Lesson 16: Methods for Selecting a Random Sample

Student Outcomes

- Students select a random sample from a population.
- Given a description of a population, students design a plan for selecting a random sample from that population.

Lesson Notes

In this lesson, students use random numbers to select a random sample. Unlike the previous lessons where students selected at random from a physical population, in this lesson the random selection is based on using random numbers to identify the specific individuals in the population that should be included in the sample. Students also design a plan for selecting a random sample to answer a statistical question about a population.

A variety of methods can be used to generate random numbers. The TI graphing calculators have a random number generator that can be used if calculators are available. For this lesson, students generate random integers. Random number generators can also be found on a number of websites. (A specific website is referenced later in this lesson that could be used to create a list of random numbers. Directions needed to use the random number generator are provided on the website.) If calculators or access to websites are not possible, simply create a random number bag. Write the numbers from 1 to 150 on small slips of paper. Students select slips of paper from the bag to form their lists of random numbers. Having access to technology makes it easier for students to concentrate on the concepts rather than counting and sorting numbers. If using technology is not possible, each student or pair of students should do only one or two examples, and the class data should be collected and displayed.

Classwork

Students read the paragraph silently.

In this lesson, you will obtain random numbers to select a random sample. You will also design a plan for selecting a random sample to answer a statistical question about a population.

Example 1 (2 minutes): Sampling Children’s Books

Introduce the data in the table, and examine the histogram.

Example 1: Sampling Children’s Books

What is the longest book you have ever read? The Hobbit has 95,022 words, and The Cat in the Hat has 830 words. Popular books vary in the number of words they have—not just the number of different words but the total number of words. The table on the following page shows the total number of words in some of those books. The histogram displays the total number of words in 150 best-selling children’s books with fewer than 100,000 words.
<table>
<thead>
<tr>
<th>Book</th>
<th>Words</th>
<th>Book</th>
<th>Words</th>
<th>Book</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black Beauty</td>
<td>59,635</td>
<td>Charlie and the Chocolate Factory</td>
<td>30,644</td>
<td>The Hobbit</td>
<td>95,022</td>
</tr>
<tr>
<td>The Catcher in the Rye</td>
<td>73,404</td>
<td>Old Yeller</td>
<td>35,968</td>
<td>Judy Moody Was in a Mood</td>
<td>11,049</td>
</tr>
<tr>
<td>The Adventures of Tom Sawyer</td>
<td>69,066</td>
<td>The Cat in the Hat</td>
<td>830</td>
<td>Treasure Island</td>
<td>66,950</td>
</tr>
<tr>
<td>The Secret Garden</td>
<td>80,398</td>
<td>Green Eggs and Ham</td>
<td>702</td>
<td>Magic Tree House Lions at Lunchtime</td>
<td>5,313</td>
</tr>
<tr>
<td>The Mouse and the Motorcycle</td>
<td>22,416</td>
<td>Little Bear</td>
<td>1,630</td>
<td>Harry Potter and the Sorcerer's Stone</td>
<td>77,325</td>
</tr>
<tr>
<td>The Wind in the Willows</td>
<td>58,424</td>
<td>The Red Badge of Courage</td>
<td>47,180</td>
<td>Harry Potter and the Chamber of Secrets</td>
<td>84,799</td>
</tr>
<tr>
<td>My Father's Dragon</td>
<td>7,682</td>
<td>Anne Frank: The Diary of a Young Girl</td>
<td>82,762</td>
<td>Junie B. Jones and the Stupid Smelly Bus</td>
<td>6,570</td>
</tr>
<tr>
<td>Frog and Toad All Year</td>
<td>1,727</td>
<td>Midnight for Charlie Bone</td>
<td>65,006</td>
<td>White Mountains</td>
<td>44,763</td>
</tr>
<tr>
<td>Book of Three</td>
<td>46,926</td>
<td>The Lion, The Witch, and the Wardrobe</td>
<td>36,363</td>
<td>Double Fudge</td>
<td>38,860</td>
</tr>
</tbody>
</table>

- In which interval is Black Beauty? The Cat in the Hat?
  - Black Beauty is in the interval 55,000 to 60,000 words, while The Cat in the Hat is in the first interval of 0 to 5,000 words.

- What is the meaning of the first bar in the histogram?
  - The first bar indicates the number of books with total number of words between 0 and 4,999.
Exercises 1–2 (3 minutes)

Students work independently on Exercises 1 and 2. Have students compare their dot plots with a neighbor.

Exercises 1–2

1. From the table, choose two books with which you are familiar, and describe their locations in the data distribution shown in the histogram.

   *Answers will vary. Sample response: I read The Mouse and the Motorcycle, and that has 22,416 words. It is below the median number of words and may be below the lower quartile. Harry Potter and the Chamber of Secrets has 84,799 words. It is one of the books with lots of words but may not be in the top quarter for the total number of words.*

2. Put dots on the number line below that you think would represent a random sample of size 10 from the number of words distribution on the previous page.

   *Answers will vary. The sample distribution might have more values near the maximum and minimum than in the center.*

Example 2 (4 minutes): Using Random Numbers to Select a Sample

Students read and answer the question posed in the Example 2.

Example 2: Using Random Numbers to Select a Sample

The histogram indicates the differences in the number of words in the collection of 150 books. How many words are typical for a best-selling children’s book? Answering this question would involve collecting data, and there would be variability in those data. This makes the question a statistical question. Think about the 150 books used to create the histogram on the previous page as a population. How would you go about collecting data to determine the typical number of words for the books in this population?

*Sample response: I would add up all of the words in the 150 books and divide by 150. This would be the mean number of words for the 150 books. As the data distribution is not symmetrical, I could also find the median of the 150 books, as it would be a good description of the typical number of words. (Note: Discuss with students that using data for all 150 books is very tedious. As a result, students may indicate that selecting a random sample of the 150 books might be a good way to learn about the number of words in these children’s books.)*

Discuss students’ suggestions for choosing a random sample. Be sure to bring out the following points: To choose a random sample, number all of the books, put the numbers in a bag, and then draw the sample from the bag. Another way is to use a random number generator, where instead of pulling numbers out of a bag, the generator selects the numbers to use in the sample. (In this discussion, observe how students make sense of random selection and generating a random sample.)
How would you choose a random sample from the collection of 150 books discussed in this lesson?

Sample response: I would make 150 slips of paper that contained the names of the books. I would then put the slips of paper in a bag and select 10 or 15 books. The number of pages of the books selected would be my sample.

The data for the number of words in the 150 best-selling children’s books are listed below. Select a random sample of the number of words for 10 books.

If necessary, explain how to use ten numbers selected from a bag that contains the numbers from 1 to 150 to select the books for the sample.

If students need more direction in finding a random sample, develop the following example: Consider the following random numbers obtained by drawing slips from a bag that contained the numbers 1 to 150: {114, 65, 77, 38, 86, 105, 50, 1, 56, 85}. These numbers represent the randomly selected books. To find the number of words in those books, order the random numbers {1, 38, 50, 56, 75, 77, 85, 86, 105, 114}. Count from left to right across the first row of the list of the number of words, then down to the second row, and so on. The sample consists of the 1st element in the list, the 38th, the 50th, and so on.

Use the above example of random numbers to help students connect the random numbers to the books selected and to the number of words in those books.

- What number of words corresponds to the book identified by the random number 1?
  - 59,635; the 1st children’s book listed has 59,635 words.

- What number of words corresponds to the book identified by the random number 38?
  - 3,252; the 38th children’s book listed has 3,252 words.

| Books 1–10  | 59,635 | 82,762 | 92,410 | 75,340 | 8,234 | 59,705 | 92,409 | 75,338 | 8,230 | 82,768 |
| Books 11–20 | 73,404 | 65,006 | 88,250 | 2,100  | 81,450 | 72,404 | 88,252 | 2,099  | 81,451 | 65,011 |
| Books 21–30 | 69,066 | 36,363 | 75,000 | 3,000  | 80,798 | 69,165 | 75,012 | 3,010  | 80,790 | 36,361 |
| Books 31–40 | 80,398 | 95,022 | 71,200 | 3,250  | 81,450 | 80,402 | 71,198 | 3,252  | 81,455 | 95,032 |
| Books 41–50 | 22,416 | 11,049 | 81,400 | 3,100  | 83,475 | 22,476 | 81,388 | 3,101  | 83,472 | 11,047 |
| Books 51–60 | 58,424 | 66,950 | 92,400 | 2,750  | 9,000  | 58,481 | 92,405 | 2,748  | 9,002  | 66,954 |
| Books 61–70 | 7,682  | 5,313  | 83,000 | 87,000 | 89,170 | 7,675  | 83,021 | 87,008 | 89,167 | 5,311  |
| Books 71–80 | 1,727  | 77,325 | 89,010 | 862   | 88,365 | 1,702  | 89,015 | 860   | 88,368 | 77,328 |
| Books 81–90 | 46,926 | 84,799 | 88,045 | 927   | 89,790 | 46,986 | 88,042 | 926   | 89,766 | 84,796 |
| Books 91–100| 30,644 | 6,570  | 90,000 | 8,410  | 91,010 | 30,692 | 90,009 | 8,408  | 91,015 | 6,574  |
| Books 101–110|15,968 |44,763 |89,210 | 510  | 9,247  | 35,940 | 89,213 | 512  | 9,249 | 44,766 |
| Books 111–120|830  |8,700 |92,040 | 7,891 | 83,150 | 838  | 92,037 | 7,889 | 83,149 | 8,705  |
| Books 121–130|702  |92,410 |94,505 | 38,860 | 81,110 | 712  | 94,503 | 87,797 | 81,111 | 92,412 |
| Books 131–140|1,630 |88,250 |97,000 | 7,549 | 8,245 | 1,632 | 97,002 | 7,547 | 8,243 | 88,254 |
| Books 141–150|47,180 |75,000 |89,241 | 81,234 | 8,735 | 47,192 | 89,239 | 81,238 | 8,739 | 75,010 |
Exercises 3–6 (10 minutes)

In this set of exercises, students select a random sample by using a random number generator or slips of paper in a bag. To generate a set of random numbers, consider directing students to a website, such as www.rossmanchance.com/applets/RandomGen/GenRandom01.htm, or to a random number generator on a graphing calculator. Be sure that the numbers generated are unique, or without replacement, so that no number is used twice. If students’ access to technology is problematic, demonstrate a random number generator for them, and have them copy down the numbers generated. If using a graphing calculator similar to the TI-84, be sure to seed the calculator by completing the command: RandSeed #, where # is any number unique to the student, such as the last four digits of a phone number. If this is not done, all of the “random” numbers may begin at the same place, and the samples will all be the same. However, once they have seeded their random number generators one time, they do not have to do this again unless the teacher wants the whole class to use the same seed so that the entire class produces the same set of numbers.

Students work in pairs with one student counting and the other recording the sample values.

3. Follow your teacher’s instructions to generate a set of 10 random numbers. Find the total number of words corresponding to each book identified by your random numbers.

   *Answers will vary.*  *Sample response: I generated random numbers 123, 25, 117, 119, 93, 135, 147, 69, 48, 46, which produces the sample 94505, 80798, 92037, 83149, 90000, 8245, 89239, 89167, 3101, 22476.*

4. Choose two more different random samples of size 10 from the data, and make a dot plot of each of the three samples.

   *Answers will vary.*  *One possible response is displayed below.*

   ![Dot Plot of Total Number of Words for Three Random Samples](image)

5. If your teacher randomly chooses 10 books from your summer vacation reading list, would you be likely to get many books with a lot of words? Explain your thinking using statistical terms.

   *Answers will vary.*  *Sample response: From my samples, it looks like I probably would get at least one book that had over 90,000 words because the maximum in each of the samples approached or exceeded 90,000 words. The three samples vary a lot, probably because the sample size is only 10. The median numbers of words for the three samples were about 86,000, 35,000, and 70,000, respectively, so it seems like at least half of the books would contain about 50,000 or more words.*
**Exercises 7–9 (19 minutes): A Statistical Study of Balance and Grade**

The following exercises ask students to develop a plan to investigate a statistical question regarding balance. Actually carrying out the activity is optional given the challenges summarized in this explanation. The goal of the activity is for students to collect real data from a random sample to help them learn about a population.

This activity may take different forms in different schools, depending on the size of the school. It is important to select a population that is large enough that taking a random sample would be a reasonable way to investigate the population. The statistical question is framed to investigate whether seventh graders have better balance than sixth graders. A sample of 10 sixth graders and a sample of 10 seventh graders would work, which would mean starting with populations of at least approximately 100 sixth graders and 100 seventh graders ideally. This may not be possible in some schools. In very small schools, teachers might find another school with which to partner for the activity. If pooling together students is not possible and the number of sixth and seventh graders is too small, use different populations, but be explicit as to which populations students are using. Clearly state that the populations of interest are all of the sixth and seventh graders in the school, and use a reasonable sample size for the selection of the random samples. Also, consider modifying the statistical question so that the groups compared represent a larger number of students in the school (e.g., students in Grades 7 and 8 compared to students in Grades 5 and 6, or whether students in Grades 7 and 8 spend more time on homework than students in Grades 5 and 6). Students can complete the exercises that involve planning this activity, even if it is not possible to actually carry out the data collection.

In completing the exercises, students have to think about how to use random numbers to select a random sample. The need for a random sample is based on the premise that it would take too much time and it would be too difficult to carry out a study using all students in the two populations. If possible, consider having students actually carry out the data collection, but that takes some time, probably two days to plan, collect, and analyze the data. If there is time to do this, it is time well spent, as the activity engages students in important aspects of the entire statistical process: beginning with a statistical question, designing a study, collecting data, analyzing the data, and using the results to answer the question.

Cell phones or stopwatches can be used for timing balance. (Some science departments have stopwatches they might lend to students.)

Students put their complete plans for Exercise 9 on chart paper or some other public display, and each group shares their thinking. The other students might anonymously write on a 3 × 5 notecard one thing they like about the plan being presented and one thing that concerns them. If the class actually carries out the activity, Exercise 9 could be done as a whole class, or the class could vote to determine which of the groups’ plans they would like to use.

**Exercise 7:** Is the following question a statistical question: Do sixth graders or seventh graders tend to have better balance?

*Yes, this is a statistical question because the data collected would vary between sixth and seventh graders.*
8. Berthio’s class decided to measure balance by finding out how long people can stand on one foot.
   a. How would you rephrase the question from Exercise 7 to create a statistical question using this definition of balance? Explain your reasoning.
      Answers will vary. Sample response: What is the typical length of time that a seventh grader can balance on one foot? The data collected to answer this question will have some variability—1 min, 2 min, 2 min, 10 sec, and so on. So, it is also a statistical question.
   b. What should the class think about to be consistent in how they collect the data if they actually have people stand on one foot and measure the time?
      Sample response: Would it make a difference if students stood on their right feet or on their left? How high do they have to hold their feet off the ground? Can they do it barefoot or with shoes on? Would tennis shoes be better than shoes with higher heels? What can we use to measure the time?

9. Work with your class to devise a plan to select a random sample of sixth graders and a random sample of seventh graders to measure their balance using Berthio’s method. Then, write a paragraph describing how you will collect data to determine whether there is a difference in how long sixth graders and seventh graders can stand on one foot. Your plan should answer the following questions:
   a. What is the population? How will samples be selected from the population? Why is it important that they be random samples?
      Sample response: The populations will be all of the sixth graders and all of the seventh graders in our school. To get a random sample, we will find the number of sixth graders, say 620, and generate a list of 199 random numbers from the set 1 to 62, that is, {4, 17, 19, 25, ...}. Then, we will go into one classroom and count off the students beginning with 1 and use student 4, 17, and 19. Then we will go into the next classroom and count off the students beginning where we left off in the first room and so on. We will do the same for the seventh graders. This will give random samples because it offers every sixth and seventh grader the same chance of being selected (if using this plan with both grades).
   b. How would you conduct the activity?
      Sample response: Students will stand for as long as they can using whichever foot they choose in their stocking or bare feet with their eyes open. We will time them to the nearest second using stopwatches from our science class. We will have students do the activity one at a time out in the hall so they cannot see each other.
   c. What sample statistics will you calculate, and how will you display and analyze the data?
      Sample response: The sample statistics will be the mean time (in seconds) standing on one foot for the sixth graders and seventh graders. We will make a dot plot of the times for the sixth graders and for the seventh graders using parallel number lines with the same scale.
   d. What would you accept as evidence that there actually is a difference in how long sixth graders can stand on one foot compared to seventh graders?
      We will compare the shape, center, and spread of the sample distributions of times for the sixth graders and do the same for the seventh graders. If the mean times are fairly close together and the spreads not that different, there is not really evidence to say one group of students has better balance.
Closing (3 minutes)

Consider posing the following questions. Allow a few student responses for each:

- Sallee argued that the set \( \{20, 24, 27, 32, 35, 40, 45, 50, 120, 500\} \) could not possibly be a random sample of ten numbers from 1 to 500 because the set had too many small numbers. Do you agree or disagree with Sallee? Explain your thinking.
  - Every possible set of ten numbers from 1 to 500 would be a possible random sample, so Sallee is not correct.

- Why is it important to choose a random sample when you are collecting data?
  - If you do not have a random sample, your sample may not reflect the population and, therefore, would not offer accurate information about the population.

Exit Ticket (4 minutes)
Lesson 16: Methods for Selecting a Random Sample

Exit Ticket

1. Name two things to consider when you are planning how to select a random sample.

2. Consider a population consisting of the 200 seventh graders at a particular middle school. Describe how you might select a random sample of 20 students from a list of the students in this population.
Exit Ticket Sample Solutions

1. Name two things to consider when you are planning how to select a random sample.
   
   **Answers will vary.**
   
   • What it means to be a random sample—that everyone in the population has the same chance to be selected.
   • Is there a way to use a random number generator to make it easier to select the sample?

2. Consider a population consisting of the 200 seventh graders at a particular middle school. Describe how you might select a random sample of 20 students from a list of the students in this population.

   **Answers will vary.** Number the students on the list from 001 to 200. Using a random number generator, get 20 different random numbers between 001 and 200, and then select the students corresponding to those numbers on the list. It would also be correct for a student to say that she would write the 200 names on slips of paper, put them in a bag, mix them well, and then select 20 names from the bag.

Problem Set Sample Solutions

Students should do Problems 1 and 3 to be sure they understand the concepts in the lesson. Problem 2(b) can be used to engage students in generating and analyzing random samples using technology. Consider having students work in pairs to generate and record the numbers in the random samples to see whether the teacher’s method of collecting homework turns out to be relatively “fair” for students.

1. The suggestions below for how to choose a random sample of students at your school were made and vetoed. Explain why you think each was vetoed.
   
   a. Use every fifth person you see in the hallway before class starts.
      
      *Students who are not in the hallway because they have a class in another part of the building would not have a chance to be selected, so the sample would not be a random sample.*

   b. Use all of the students taking math the same time that your class meets.
      
      *The students not taking math at that time would not have a chance of being selected, so the sample would not be a random sample.*

   c. Have students who come to school early do the activity before school starts.
      
      *The sample would not be a random sample because some students would not be able to get to school early, so they could not be selected.*

   d. Have everyone in the class find two friends to be in the sample.
      
      *Choosing people that members of the class know would not be a random sample because people that members of the class do not know have no chance to be chosen.*
2. A teacher decided to collect homework from a random sample of her students rather than grading every paper every day.
   a. Describe how she might choose a random sample of five students from her class of 35 students.
      Sample response: You could assign each student a number from 1 to 35, generate five random numbers from 1 to 35, and choose the corresponding students.
   b. Suppose every day for 75 days throughout an entire semester she chooses a random sample of five students. Do you think some students will never get selected? Why or why not?
      Sample response: Over that many days, it should almost even out. If you think about 375 numbers generated all together with each number from 1 to 35 having an equal chance of showing up each time, then each number should be in the overall set about 10 or 11 times. I generated 75 random samples of numbers from 1 to 35 and looked at how the numbers showed up. Every number from 1 to 35 showed up at least three times, and most of the numbers showed up about 10 or 11 times.

3. Think back to earlier lessons in which you chose a random sample. Describe how you could have used a random number generator to select a random sample in each case.
   a. A random sample of the words in the poem “Casey at the Bat”
      Sample response: You could have generated the random numbers from 1 to 0066 for the block of words and the random numbers 1 to 0000 to choose a word in the block. Or you could number all of the words from 1 to 99000 and then generate random numbers between 1 and 99000 to choose the words.
   b. A random sample of the grocery prices on a weekly flyer
      Sample response: Instead of cutting out all of the prices and putting them in a bag, you could just number them on the flyer and use the random number generator to select numbers to identify the items in the sample and use the price of those items.

4. Sofia decided to use a different plan for selecting a random sample of books from the population of 119900 top-selling children’s books from Example 2. She generated ten random numbers between 1 and 100,000 to stand for the possible number of pages in any of the books. Then, she found the books that had the number of pages specified in the sample. What would you say to Sofia?
   Sample response: She would have to reject the numbers in the sample that referred to pages that were not in her list of 150 books. For example, if she gets the random numbers 4 or 720, she would have to generate new numbers because no books on the list had either 4 or 720 pages. She would have to throw out a lot of random numbers that did not match the number of pages in the books in the list. It would take her a long time. But if there were no two books that had the same total number of words in the population, it would be a random sample if she wanted to do it that way. However, because there are quite a few books that have the same number of words as other books in the population, this method would not work for selecting a random sample of the books.

5. Find an example from a newspaper, a magazine, or another source that used a sample. Describe the population, the sample, the sample statistic, how you think the sample might have been chosen, and whether or not you think the sample was random.
   Responses will vary depending on the articles students find. For example, “an estimated 60% of the eligible children in Wisconsin did not attend preschool in 2009.” The population would be all of the children in Wisconsin eligible for preschool in 2009, and the sample would be the ones selected by the study. The sample statistic would be 60%. The article did not tell how the sample was chosen but said the source was from the Census Bureau, so it was probably a random sample.