresented in the sample.
 - b. Number all of the letters in the words in a paragraph of the book, cut out the numbers, and put them in a bag. Then, choose a random set of numbers from the bag to identify which letters you will use.

This would give you a random sample of the letters.
 - c. Have a family member or friend write down a list of their favorite words and count the number of times each of the letters occurs.

This would not be a random sample. They might like words that rhyme or that all start with the same letter. The list might also include words not in the book.
2. Indicate whether the following are random samples from the given population and explain why or why not.
 - a. Population: All students in school; sample includes every fifth student in the hall outside of class.

Sample response: No, because not everyone in school would be in our hall before class—our hall only has sixth graders in it, so the seventh and eighth graders would not have a chance to be chosen.
 - b. Population: Students in your class; sample consists of students that have the letter "s" in their last name.

Sample response: No, because students that do not have the letter "s" in their last name would not have a chance to be chosen
 - c. Population: Students in your class; sample selected by putting their names in a hat and drawing the sample from the hat.

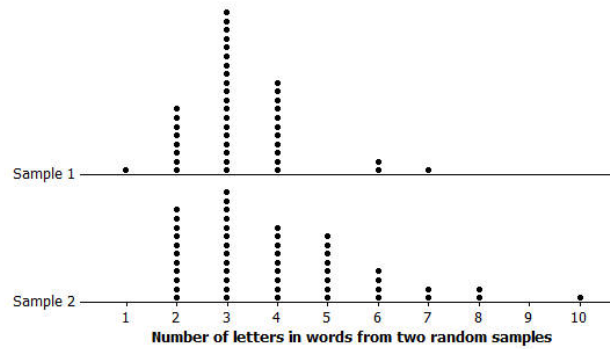
Sample response: Yes, everyone would have the same chance to be chosen.
 - d. Population: People in your neighborhood; sample includes those outside in the neighborhood at 6:00 p.m.

Sample response: No, because people who are not in the neighborhood at that time have no chance of being selected.

- e. Population: Everyone in a room; sample selected by having everyone tosses a coin and those with heads are the sample.

Sample response: Yes, everyone would have the same chance to be chosen.

3. Consider the two sample distributions of the number of letters in randomly selected words shown below:



- a. Describe each distribution using statistical terms as much as possible.

Answers will vary; the top distribution seems to have both a median and balance point, or mean, at 3, with a minimum of 1 letter in a word and a maximum of 7 letters. Most of the words in the sample were 2 to 4 letters long. The bottom distribution seems more skewed with the median of about 4 letters. The smallest number of letters was 2 and the largest was 10 letters. Most of the letters in this sample had between 2 and 5 letters.

- b. Do you think the two samples came from the same poem? Why or why not?

Sample response: The samples could have come from the same poem, but the distributions seem different both with respect to shape and to a measure of center, so it seems more likely that they were from two different populations.

4. What questions about samples and populations might you want to ask if you saw the following headlines in a newspaper?

- a. "Peach Pop is the top flavor according to 8 out of 10 people."

Sample response: How the people were selected? How many people were surveyed? What were the choices and how many did not like Peach Pop?

- b. "Candidate X looks like a winner! 10 out of 12 people indicate they will vote for Candidate X."

Sample response: How was the sample chosen? Were the people selected at random or were they friends of candidate X?

- c. "Students Overworked. Over half of 400 people surveyed think students spend too many hours on homework."

Sample response: Who was surveyed and how were they selected? Was the survey given to students in a school?

- d. "Action/adventure was selected as the favorite movie type by an overwhelming 75% of those surveyed."

Sample response: Was the survey given at a movie theater showing an action/adventure movie where people were there because they like that kind of movie?



Handout 1

1 Casey at the Bat

The Outlook wasn't brilliant for the Mudville nine that day: The score stood four to two, 2 with but one inning more to play. And then when Cooney died at first, and Barrows did the same, A 3 sickly silence fell upon the patrons of the game.

A straggling few got up to go in deep despair. The 4 rest Clung to that hope which springs eternal in the human breast; They thought, if only Casey could get but 5 a whack at that - We'd put up even money, now, with Casey at the bat.

But Flynn preceded Casey, as 6 did also Jimmy Blake, And the former was a lulu and the latter was a cake; So upon that stricken 7 multitude grim melancholy sat, For there seemed but little chance of Casey's getting to the bat.

But Flynn let drive 8 a single, to the wonderment of all, And Blake, the much despised, tore the cover off the ball; And when 9 the dust had lifted, and the men saw what had occurred, There was Jimmy safe at second and Flynn a 10 hugging third.

Then from five thousand throats and more there rose a lusty yell; It rumbled through the valley, it 11 rattled in the dell; It knocked upon the mountain and recoiled upon the flat, For Casey, mighty Casey, was advancing 12 to the bat.

There was ease in Casey's manner as he stepped into his place; There was pride in Casey's 13 bearing and a smile on Casey's face. And when, responding to the cheers, he lightly doffed his hat, No stranger 14 in the crowd could doubt 'twas Casey at the bat.

Ten thousand eyes were on him as he rubbed his 15 hands with dirt; Five thousand tongues applauded when he wiped them on his shirt. Then while the writhing pitcher ground 16 the ball into his hip, Defiance gleamed in Casey's eye, a sneer curled Casey's lip.

And now the leather covered 17 sphere came hurtling through the air, And Casey stood a-watching it in haughty grandeur there. Close by the sturdy batsman 18 the ball unheeded sped- "That ain't my style," said Casey. "Strike one," the umpire said.

From the benches, black with 19 people, there went up a muffled roar, Like the beating of the storm waves on a stern and distant shore. 20 "Kill him! Kill the umpire!" shouted someone on the stand; And its likely they'd a-killed him had not Casey raised 21 his hand.



With a smile of Christian charity great Casey's visage shone; He stilled the rising tumult; he bade the \22\ game go on; He signaled to the pitcher, and once more the spheroid flew; But Casey still ignored it, and \23\ the umpire said, "Strike two."

"Fraud!" cried the maddened thousands, and echo answered fraud; But one scornful look from Casey \24\ and the audience was awed. They saw his face grow stern and cold, they saw his muscles strain, And they \25\ knew that Casey wouldn't let that ball go by again.

The sneer is gone from Casey's lip, his teeth are \26\ clenched in hate; He pounds with cruel violence his bat upon the plate. And now the pitcher holds the ball, \27\ and now he lets it go, And now the air is shattered by the force of Casey's blow.

Oh, somewhere \28\ in this favored land the sun is shining bright; The band is playing somewhere, and somewhere hearts are light, And \29\ somewhere men are laughing, and somewhere children shout; But there is no joy in Mudville - mighty Casey has struck out.

by Ernest Lawrence Thayer