Random Sampling

ACTIVITY 37

Part-Time Jobs
Lesson 37-1 Surveys

Learning Target:

- Explain why random sampling is advantageous when conducting a survey.

LEARNING STRATEGIES: Close Reading, Questioning the Text, Role Play, Summarizing, Paraphrasing, Debriefing, Discussion Groups

Jorge is a member of the student government at a large school with over 2500 students. The student government would like to recommend that students with part-time jobs be permitted to get a class credit in business. Knowing that Jorge is a good statistics student, the student government asked him to estimate the proportion of students at the school who have part-time jobs.

1. What difficulties might Jorge encounter if he tries to ask every student about having a part-time job?

   Meeting 2500 students in person would take a very long time and it would be difficult to get to see them all.

Sometimes you may want to know some characteristic of a large population, such as the median income of households in your state or the proportion of students at a large school who have part-time jobs. Since it is often difficult or impossible to survey everyone in the population, you may wish to ***survey*** a ***sample*** of the population and infer conclusions from the sample about the population.

Jorge considers different methods for obtaining a sample.

2. Jorge is thinking about posting the question, "Do you have a part-time job?" on Facebook and collecting ***responses*** to his post. He knows that not everyone will reply, but he thinks he'll still get a large number of responses. Explain why, even if a large number of people replied (even as much as half of the student body), Jorge would be unwise to suppose that the proportion of people who posted that they have a part-time job is the same as the proportion of all students who have a part-time job.

   Sample answer: People who work part-time jobs may not have as much time to post on Facebook, and so they would be underrepresented among all the people who answered the question on Facebook.

My Notes

MATH TERMS

A **survey** is a study in which subjects are asked a question or series of questions.

An answer provided by a subject to a survey question is called a **response**.

MATH TERMS

A **sample** is part of a population of interest. Data are collected from the individuals in the sample.

Common Core State Standards for Activity 37

HSS-IC.A.1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

HSS-IC.B.3 Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

ACTIVITY 37

Investigative

Activity Standards Focus

In this activity, students investigate the process of sampling a population and identify possible bias in samples. Students apply these concepts and more in exploring experimental and observational studies.

Lesson 37-1

PLAN

Materials:

- table of random digits

Pacing: 1 class period

Chunking the Lesson

#1 #2–4 #5–6
#7 #8 #9
Check Your Understanding
Lesson Practice

TEACH

Bell-Ringer Activity

Pose a question of interest that applies to the population in your school's city, such as "Would you be in favor of ending Saturday mail delivery?" Ask students how they would determine the proportion of their city's population that would be in favor of the proposal in the question.

Introduction, 1 Shared Reading, Marking the Text, Summarizing, Interactive Word Wall After discussing responses to the Bell-Ringer Activity, students will recognize difficulties in surveying an entire population. Some students may consider surveying the school population in Item 1 a possibility, so stress the difficulties of surveying any large population. Make sure that students read the Math Terms signal box for *survey* before answering Item 1.

2 Marking the Text, Summarizing, Sharing and Responding Have students read the Math Terms signal box on *sample* prior to completing Item 2. Students will have varying experiences with social media, so be prepared for a lively discussion regarding bias in this situation.

3–4 Marking the Text, Summarizing, Sharing and Responding Have students answer Items 3 and 4 prior to having them read the Math Terms signal box on *bias*. Students will recognize that the samples chosen favor one response, allowing the students to develop an intuitive sense of the concept of bias.

Paragraphs Shared Reading, Marking the Text, Summarizing, Interactive Word Wall The concept of a *representative sample* may be initially difficult for some students to understand, so query the class for characteristics of representative samples—as compared to biased samples—of specific populations.

TEACHER to TEACHER

To build on the Bell-Ringer Activity, you may wish to address the concept of a representative sample for the students' home city. Consider different demographic categories—age, ethnicity, income, or other characteristics identified by your students—as you identify what may be a truly representative sample for the city's population.

TEACHER to TEACHER

Make sure that the concept of Simple Random Sample (SRS) is understood in terms of the probability of being chosen.

My Notes

3. Jorge is on the football team at his school and is thinking of asking everyone on the football team if they have a part-time job. Why might this give him a poor estimate of the actual proportion of students at his school with part-time jobs?

Sample answer: Students on the football team would be less likely than other students to have part-time jobs, since football practice takes up so much time. So, if Jorge assumed the football team represented all students in regards to having part-time jobs, he would probably end up underestimating the actual number of part-time jobs.

4. Jorge is considering standing beside an exit of the school one day after the last class is over and asking every student who passes by if he or she has a part-time job. How might this method produce an inaccurate estimate of the actual proportion of students at his school with part-time jobs?

Sample answer: Maybe students who have part-time jobs have to leave school early to go to work. If so, then by waiting for the last class to be over, Jorge will already have missed a lot of them, which would make them undercounted.

MATH TERMS

A sample shows **bias** if the composition of the sample favors certain outcomes.

MATH TERMS

A **simple random sample (SRS)** is a sample in which all members of a population have the same probability of being chosen for the sample.

Sampling can give very good results even if only a small sample of the population is surveyed, but it is critical that the sample be *representative* of the population with respect to the survey question. If the design of a sample favors one outcome over another, the sample is said to be ***biased***. Each of Jorge's sampling methods described in Items 2, 3, and 4 display bias, and your responses indicate how this bias was manifested in the results.

How can you be sure that a sample is representative of the population? Many methods of sampling people could produce samples of people that would tend to favor one type of survey response over another.

One way to avoid favoring some types of response over others is to sample people at random, with every person being equally likely to be chosen. Such a sample is called a ***simple random sample***, abbreviated **SRS**. A simple random sample is impartial because it does not favor anyone over anyone else. When a simple random sampling process is used to select members from a population, then everyone is as likely to be included in the sample as everyone else, and one person's inclusion in the sample has no effect on anyone else's inclusion in the sample.

Lesson 37-1
Surveys

My Notes

5. There was bias in each of the sampling methods described in Items 2, 3, and 4 of this activity. Describe how a simple random sample would have avoided such bias.

Sample answer: In each of the earlier sampling methods, certain common characteristics existed in each sample, introducing bias. These samples were quite possibly going to lead to underestimating the proportion of students who have part-time jobs.

If everyone is being chosen at random, then the sample would likely end up looking like a cross section of the student body.

6. Jorge has access to a full roster of all 2500 students at his school. One way to get a simple random sample of students would be for him to write the names of all 2500 students on index cards, put the cards into a large cardboard box and mix them up thoroughly, and then to draw out the desired number of names at random. What difficulties might Jorge encounter in his attempt to take a simple random sample in this way?

This method requires much time and effort. It would also be discouraging to write down so many names that would not be chosen to be part of the sample.

Another way to get a simple random sample is to number the list of students from 1 to 2500, and then use technology to randomly generate integers between 1 and 2500 until you have the desired sample size. For example, on TI-84 calculators, the following command generates a random integer between 1 and 2500:

randInt(1,2500)

Use the command to generate random integers that are matched up with the numbered list (ignoring repeated numbers) until you have identified all those names chosen to be in your sample.

7. Use your graphing calculator to choose 20 random integers between 1 and 100. Write the calculator syntax and your 20 random integers.

randInt(1,100); Answers will vary.

TECHNOLOGY TIP

To find the *randInt(* function on a TI-84 calculator, press the MATH button, scroll to *PRB*, and then choose *randInt(*.

ACTIVITY 37 Continued

5–6 Think-Pair-Share, Activating Prior Knowledge, Sharing and Responding For students to respond meaningfully to Item 5, they must have a good understanding of the concept of SRS. Discuss with students *how* to choose an SRS. Such discussions may provide an appropriate segue into Item 6. While students may appreciate this random approach of choosing the SRS, they will likely also see the difficulties with the method.

Technology Tip

There are various operating systems for graphing calculators (particularly among newer models) that may have different methods of entering the parameters for the commands to generate random numbers. Regardless of the operating system, the syntax is the same once the command is entered on into the home screen. If there is a different graphing calculator model available for use in your class, it is likely that analogous functions for generating random numbers exist for that calculator. For additional technology resources, visit SpringBoard Digital.

7 Create Representations, Sharing and Responding Students will get different groups of 20 numbers for Item 7 as they use the random integer function of their graphing calculators. However, there is one exception. If students are using calculators whose memory has just been reset, students *will* get exactly the same numbers. Each calculator uses the same initial value for its random number algorithm, making any such random number selection not truly random.

8 Close Reading, Marking the Text, Create Representations, Group Presentation This item is likely a student's initial exposure to random digit tables, and students may find it interesting that volumes of such tables exist (and were the only option prior to the introduction of modern technology). Carefully read the introductory paragraph so that students understand the procedure for identifying the random digits.

Teacher to Teacher

When students compare the efficiency of the random integer method using a graphing calculator versus using a random digits table they will recognize that using the random digits table will be more time consuming. This occurs not only because the numbers are generated by visual inspection, but because only 25% of the possible four-digit clusters will yield a number within our range of 0001 to 2500. Those outside of that range are ignored.

9 Shared Reading, Summarizing, Think-Pair-Share, Debriefing Students should focus on the method of generating the sample, not on the composition of the sample. When students share their individual responses to this item, monitor the discussion to ensure that the focus is on an unbiased method of selecting the sample.

Another method for generating random numbers from 1 to 2500 involves using a random digits table. Since the largest number in this range has four digits, you need to represent all numbers from 1 to 2500 as four-digit numbers. For example, 23 would be represented as 0023, and 798 would be represented as 0798. Then choose a line of the table at random and begin inspecting clusters of four digits. When a four-digit number matches one on Jorge's list, that name is selected as part of the sample. If a number is not on the list, then it is disregarded, as are repeated occurrences of the same number.

Random digits

Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915
115	61041	77684	94322	24709	73698	14526	31893	32592
116	14459	26056	31424	80371	65103	62253	50490	61181
117	38167	98532	62183	70632	23417	26185	41448	75532
118	73190	32533	04470	29669	84407	90785	65956	86382
119	95857	07118	87664	92099	58806	66979	98624	84826
120	35476	55972	39421	65850	04266	35435	43742	11937
121	71487	09984	29077	14863	61683	47052	62224	51025
122	13873	81598	95052	90908	73592	75186	87136	95761
123	54580	81507	27102	56027	55892	33063	41842	81868
124	71035	09001	43367	49497	72719	96758	27611	91596
125	96746	12149	37823	71868	18442	35119	62103	39244
126	96927	19931	36089	74192	77567	88741	48409	41903
127	43909	99477	25330	64359	40085	16925	85117	36071
128	15689	14227	06565	14374	13352	49367	81982	87209
129	36759	58984	68288	22913	18638	54303	00795	08727
130	69051	64817	87174	09517	84534	06489	87201	97245

8. Beginning at line 122 on the random digit table, identify the first five numbers that would correspond to names on Jorge's list. Compare this method to using the random integer generator on the graphing calculator.

The numbers 1387, 0529, 0908, 1369, and 0815 would be chosen. This method would be more time-consuming, as there are a large number of selections that do not match the range of numbers on the list using the random digits table.

9. Suppose that Jorge uses the random number generator on his graphing calculator to choose an SRS of 100 students at his school. He then surveys these students to determine whether they have part-time jobs. He notices that two of the 100 students in his sample are friends who both have part-time jobs working at the local auto garage. Jorge is worried about the over-inclusion of people with part-time jobs in his sample. Should he be concerned?

If we have a method of sampling people that is impartial, and using a random number generator on a calculator is impartial, then even if we happen to get a person here or there who we know has (or does not have) a part-time job, that does not mean there is a problem with the sampling method.

My Notes

Check Your Understanding

10. Describe a sampling method that Jorge might have thought about using that would have likely *overestimated* the fraction of students at his school who hold part-time jobs.

11. Priscilla is a junior at the same high school. She would like to survey a simple random sample of the 600 juniors in her class to determine preferences for class T-shirt designs. Describe how she could create a SRS of 50 students using a random digits table and using a graphing calculator.

LESSON 37-1 PRACTICE

Veronica wanted to know how many students in the sophomore class at her school learned a language other than English as their first language. There were 450 sophomores in the sophomore class, too many for Veronica to question each of them, so she prepared 50 questionnaires to distribute to some of the students in the class.

12. In Veronica's survey, what is the population? What is the question of interest? What is the sample?

13. Veronica chooses two classes near her homeroom in which to distribute the questionnaires. One has 25 students and is for first-year Spanish learners, and the other has 25 students and is for ELL (English language learner) students. Why is this selection of students not a simple random sample? What type of bias may exist in this sample?

14. Describe how Veronica could create a simple random sample of 50 students from the sophomore class in two different manners, without using technology.

15. Describe how Veronica could use technology to create a simple random sample.

Check Your Understanding

Debrief students' answers to these items to ensure that they understand the concept of bias in selecting a sample and methods of choosing a simple random sample (SRS) to avoid such bias.

Answers

10. Possible answer: If Jorge went to a business class and surveyed the students there, or if he went to a business-related club such as Future Business Leaders of America and surveyed those students, he would probably over-represent people with part-time jobs, since people in those groups are already interested in business at a younger age than most.

11. Priscilla would first have to assign each of the 600 students a number from 1 to 600. To determine the SRS using the random digits table, she would have to choose a random line from the table and use three digit clusters to find the first 50 clusters that matched numbers from 001 to 600. To determine the SRS using a graphing calculator, she would have to enter randInt (1,600) and choose the first 50 numbers, excluding repeated occurrences of the same number.

ASSESS

Students' answers to Lesson Practice problems will provide you with a formative assessment of their understanding of the lesson concepts and their ability to apply their learning.

See the Activity Practice for additional problems for this lesson. You may assign the problems here or use them as a culmination for the activity.

LESSON 37-1 PRACTICE

12. The population is the 450 students in the sophomore class. The question of interest is, "What proportion of the students in the sophomore class learned a language other than English as their first language?" The sample is the group of 50 students chosen to take the survey.

13. This sample is a *convenience sample,* chosen because it was convenient for Veronica to choose students near her homeroom. It does not represent a simple random sample because those students with classes near Veronica's homeroom had a greater probability of being chosen for the sample. The results would be biased because all students in the ELL class would likely have learned a language other than English first, while the Spanish class would likely have many students who learned English first.

14. Veronica could write names of students on identical slips of paper, put all names into a container, mix the slips of paper well, and randomly choose 50 slips of paper and record those names as her sample. Alternatively, she could assign each student a number from 001 to 450 (or 000 to 449) and use the random digits table, with clusters of three digits, to identify the first fifty numbers that match a student's number, ignoring repeats and numbers outside the range 001–450.

15. She could assign each student a number from 1 to 450 and use a random integer generator (randInt(*1, 450*) on the TI-84) and select the first 50 numbers (ignoring repeats).

ADAPT

Check students' answers to the Lesson Practice to ensure that they understand concepts associated with choosing a simple random sample by avoiding bias. If students struggle with these concepts, use the class as a population and address methods of choosing a sample of four students in both biased and unbiased manners.

Lesson 37-2

PLAN

Pacing: 1 class period

Chunking the Lesson

#1–3 #4–5
#6–7 #8
Check Your Understanding
Lesson Practice

TEACH

Bell-Ringer Activity

Have students read the Math Terms signal box and ask them to describe an experiment from their science class identifying these terms: *experiment, treatment, explanatory variable, response variable.*

Teacher to Teacher

The Bell-Ringer Activity may be a good opportunity to incorporate some interdisciplinary collaboration with the science department in your school. You may wish to present a description of an experiment with which students are familiar and have them identify the treatment, the explanatory variable, and the response variable.

1–3 Shared Reading, Summarizing, Interactive Word Wall, Activating Prior Knowledge Students may well be familiar with the concept of an experiment, but may use different vocabulary (e.g., independent variable for explanatory variable, dependent variable for response variable). Items 2 and 3 help students recognize the limits on inference from the results of an experiment.

My Notes

MATH TERMS

An **experiment** applies a *treatment* (a condition administered) to *experimental units* to observe an effect.

The **explanatory variable** is what is thought to be the cause of different outcomes in the experiment. In simple experiments, the explanatory variable is simply the presence or absence of the treatment.

The effect of the explanatory variable is called the **response variable**.

Learning Target:

- Explain why random allocation of treatments is critical to a good experiment.

LEARNING STRATEGIES: Close Reading, Questioning the Text, Role Play, Summarizing, Paraphrasing, Debriefing, Discussion Groups

For a science fair project, Zack and Matt wanted to estimate how the rebound of a tennis ball is changed if it is soaked in water overnight and then allowed to dry out. They would have liked to get a random sample of tennis balls on which to perform an experiment, but they realized such a sample was impossible. Instead, their physical education coach gave them 20 used tennis balls as their sample.

1. Consider the definition of ***experiment***. Identify the ***explanatory*** and ***response variables***, the experimental units, and the treatment in Zack and Matt's experiment.
   The experimental units are the 20 used tennis balls. The treatment is soaking the balls and letting them dry out. The explanatory variable is whether or not the balls are soaked. The response variable is the rebound of the ball.

2. Why was it impossible for Zack and Matt to get a random sample of all tennis balls?
   Many reasons exist for this impossibility. They have no idea how many tennis balls exist (a number which is constantly changing anyway). They cannot get a list of all tennis balls in existence, and they cannot access tennis balls at all locations for their project. Furthermore, most tennis balls are owned by people who would not relinquish possession.

3. Zack and Matt decided to perform their rebound experiment on the 20 tennis balls their gym coach gave them. What limits on their conclusions would exist by performing the experiment with these balls?
   Since their sample of tennis balls is not a simple random sample, we cannot be sure that they are representative of all tennis balls. In fact, we know they won't be representative because they are all old, used tennis balls. The consequence of this is that Zack and Matt will really only be able to estimate the effect on the rebound on old, used tennis balls. In fact, the balls will probably share other characteristics that will further limit the population of all tennis balls of which their sample is representative. Since there might even be common characteristics that Zack and Matt won't be able to identify, it is only possible for Zack and Matt to draw conclusions about the rebound effect on the 20 tennis balls they have.

Lesson 37-2
Experiments

My Notes

Zack and Matt planned to take their 20 tennis balls and put them into two groups of ten. The balls in one group would be soaked in water overnight and then allowed to dry out, while the others would just stay dry. They would then measure the rebound of all the tennis balls and compare the data for the two groups.

4. To determine which balls should be soaked and which would remain dry, Matt thought it best to use a ***completely randomized design***. Describe a process that would provide a completely randomized design for this experiment.

Assign each ball a number 1–20 and use technology to choose ten random integers from that range. Those numbers would identify the balls assigned to the treatment group. Alternatively, assign each ball a number from 01 to 20 and use clusters of two numbers from the random digits table.

MATH TERMS

A **completely randomized design** implies that all experimental units have the same probability of being selected for application of the treatment.

5. Zack noticed that ten of the balls their coach gave them were Wilson brand balls, and the other ten were Dunlop brand balls. He thought that they should let the ten Wilson balls be the ones soaked in water and the ten Dunlop balls be the ones that stayed dry. What reasons might Matt have to disagree with Zack?

It is possible that Dunlop balls and Wilson balls are different in their natural rebound. Given that possibility, if Zack and Matt were to find a difference in the rebound of the soaked balls and the dry balls, they wouldn't know if it was due to the effect of overnight soaking or due to the difference in brands.

6. Matt suggested that it would be better to group all the Wilson balls and randomly choose five to be soaked in water. Similarly, he would group all the Dunlop balls and randomly choose five to be soaked in water. Why is this ***randomized block design*** a good strategy?

When sampling from a population, it is desirable for the sample to be representative of the population. With two different types of balls serving as the block (to avoid the lurking variable of brand difference), such a selection would provide random assignment within each group, with five soaked and five dry balls of each type.

MATH TERMS

A **randomized block design** involves first grouping experimental units according to a common characteristic, and then using random assignment within each group.

ACTIVITY 37 Continued

4–5 Close Reading, Marking the Text, Think-Pair-Share, Quickwrite, Self Revision/Peer Revision Have students read the Math Terms signal box on *completely randomized design* before they attempt these items. Suggest appropriate reading strategies students can utilize to ensure understanding. After students have written responses to both items, have them discuss in their groups and revise their responses as needed.

6 Shared Reading, Summarizing Have students read the Math Terms signal box on *randomized block design* before reading Item 6. Students will have an intuitive sense for the reason behind such a design in this situation, especially after debriefing responses to Item 5.

7 Close Reading, Summarizing, Paraphrasing, Sharing and Responding, Group Presentation
Students may initially consider such a design effective, so emphasize that there may be a reason for criticizing the design. Once students are familiar with the item, allow for small group discussion and ask each group to present their response. Diversity of multiple correct results should be encouraged and celebrated.

8 Marking the Text, Think-Pair-Share, Debriefing, Graphic Organizer Have students read the Math Terms signal box on *matched pairs design* prior to reading and responding to Item 8. A web diagram, created by and shared with all groups, may enable students to compare all types of experimental design and to recognize that matched pairs design is a specific type of block design.

TEACHER to TEACHER

This lesson is intended to expose students to some select elements of experimental design. A thorough exploration of experimental design may be an entire course (or more), so some degree of student discomfort is expected.

Check Your Understanding

Debrief students' answers to this item to ensure that they understand the essential components of an experiment and the elements of effective experimental design.

Answers

9. The random processes are similar because they are both aimed at impartiality—Jorge wants people in his survey to be just as likely to have a part-time job as any student at his school, and Matt and Zack want the tennis balls in one treatment group to be just as bouncy (initially) as those in the other group. In both studies, randomness gives the researchers the impartiality they want. The way that they are different is that Jorge was randomly sampling people from a large population, but Matt and Zack already had the tennis balls they were going to use, and they were using a random process to divide the tennis balls into two treatment groups.

My Notes

MATH TERMS

A **matched pairs design** involves creating blocks that are pairs. In each pair, one unit is randomly assigned the treatment. Sometimes, both treatments may be applied, and the order of application is randomly assigned.

7. Matt and Zack thought about first measuring the rebound of all 20 dry tennis balls on a tennis court. Then they would soak all of the balls in water overnight and let them dry out. Finally, they would measure the rebound again on the tennis court. They could then see for each individual ball how much its rebound was changed by being soaked in water overnight. This strategy might be effective in accomplishing their research goal, but a critic of their experiment could point out that the change in rebound could be due to something other than having been soaked in water. Can you think of such a possible explanation?
Answers will vary, but responses must name something that the measurements on dry balls all had in common that they didn't share with the measurements made on the soaked balls. Since Matt and Zack planned to measure rebound outdoors on a tennis court, one possible response is that the outdoor temperature or humidity might be different on the two days. If they see a difference in rebound, it might be due to different outdoor conditions rather than on the soaking and drying treatment to which the balls were subjected.

8. Describe how a ***matched pairs design*** may alleviate the potential problems identified in Item 6. Why would it be impossible to have matched pairs in which the order of treatment is randomized?
By blocking the balls into matched pairs, the variability of conditions on two different days would be eliminated. It would be impossible to change the order of treatment because if a ball was soaked in water first, it would not be possible to then test the rebound of the ball as "unsoaked."

My Notes

Check Your Understanding

9. A random process was recommended to Jorge when he wanted to estimate how many students at his school hold part-time jobs. A random process was also recommended to Matt and Zack when they wanted to estimate the effect of waterlogging on tennis ball rebound. Explain how these two random processes are similar and how they are different.

LESSON 37-2 PRACTICE

A medical researcher wanted to determine the effect of a new drug on a specific type of cancer. He recruited 50 female and 50 male cancer patients, each diagnosed with this specific cancer that had progressed to the same stage. The anticipated effect of the drug was a 50% reduction in the size of the tumor within 4 weeks of treatment. All subjects would receive an injection, but some would receive the drug and others would receive a ***placebo***.

10. Describe a completely randomized experiment that the researcher could perform with these subjects.

11. Describe an experiment that would incorporate a block design and the purpose of the block design.

12. Describe an experiment that would incorporate a matched-pairs design and the purpose of the matched-pairs design.

13. A ***single-blind study*** is one in which either the person conducting the experiment or the subjects have knowledge of the treatment, but not both. A ***double-blind study*** is one in which neither the person conducting the experiment nor the subjects have knowledge of the treatment. Describe an advantage of a double-blind study in the cancer researcher's study.

ACADEMIC VOCABULARY

A ***placebo*** is a treatment applied to an experimental subject that appears to be the experimental treatment, but in fact is a treatment known to have no effect.

13. The psychological effect of a patient's knowledge of receipt of a drug may introduce another variable into the study. That is why a placebo is used to mask patient's knowledge of whether or not they have received the drug. Similarly, a researcher's knowledge of which patients are receiving the drug and which are not may influence the researcher's behavior toward patients. This behavior may reintroduce the variable that the placebo seeks to avoid.

ASSESS

Students' answers to Lesson Practice problems will provide you with a formative assessment of their understanding of the lesson concepts and their ability to apply their learning.

See the Activity Practice for additional problems for this lesson. You may assign the problems here or use them as a culmination for the activity.

LESSON 37-2 PRACTICE

10. The researcher could select from the 100 subjects a simple random sample of size 50 to receive the treatment, while the remaining 50 subjects receive a placebo. After four weeks, image all 100 patients and measure the size of tumors to compare to imaging done prior to the experiment.

11. Given the limited information about the subjects, the most obvious type of blocking would be to form two blocks, one female and one male. Choose a random sample of 25 from each group to receive the treatment, while the remaining patients receive the placebo. The purpose of such blocking may be to recognize different responses to the drug among male and female populations.

12. In this case, the best pairs to match would be those of the same gender. Select such pairs using a randomization process (e.g. select two random digits from 1 to 50 at a time), then randomly select and assign the treatment to one member of the pair. Compare images of both patients after four weeks to determine the effect of the drug.

ADAPT

Check students' answers to the Lesson Practice to ensure that they understand the basics of experiment and the importance of randomization in experimental design. If students struggle with these concepts, seek the assistance of a science colleague who may be able to provide additional context aligned to current science curricular topics.

Lesson 37-3

PLAN

Pacing: 1 class period

Chunking the Lesson

#1–2 #3 #4–5

Lesson Practice

TEACH

Bell-Ringer Activity

Ask students to consider a study that concluded that students who studied more than five hours a week had higher grade point averages than those who studied less than five hours a week. Is a treatment applied to the subjects in this study? Ask students to explain their response.

1–2 Shared Reading, Paraphrasing, Interactive Word Wall, Think-Pair-Share, Debriefing Make sure that students read the Math Terms and Math Tip signal boxes on *observational study* and *causation* before answering Items 1 and 2. Students should recognize that without a treatment imposed by the researcher, as is the case in an observational study, no causation can be established.

My Notes

MATH TERMS

In an **observational study,** a researcher observes and records measurements of variables of interest but does not impose a treatment.

MATH TIP

The results of an *observational study* can only imply an *association*. The results of an *experiment*, by imposing a condition, can imply *causation*.

Learning Target:

- Identify a confounding variable in an observational study.

LEARNING STRATEGIES: Close Reading, Questioning the Text, Role Play, Summarizing, Paraphrasing, Debriefing, Discussion Groups

Rebecca read an article online with the headline, "Survey shows that among employed Americans, people who text frequently tend to have lower-paying jobs than those who do not." Rebecca immediately sent a text message to her friend Sissy:

"OMG cc! txting makes u have less $$$! 2 bad 4 us!!!"

1. Why is the study referenced by the article that Rebecca read an ***observational study*** and not an experiment?

   In an experiment, researchers would have imposed a treatment on subjects. In this study, no treatment is imposed by the researcher.

2. While it is possible that Rebecca is correct, the statement she read didn't say that texting caused people to have lower incomes, only that people who frequently text have lower incomes. Give another possible explanation for why those who text frequently may have lower-paying jobs.

   Answers will vary, but responses should name a characteristic that plausibly could be associated with frequent texting and with lower-paying jobs. One answer may be that people who text frequently tend to be younger people, and younger people tend to have less work experience than older people, and therefore earn less money. The variable causing lower salaries could be less work experience, which happens to be associated with youth who text more frequently.

If a study reports an association between two factors, and the researcher merely observed the association between the two variables without applying a treatment, then the researcher cannot determine if one of the factors directly caused the other. A third unmeasured variable that may be associated with both of the measured variables is called a ***confounding variable***. This variable is "confounded with" one of the other two, and therefore is a potential explanation of the association.

Lesson 37-3
Observational Studies

My Notes

3. A 2010 study reported that people who take long vacations tend to live longer than people who do not. One possible explanation is that vacations are good for you, improving your health and increasing your lifespan. Describe another potential explanation for the association, and identify a confounding variable.

Answers will vary, but responses should identify a characteristic that could reasonably be associated with both taking long vacations and living longer. One possible answer may be that people who take long vacations are likely wealthier people, since poorer people cannot afford long vacations as easily. Wealthier people may also be able to afford better health care than poorer people. Therefore, a possible reason for the association between taking long vacations and living a long time is that wealth enables both phenomena.

A study published in the Journal of the American Medical Association showed that among a group of people who were hospitalized for bicycling accidents, the prevalence of elevated blood alcohol levels was significantly greater than it was among bicyclists who were stopped by the side of the road and who agreed to participate in the study by having their blood alcohol level measured.

4. Is there reason to believe that the actual proportion of (non-hospitalized) bicyclists who have elevated blood alcohol levels might be greater than what was estimated by recruiting bicyclists by the side of the road?

The bicyclists had to agree to participate in the study, and it is not at all unreasonable to think that people with elevated blood alcohol levels would be less likely to agree to participate in the study (since it required checking their blood alcohol) than would people who had not been drinking alcohol.

ACTIVITY 37 Continued

3 Shared Reading, Paraphrasing, Interactive Word Wall, Think-Pair-Share, Group Presentation, Debriefing Students will explore the concept of a confounding variable. Allow students the opportunity to develop a third variable to explain the correlation identified in this item, and then have groups share their responses. Ask each group for explanation of how their variable can connect the length of vacations and health.

4–5 Shared Reading, Paraphrasing, Think-Pair-Share, Group Presentation, Debriefing Each of these items addresses the idea that bias in sample selection can provide a confounding variable in an observational study.

4–5 (continued) Students should recognize that *voluntary* bias (the self-selection of the bicyclists who stopped to answer questions) will affect the results of the study.

ASSESS

Students' answers to Lesson Practice problems will provide you with a formative assessment of their understanding of the lesson concepts and their ability to apply their learning.

See the Activity Practice for additional problems for this lesson. You may assign the problems here or use them as a culmination for the activity.

LESSON 37-3 PRACTICE

6. The population is all people in the small town. The question of interest is whether increased sales of ice cream are associated with an increase in crime.

7. This is an observational study as no treatment is imposed. The researcher is not providing ice cream to the population; the researcher is merely gathering information.

8. An association between ice cream sales and crime rate does not mean that eating ice cream makes anyone more likely to commit crimes. Observational studies cannot determine causation; only experiments can determine causation. It is likely that crime and ice cream sales both tend to go up whenever the weather is warmer.

ADAPT

Check students' answers to the Lesson Practice to ensure that they understand the concept of observational study, its relation to causation, and the possibility of confounding variables. If students struggle with these concepts, find results of observational studies in current events from online or print news sources. Have students identify the correlation and possible confounding variables for these studies.

My Notes

5. The study included a caution about its conclusions, mentioning that the use of bicycle helmets was significantly more common among the people stopped by the side of the road than it was among those who were hospitalized. Why is that relevant to the conclusions one might draw from this study?

Among bicyclists, there is an apparent association between elevated blood alcohol and being hospitalized. It is very possible that bicycling while intoxicated leads you to be more reckless, and therefore causes you to be injured and hospitalized. But it is also possible that people who bicycle while intoxicated are the same kind of people who would refuse to wear a bicycle helmet. The accident rate might possibly be the same for intoxicated and non-intoxicated bicyclists, but the failure to wear a helmet among the intoxicated people would make them more likely to be injured and hospitalized than bicyclists who are not intoxicated.

LESSON 37-3 PRACTICE

The crime rate in a small town was shown to be significantly higher whenever ice cream sales were higher. A town councilman was baffled by this, but nevertheless advocated closing down ice cream parlors to lower crime.

6. Identify the population and the question of interest in this study.

7. Was the ice cream crime rate study an experiment or an observational study? Explain your decision.

8. Write a letter to the councilman explaining why his position on closing ice cream parlors may be based on faulty reasoning. Include a potential confounding variable in your letter.

ACTIVITY 37 PRACTICE

Write your answers on notebook paper.
Show your work.

Lesson 37-1

Following an online article about sunbathing posted on a website for teenagers, a poll asked the reader whether he or she regularly sunbathes. 81% of those who responded clicked on "Yes."

1. In this survey, what is question of interest?
2. What is the population that the survey seeks to represent?
3. What is the sample for this survey?
4. Is the sample representative of the population? Is it a simple random sample?
5. What bias may be apparent in the survey?
6. Describe how the bias in this survey may influence the results.

Lesson 37-2

A study was conducted to see whether drinking eight glasses of water daily would reduce the risk of catching a cold. Forty volunteers who participated in the study were randomly assigned to one of two groups. Those in one group were told not to change any aspect of their daily lives. Those in the other group were instructed to drink at least eight glasses of water daily. At the end of several months, the proportion of people who had caught a cold during that time period was significantly lower among those who drank at least eight glasses of water than among those who didn't. Since this was a randomized experiment, the researchers conducting the experiment thought that the only difference between the two groups of subjects was their water consumption, and, therefore, that drinking eight glasses of water daily can reduce your risk of getting a cold.

7. Why is this study an experiment as opposed to an observational study?
8. Describe a method that the researchers could have used to randomly assign members to each group.
9. What was the treatment in this experiment? What were the explanatory variable and the response variable?
10. Critics of the study identified something other than drinking water that made the two groups of subjects different from one another. What confounding variable may have influenced the results?
11. How could the experiment have been modified to eliminate the problem?

ACTIVITY PRACTICE

1. What is the proportion of teenagers who sunbathe regularly?
2. The population is teenagers, but likely not all teenagers in the world. Limitations with respect to language and exposure to technology exist, therefore limiting the population to English-speaking, relatively affluent teenagers.
3. The sample consists of those teenagers who responded to the survey question.
4. This is not a simple random sample as all teenagers in the population are not equally likely to be in the sample. Therefore, it is not representative of the population.
5. Those teenagers in the sample chose to be included in the sample, so there is voluntary bias evident in the sample.
6. Those who responded to the survey were probably not representative of teenagers in their sunbathing habits. They had finished reading an article about sunbathing and cared enough to respond to the survey, which means that sunbathing interests them. So the rate among a random sample of teenagers would likely be lower than 81%.
7. One part of the sample is told to drink eight glasses of water per day. This constitutes a treatment applied to some subjects, while others do not receive this treatment. Therefore, this study is an experiment.
8. A random selection process using names on identical slips of paper, a random digits table, or technology should be described.
9. The treatment is drinking eight glasses of water per day. The explanatory variable is whether or not 8 glasses of water were drunk each day. The response variable is the number of colds suffered during the several-month-long period of the study.
10. Subjects in the group that drank a lot of water must have also had to use the bathroom more frequently during the day, likely washing their hands more frequently than those subjects who were not drinking additional water. Frequent hand washing is known to reduce the chance of catching a cold.
11. The subjects in both groups could be instructed at the outset to be sure to wash their hands at least five times a day.

12. A population of parents with young infants would be identified and data would be collected about music exposure during the period of infancy. Additional data on intelligence would then be collected later in life for each of these subjects, and an association would be sought between classical music and intelligence.
13. Parents who are very concerned about the intelligence of their babies probably do lots of things to try to make them smarter—read to them, buy them brain games as presents, etc. If they thought that playing classical music for them would help, they would probably do that, too. So the confounding variable could be the level of concern parents have and the overall effort they exert to improve their children's brainpower. Listening to classical music might have nothing to do with babies becoming smarter children, but the association is there because those same parents did other things to increase their children's intelligence.
14. The question of interest is whether exposure to classical music as an infant causes an increase in intelligence later in life. The experimental units are the infants, and the treatment would be exposure to classical music.
15. Answers will vary, but responses should include random selection (using an appropriate method) of ten infants to receive the treatment, with the remainder not receiving the treatment. Specific details of the treatment must be identified: selection, daily duration, and volume of music. Additionally, length of time of application of the treatment—weeks or months—must be discussed. Then indicators of intelligence must be identified and measured at various intervals over a period of years.
16. Since Bruno obtained his population with the help of a local daycare, it is likely that his subjects were from a relatively limited population living in the same location. Due to common yet unmeasured influences on this population, his results would be limited to just that local population.

ADDITIONAL PRACTICE

If students need more practice on the concepts in this activity, see the Teacher Resources at SpringBoard Digital for additional practice problems.

Lesson 37-3

For many years it was believed that playing classical music for infants was associated with these same people being smarter as older children and adults. Several early studies seemed to support this idea.

12. Valentina read one such study that claimed to be an observational study, not an experiment. Explain how such a study would be designed to be an observational study.
13. Identify a likely confounding variable in such a study, and explain how it could be responsible for the apparent association between listening to classical music and being smarter.

Bruno considered the classical music theory as well, but thought that an experiment would be better suited to test this theory.

14. For such an experiment, identify the question of interest, the experimental units, and the treatment.
15. With the help of a local daycare center, Bruno was able to identify 20 parents with infants between the age of 1 month and 2 months. Describe, in detail, an experiment that would test the question of interest.

MATHEMATICAL PRACTICES
Reason Abstractly and Quantitatively

16. Suppose Bruno's experiment reveals a significant increase in intelligence for those children who listened to classical music. What limitations may exist in the interpretation of the results?

Normal Models, Surveys, and Experiments

RESEARCHING READERS

Embedded Assessment 1

Use after Activity 37

1. A researcher in psychology measured the reading skill, on a scale of 1 to 100 of a random sample of 16 fifth-graders at a school. The skill levels were as follows:

51	82	65	69	69	71	58	72
68	76	56	61	77	64	63	71

Assume that it is reasonable to model the distribution of reading skill levels of all fifth-graders at the school as approximately normal.

a. Estimate the proportion of fifth-graders at the school with reading skill levels at or below 55.

b. Estimate the proportion of fifth-graders at the school with reading skill levels between 60 and 70.

c. Estimate the reading skill level that a fifth-grader would have if his or her score was in the 95th percentile of reading skill levels for fifth-graders at the school.

d. Create a data display and explain how it supports or conflicts with the assumption of an approximately normal distribution for this data set.

2. A study was done in which volunteer subjects were divided into two groups at random. Subjects in the first group read realistic news stories about fictitious politicians and their political activities. Subjects in the second group read the same stories, but they also read stories about scandals involving the politicians. After several weeks, the subjects were asked to recall information about the politicians. The subjects in the second group recalled more about the activities of the politicians than did the subjects in the first group.

a. Identify the treatment, explanatory variable, and response variable in this experiment.

b. What might the researchers conclude as a result of this study?

c. Suppose that researchers used a block design in the experiment, placing subjects who regularly read news stories in one group and those who did not regularly read news stories in another group. Explain how this may have changed the conclusions that could be drawn from this study.

3. An online survey on a vegetable gardening website found that respondents who planted after April 1 had greater yields than those who planted before April 1.

a. Describe why this survey is an example of an observational study and not an experiment.

b. Brianna read the survey results and commented, "Planting after April 1 must cause vegetables yields to be greater." Describe the flaw in her statement.

c. Why might someone be skeptical about the results of such a survey?

Common Core State Standards for Embedded Assessment 1

HSS-ID.A.1	Represent data with plots on the real number line (dot plots, histograms, and box plots).
HSS-ID.A.2	Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
HSS-ID.A.3	Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).
HSS-ID.A.4	Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Recognize that there are data sets for which such a procedure is not appropriate. Use calculators, spreadsheets, and tables to estimate areas under the normal curve.

Embedded Assessment 1

Assessment Focus

- Properties of normal distributions
- Sampling techniques in studies
- Characteristics of experimental studies
- Characteristics of observational studies

Materials

- Standard Normal Table

Teacher to Teacher

Students will need their graphing calculators or the Standard Normal Table to complete Item 1. Be aware that some minor differences in numerical answers may occur as a result of these different methods. (Answers provided are found using a graphing calculator.)

Answer Key

1. Using mean = 67.0625 and standard deviation ≈ 8.1443,

a. 0.0693, or about 6.9%

b. 0.4479, or about 44.8%

c. The score must be 81; the actual skill level ≈ 80.46, and since scores are reported as integers, a score of 80 would be in the 94th percentile.

d. Displays may vary in type and style, but the distribution is bell-shaped and symmetric, and therefore, may be normal.

2. a. The treatment is reading stories about politicians, the explanatory variable is reading stories about scandals, and the response variable is recalling the politicians.

b. Researchers may conclude that reading stories about scandals involving politicians increases the likelihood that such politicians are remembered.

c. Whether or not subjects regularly read news stories may be a lurking variable and results may be a result of reading habits, not the type of stories.

3. a. The researchers did not impose a treatment on the subjects; they merely gathered data from a survey.

b. Observational studies can only infer correlation, not causation; therefore, Brianna's statement is incorrect.

c. Participants in the survey were self-selected as they chose to respond to the online survey. Therefore, voluntary bias may influence the results of the survey.