Lesson 13: Populations, Samples, and Generalizing from a Sample to a Population

Classwork

In this lesson, you will learn about collecting data from a sample that is selected from a population. You will also learn about summary values for both a population and a sample and think about what can be learned about the population by looking at a sample from that population.

Exercises 1–4: Collecting Data

1. Describe what you would do if you had to collect data to investigate the following statistical questions using either a sample statistic or a population characteristic. Explain your reasoning in each case.
   a. How might you collect data to answer the question, “Does the soup taste good?”
   b. How might you collect data to answer the question, “How many movies do students in your class see in a month?”
   c. How might you collect data to answer the question, “What is the median price of a home in our town?”
   d. How might you collect data to answer the question, “How many pets do people own in my neighborhood?”
   e. How might you collect data to answer the question, “What is the typical number of absences in math classes at your school on a given day?”
f. How might you collect data to answer the question, “What is the typical life span of a particular brand of flashlight battery?”

g. How might you collect data to answer the question, “What percentage of girls and of boys in your school have a curfew?”

h. How might you collect data to answer the question, “What is the most common blood type of students in my class?”

A population is the entire set of objects (people, animals, plants, etc.) from which data might be collected. A sample is a subset of the population. Numerical summary values calculated using data from an entire population are called population characteristics. Numerical summary values calculated using data from a sample are called statistics.

2. For which of the scenarios in Exercise 1 did you describe collecting data from a population and which from a sample?

3. Think about collecting data in the scenarios above. Give at least two reasons you might want to collect data from a sample rather than from the entire population.

4. Make up a result you might get in response to the situations in Exercise 1, and identify whether the result would be based on a population characteristic or a sample statistic.
   a. Does the soup taste good?
b. How many movies do your classmates see in a month?

c. What is the median price of a home in our town?

d. How many pets do people in my neighborhood own?

e. What is the typical number of absences in math classes at your school on a given day?

f. What is the typical life span of a particular brand of flashlight batteries?

g. What percentage of girls and of boys in your school that have a curfew?

h. What is the most common blood type of my classmates?

Exercise 5: Population or Sample?

5. Indicate whether the following statements are summarizing data collected to answer a statistical question from a population or from a sample. Identify references in the statement as population characteristics or sample statistics.

a. 54% of the responders to a poll at a university indicated that wealth needed to be distributed more evenly among people.
b. Are students in the Bay Shore school district proficient on the state assessments in mathematics? In 2013, after all the tests taken by the students in the Bay Shore schools were evaluated, over 52% of those students were at or above proficient on the state assessment.

c. Does talking on mobile phones while driving distract people? Researchers measured the reaction times of 38 study participants as they talked on mobile phones and found that the average level of distraction from their driving was rated 2.25 out of 5.

d. Did most people living in New York in 2010 have at least a high school education? Based on the data collected from all New York residents in 2010 by the United States Census Bureau, 84.6% of people living in New York had at least a high school education.

e. Were there more deaths than births in the United States between July 2011 and July 2012? Data from a health service agency indicated that there were 2% more deaths than births in the U.S. during that timeframe.

f. What is the fifth best-selling book in the United States? Based on the sales of books in the United States, the fifth best-selling book was *Oh, the Places You’ll Go!* by Dr. Seuss.
Exercises 6–8: A Census

6. When data are collected from an entire population, it is called a census. The United States takes a census of its population every ten years, with the most recent one occurring in 2010. Go to http://ri.essortment.com/unitedstatesce_rita.htm to find the history of the U.S. census.
   a. Identify three things that you found to be interesting.

   b. Why is the census important in the United States?

   a. How many people were living in New York for the 2010 census?

   b. Estimate the ratio of those 65 and older to those under 18 years old. Why is this important to think about?

   c. Is the ratio a population characteristic or a statistic? Explain your thinking.

8. The American Community Survey (ACS) takes samples from a small percentage of the U.S. population in years between the censuses. (www.census.gov/acs/www/about_the_survey/american_community_survey/)
   a. What is the difference between the way the ACS collects information about the U.S. population and the way the U.S. Census Bureau collects information?
b. In 2011, the ACS sampled workers living in New York about commuting to work each day. Why do you think these data are important for the state to know?

c. Suppose that from a sample of 200,000 New York workers, 32,400 reported traveling more than an hour to work each day. From this information, statisticians determined that between 16% and 16.4% of the workers in the state traveled more than an hour to work every day in 2011. If there were 8,437,512 workers in the entire population, about how many traveled more than an hour to work each day?

d. Reasoning from a sample to the population is called making an inference about a population characteristic. Identify the statistic involved in making the inference in part (c).

e. The data about traveling time to work suggest that across the United States typically between 79.8% and 80% of commuters travel alone, 10% to 10.2% carpool, and 4.9% to 5.1% use public transportation. Survey your classmates to find out how a worker in their family gets to work. How do the results compare to the national data? What might explain any differences?
Lesson Summary

The focus of this lesson was on collecting information either from a population, which is the entire set of elements in the group of interest, or a subset of the population, called a sample. One example of data being collected from a population is the U.S. Census, which collects data from every person in the United States every ten years. The Census Bureau also samples the population to study things that affect the economy and living conditions of people in the U.S. in more detail. When data from a population are used to calculate a numerical summary, the value is called a population characteristic. When data from a sample are used to calculate a numerical summary, the value is called a sample statistic. Sample statistics can be used to learn about population characteristics.

Problem Set

1. The lunch program at Blake Middle School is being revised to align with the new nutritional standards that reduce calories and increase servings of fruits and vegetables. The administration decided to do a census of all students at Blake Middle School by giving a survey to all students about the school lunches.

   http://frac.org/federal-foodnutrition-programs/school-breakfast-program/school-meal-nutrition-standards

   a. Name some questions that you would include in the survey. Explain why you think those questions would be important to ask.

   b. Read through the paragraph below that describes some of the survey results. Then, identify the population characteristics and the sample statistics.

      About \( \frac{3}{4} \) of the students surveyed eat the school lunch regularly. The median number of days per month that students at Blake Middle School ate a school lunch was 18 days. 36\% of students responded that their favorite fruit is bananas. The survey results for Tanya’s seventh-grade homeroom showed that the median number of days per month that her classmates ate lunch at school was 22, and only 20\% liked bananas. The fiesta salad was approved by 78\% of the group of students who tried it, but when it was put on the lunch menu, only 40\% of the students liked it. Of the seventh graders as a whole, 73\% liked spicy jicama strips, but only 2 out of 5 of all the middle school students liked them.

2. For each of the following questions: (1) describe how you would collect data to answer the question, and (2) describe whether it would result in a sample statistic or a population characteristic.

   a. Where should the eighth-grade class go for their class trip?

   b. What is the average number of pets per family for families that live in your town?

   c. If people tried a new diet, what percentage would have an improvement in cholesterol reading?

   d. What is the average grade point of students who got accepted to a particular state university?

   e. What is a typical number of home runs hit in a particular season for major league baseball players?
3. Identify a question that would lead to collecting data from the given set as a population, and one where the data could be a sample from a larger population.
   a. All students in your school
   b. Your state

4. Suppose that researchers sampled attendees of a certain movie and found that the mean age was 17 years old. Based on this observation, which of the following would be most likely.
   a. The mean ages of all of the people who went to see the movie was 17 years old.
   b. About a fourth of the people who went to see the movie were older than 51.
   c. The mean age of all people who went to see the movie would probably be in an interval around 17 years of age, i.e., between 15 and 19.
   d. The median age of those who attended the movie was 17 years old as well.

5. The headlines proclaimed: “Education Impacts Work-Life Earnings Five Times More Than Other Demographic Factors, Census Bureau Reports.” According to a U.S. Census Bureau study, education levels had more effect on earnings over a 40-year span in the workforce than any other demographic factor. [link]
   a. The article stated that the estimated impact on annual earnings between a professional degree and an eighth-grade education was roughly five times the impact of gender, which was $13,000. What would the difference in annual earnings be with a professional degree and with an eighth-grade education?
   b. Explain whether you think the data are from a population or a sample, and identify either the population characteristic or the sample statistic.
The History of the United States Census

The word census is Latin in origin and means to tax. The first U.S. census took place over 200 years ago, but the United States is certainly not the first country to implement a census. Based on archaeological records, it appears that the ancient Egyptians conducted a census as early as 3000 B.C.

The U.S. census is mandated by the U.S. Constitution in Article I, Section 2, which states, in part, “Representatives and direct Taxes shall be apportioned among the several States . . . according to their respective Numbers . . . . The Number of Representatives shall not exceed one for every thirty thousand, but each State shall have at Least one Representative . . . .” The Constitution then specifies how to calculate the number of people in each state and how often the census should take place.

The U.S. census has been conducted every ten years since 1790, but as time has passed, our census has evolved. Not only have the types of questions changed but also the manner in which the data are collected and tabulated. Originally, the census had only a few questions, the purpose of which was to discern the number of people in each household and their ages. Presumably, this data was used to determine the number of men in each state who were available to go to war. Federal marshals were charged with the task of conducting this first census. After collecting data from their respective jurisdictions, the marshals sent the data to President Washington.

As time has passed, more questions have been added to the U.S. census. Today, the census includes questions designed to collect data in various fields such as manufacturing, commerce, and transportation to name a few. Data that were once manually tabulated are now processed by computers. Home visits by census officials were once the norm, but now the census is conducted primarily through the U.S. Postal Service. Each household in the U.S. receives in the mail a copy of the census questionnaire to be completed by its head of household who then mails it back to the Census Bureau. Home visits are paid only to those individuals who do not return the questionnaire by the specified deadline.

The census is an important part of our Constitution. Today, the census not only tells us the population of each state, thereby determining the number of representatives that each state will have in the House of Representatives, but it also provides the U.S. government with very useful data that paint a picture of the current state of our population and how it has changed over the decades.

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.

<table>
<thead>
<tr>
<th>Exercise 4 Mean</th>
<th>Exercise 6 Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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:

![Sample Distributions](image)

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.”
Lesson 15: Random Sampling

Classwork

In this lesson, you will investigate taking random samples and how random samples from the same population vary.

Exercises 1–5: Sampling Pennies

1. Do you think different random samples from the same population will be fairly similar? Explain your reasoning.

2. The plot below shows the number of years since being minted (the penny age) for 150 pennies that JJ had collected over the past year. Describe the shape, center, and spread of the distribution.

![Dot Plot of Population of Penny Ages](image-url)
3. Place ten dots on the number line that you think might be the distribution of a sample of 10 pennies from the jar.

4. Select a random sample of 10 pennies, and make a dot plot of the ages. Describe the distribution of the penny ages in your sample. How does it compare to the population distribution?

5. Compare your sample distribution to the sample distributions on the board.
   a. What do you observe?
   b. How does your sample distribution compare to those on the board?

Exercises 6–9: Grocery Prices and Rounding

6. Look over some of the grocery prices for this activity. Consider the following statistical question, “Do the store owners price the merchandise with cents that are closer to a higher dollar value or a lower dollar value?” Describe a plan that might answer that question that does not involve working with all 100 items.
7. Do the store owners price the merchandise with cents that are closer to a higher dollar value or a lower dollar value? To investigate this question in one situation, you will look at some grocery prices in weekly flyers and advertising for local grocery stores.
   a. How would you round $3.49 and $4.99 to the nearest dollar?
   b. If the advertised price was three for $4.35, how much would you expect to pay for one item?
   c. Do you think more grocery prices will round up or round down? Explain your thinking.

8. Follow your teacher’s instructions to cut out the items and their prices from the weekly flyers, and put them in a bag. Select a random sample of 25 items without replacement, and record the items and their prices in the table below.

<table>
<thead>
<tr>
<th>Item</th>
<th>Price</th>
<th>Rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Price</th>
<th>Rounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example of chart suggested:

<table>
<thead>
<tr>
<th>Student</th>
<th>Number of times prices were rounded to the higher value</th>
<th>Number of times the prices were rounded to the lower value</th>
<th>Percent of prices rounded up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bettina</td>
<td>20</td>
<td>5</td>
<td>80%</td>
</tr>
</tbody>
</table>

9. Round each of the prices in your sample to the nearest dollar, and count the number of times you rounded up and the number of times you rounded down.
   a. Given the results of your sample, how would you answer the question: Are grocery prices in the weekly ads at the local grocery closer to a higher dollar value or a lower dollar value?

   b. Share your results with classmates who used the same flyer or ads. Looking at the results of several different samples, how would you answer the question in part (a)?

   c. Identify the population, sample, and sample statistic used to answer the statistical question.

   d. Bettina says that over half of all the prices in the grocery store will round up. What would you say to her?
Lesson Summary

In this lesson, you took random samples in two different scenarios. In the first scenario, you physically reached into a jar and drew a random sample from a population of pennies. In the second scenario, you drew items from a bag and recorded the prices. In both activities, you investigated how random samples of the same size from the same population varied. Even with sample sizes of 10, the sample distributions of pennies were somewhat similar to the population (the distribution of penny ages was skewed right, and the samples all had 0 as an element). In both cases, the samples tended to have similar characteristics. For example, the samples of prices from the same store had about the same percent of prices that rounded to the higher dollar value.

Problem Set

1. Look at the distribution of years since the pennies were minted from Example 1. Which of the following box plots seem like they might not have come from a random sample from that distribution? Explain your thinking.

   ![Box Plots of Three Random Samples of Penny Ages](image)

2. Given the following sample of scores on a physical fitness test, from which of the following populations might the sample have been chosen? Explain your reasoning.

   ![Dot Plots of Four Populations and One Sample](image)
3. Consider the distribution below:

![Distribution Plot](image)

a. What would you expect the distribution of a random sample of size 10 from this population to look like?

b. Random samples of different sizes that were selected from the population in part (a) are displayed below. How did your answer to part (a) compare to these samples of size 10?

![Dot Plots of Five Samples of Different Sizes](image)

c. Why is it reasonable to think that these samples could have come from the above population?

d. What do you observe about the sample distributions as the sample size increases?

4. Based on your random sample of prices from Exercise 2, answer the following questions:

a. It looks like a lot of the prices end in 9. Do your sample results support that claim? Why or why not?

b. What is the typical price of the items in your sample? Explain how you found the price and why you chose that method.
5. The sample distributions of prices for three different random samples of 25 items from a grocery store are shown below.
   
   a. How do the distributions compare?

   ![Dot Plots of Three Samples]

   b. Thomas says that if he counts the items in his cart at that grocery store and multiplies by $2.00, he will have a pretty good estimate of how much he will have to pay. What do you think of his strategy?