Lesson 14: Selecting a Sample

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: What is Random?

1. Write down a sequence of heads/tails you think would typically occur if you tossed a coin 20 times. Compare your sequence to the ones written by some of your classmates. How are they alike? How are they different?

2. Working with a partner, toss a coin 20 times, and write down the sequence of heads and tails you get.
   a. Compare your results with your classmates.

   b. How are your results from actually tossing the coin different from the sequences you and your classmates wrote down?

   c. Toni claimed she could make up a set of numbers that would be random. What would you say to her?

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?

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.

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?

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.

   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.

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.

8. Do you think the 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.

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 of the mean of the population. Explain your answer.

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?

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.

Lesson Summary

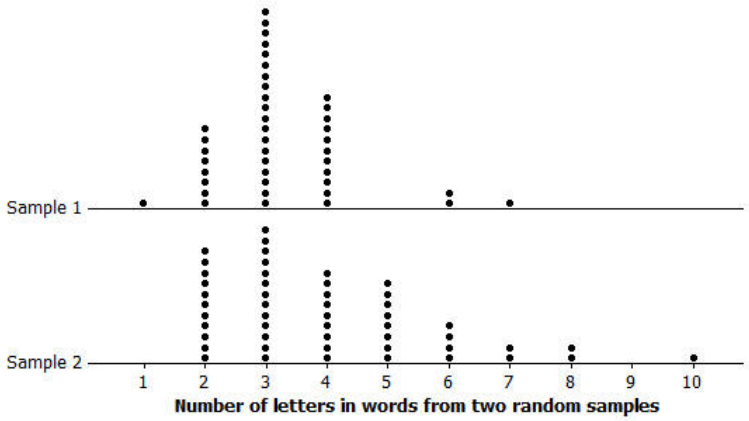
When choosing a sample, you want the sample to be representative of a population. When you try to select a sample just by yourself, you do not usually do very well, like the words you chose from the poem to find the mean number of letters. One way to help ensure that a sample is representative of the 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.

Problem Set

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.
   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.
   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.

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.
   b. Population: Students in your class; sample consists of students that have the letter "s" in their last name.
   c. Population: Students in your class; sample selected by putting their names in a hat and drawing the sample from the hat.
   d. Population: People in your neighborhood; sample includes those outside in the neighborhood at 6:00 p.m.
   e. Population: Everyone in a room; sample selected by having everyone toss a coin, and those that result in heads are the sample.

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

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."
   b. "Candidate X looks like a winner! 10 out of 12 people indicate they will vote for Candidate X."
   c. "Students overworked. Over half of 400 people surveyed think students spend too many hours on homework."
   d. "Action/adventure was selected as the favorite movie type by an overwhelming 75% of those surveyed."