



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 will 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 will be based on using random numbers to identify the specific individuals in the population that should be included in the sample. Students will 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 will generate random integers. Random number generators can also be found on a number of websites. (A specific website is referenced below that could be used to create a list of random numbers. Directions needed to use the random number generator are provided at the website.) If calculators or access to websites are not possible, simply create a random number bag. Write the numbers from to on small slips of paper. Students will select slips of paper from the bag to form their list of random numbers. Having access to technology will make it easier for students to concentrate on the concepts rather than counting and sorting numbers. If you are not able to use technology, each student or pair of students should do only one or two examples, and the class data should be collected and displayed.

Classwork

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

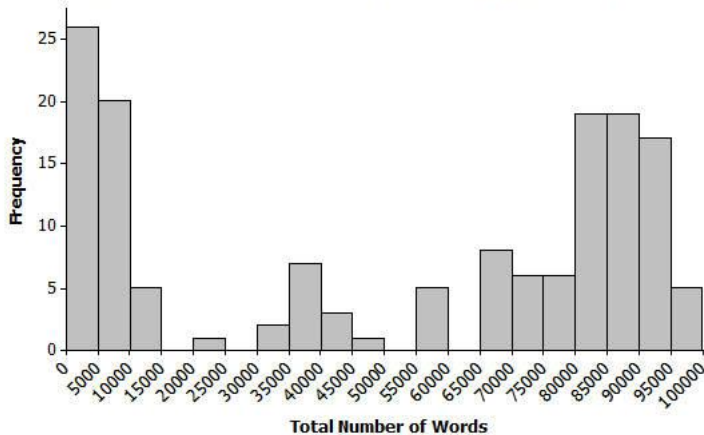
Introduce the data in the table and examine the histogram.

Example 1: Sampling Children's Books

What is the longest book you have ever read? *The Hobbit* has _____ words and *The Cat in the Hat* has _____ 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 below shows the total number of words in some of those books. The histogram displays the total number of words in _____ best-selling children's books with fewer than _____ words.

Book	Words	Book	Words	Book	Words
<i>Black Beauty</i>		<i>Charlie and the Chocolate Factory</i>		<i>The Hobbit</i>	
<i>Catcher in the Rye</i>		<i>Old Yeller</i>		<i>Judy Moody was in a Mood</i>	
<i>The Adventures of Tom Sawyer</i>		<i>Cat in the Hat</i>		<i>Treasure Island</i>	
<i>The Secret Garden</i>		<i>Green Eggs and Ham</i>		<i>Magic Tree House Lions at Lunchtime</i>	
<i>Mouse and the Motorcycle</i>		<i>Little Bear</i>		<i>Philosopher's Stone</i>	
<i>The Wind in the Willows</i>		<i>Red Badge of Courage</i>		<i>Chamber of Secrets</i>	
<i>My Father's Dragon</i>		<i>Anne Frank: The Diary of a Young Girl</i>		<i>Junie B. Jones and the Stupid Smelly Bus</i>	
<i>Frog and Toad All Year</i>		<i>Midnight for Charlie Bone</i>		<i>White Mountains</i>	
<i>Book of Three</i>		<i>The Lion, The Witch and the Wardrobe</i>		<i>Double Fudge</i>	

Total Number of Words for 150 Best-Selling Children's Books



- In which interval is *Black Beauty*? *Cat in the Hat*?
 - *Black Beauty* is in the interval _____ to _____ words, while *Cat in the Hat* is in the first interval of _____ to _____ 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 _____ and _____.*

Exercises 1–2 (3 minutes)

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

Exercises 1–2

- 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 _____ words. It is below the median number of words and maybe below the lower quartile. *The Chamber of Secrets* has _____ words. It is one of the books with lots of words, but maybe not in the top quarter for the total number of words.*
- Put dots on the number line below that you think would represent a random sample of size _____ from the number of words distribution above.

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

Example 2 (3–5 minutes): Using a Random Number Generator to Select a Sample

Example 2: Using Random Numbers to Select a Sample

The histogram indicates the differences in the number of words in the collection of _____ 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 that data. This makes the question a statistical question. Think about the _____ books used to create the histogram above 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 _____ books and divide by _____. This would be the mean number of words for the _____ books. As the data distribution is not symmetrical, I could also find the median of the _____ books, as it would be a good description of the typical number of words. (Note: Discuss with students that using data for all _____ books is very tedious. As a result, students may indicate that selecting a random sample of the _____ books might be a good way to learn about the number of words in these children’s books.)

MP.1 Discuss students’ suggestions for choosing a random sample. Be sure to bring out the following points: to choose a random sample, you could number all of the books, put the numbers in a bag, and then draw your 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 _____ books discussed in this lesson?

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

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

If necessary, explain how to use ten numbers selected from a bag that contains the numbers from _____ to _____ 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 _____ to _____ :
 _____ . These numbers represent the randomly selected books. To find the number of words in those books, order the random numbers _____ . 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 will consist 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 _____ ?
 - _____ ; the first children’s book listed has _____ words.
- What number of words corresponds to the book identified by the random number _____ ?
 - _____ ; the 38th children’s book listed has _____ words.

Books 1–10
Books 11–20
Books 21–30
Books 31–40
Books 41–50
Books 51–60
Books 61–70
Books 71–80
Books 81–90
Books 91–100
Books 101–110
Books 111–120
Books 121–130
Books 131–140
Books 141–150

Exercises 3–6 (10–12 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 site, 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 you generate. If you use 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 generator one time, they do not have to do this again unless you are in a situation where you want the whole class to use the same seed so that the entire class will produce the same set of numbers.

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

Exercises 3–6

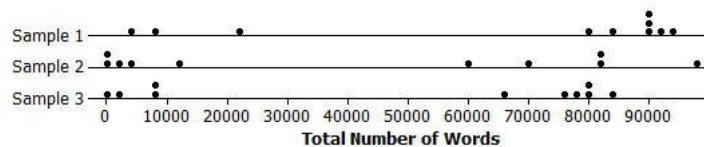
- Follow your teacher’s instructions to generate a set of random numbers. Find the total number of words corresponding to each book identified by your random numbers.

Answers will vary. Sample response: I generated random numbers which produces the sample

- Choose two more different random samples of size 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



- If your teacher randomly chooses 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 words because the maximum in each of the samples approached or exceeded words. The three samples vary a lot, probably because the sample size is only . The median numbers of words for the three samples were about , , and , respectively, so it seems like at least half of the books would contain about or more words.

6. If you were to compare your samples with your classmates' samples, do you think your answer to Exercise 5 would change? Why or why not?

Answers will vary. Sample response: The sample size is pretty small so different samples might be different. I still think that I would have to read some books with a lot of words because of the shape of the population distribution.

Exercises 7–9 (19 minutes): A Statistical Study of Balance

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 sixth graders and a sample of seventh graders would work, which would mean starting with populations of at least approximately sixth graders and seventh graders ideal. 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, you can 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. You may also modify the statistical question so that the groups compared represent a larger number of the students in your school (for example, students in Grades 7 and 8 compared to students in Grades 5 and 6 or do 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, you may want students to actually carry out the data collection, but that will take some time, probably two days to plan, collect, and analyze the data. If you can find the time to do this, it will be 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 the students.)

MP.

Students should put their complete plans for Exercise 9 on chart paper or some other public display, and each group should share their thinking. The other students might anonymously write on a 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.

Exercises 7–9: A Statistical Study of Balance and Grade

7. Is the following question a statistical question, "Do sixth graders or seventh graders tend to have better balance?" Why or why not?

Yes, this is a statistical question because it would be answered by collecting data on balance and there would be variability in the data collected. It would be important to think about how balance will be measured.

8. Berthio's class decided to measure balance by finding out how long people can stand on one foot.
- How would you rephrase the question above to create a statistical question using this definition of balance? Explain your reasoning.
Answers will vary. Sample response: Can sixth graders balance on one foot longer than seventh graders? The data collected to answer this question will have some variability – min., min., min. sec., and so on. So, it is also a statistical question.
 - 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 foot or on their left? How high do they have to hold the foot 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 leg. Your plan should answer the following questions:
- What is the population? How will samples be selected from the population? And, 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 , and generate a list of random numbers from the set to , i.e., . Then we will go into one classroom and count off the students beginning with and use student , and ; 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).
 - How would you conduct the activity?
Sample response: Students will stand for as long as they can using whichever leg they choose in their stocking or bare feet with their eyes open. We will time them to the nearest second using stop watches 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.
 - What sample statistics will you calculate, and how you would display and analyze the data?
Sample response: The sample statistics will be the mean time (in seconds) standing on one leg for the sixth graders, seventh graders, and the corresponding MADs. 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.
 - What you would accept as evidence that there actually is a difference in how long sixth graders can stand on one leg 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 has better balance.

Closing (4 minutes)

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

- Sallee argued that the set _____ could not possibly be a random sample of ten numbers from _____ to _____ 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 _____ to _____ 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.*

Lesson Summary

In this lesson, you collected data on total number of words by selecting a random sample of children's books. You also observed that several different samples of the same size had some characteristics in common with each other and with the population. In the second activity, you designed a statistical study. First, you considered a statistical question. Then, you went through the statistical process beginning with the question, and then thinking about how to choose a random sample, how students would take part, what data you would collect to answer the question, and how you would display and analyze the data.

Exit Ticket (4 minutes)

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 will have 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 seventh graders at a particular middle school. Describe how you might select a random sample of students from a list of the students in this population.

Answers will vary. Number the students on the list from 1 to 100. Using a random-number generator, get different random numbers between 1 and 100, and then select the students corresponding to those numbers on the list. It would also be correct for a student to say that they would write the names on slips of paper, put them in a bag, mix them well, and then select 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. You might have them work in pairs to generate and record the numbers in the random samples to see whether the teacher’s method of collecting homework will turn out to be relatively “fair” for the 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 as 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 _____ students.

Sample response: You could assign each student a number from _____ to _____, generate five random numbers from _____ to _____ and choose the corresponding students.

b. Suppose every day for _____ 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 _____ numbers generated all together with each number from _____ to _____ having an equal chance of showing up each time, then each number should be in the overall set about _____ or _____ times. I generated _____ random samples of numbers from _____ to _____ and looked at how the numbers showed up. Every number from _____ to _____ showed up at least three times and most of the numbers showed up about _____ or _____ 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 _____ to _____ for the block of words and the random numbers _____ to _____ to choose a word in the block. Or, you could number all of the words from _____ to _____ and then generate random numbers between _____ and _____ 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 _____ top-selling children’s books from Example 2. She generated ten random numbers between _____ and _____ 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 _____ books. For example, if she gets the random numbers _____ or _____, she would have to generate new numbers because no books on the list had either _____ or _____ 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, 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 _____ 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 _____. 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.