

ACTIVITY SHEET: SAMPLING

This activity sheet includes exercises to assess students' understanding of important concepts presented in the *Sampling* lesson.

Sampling

A data set is not provided for these exercises.

Exercise 1

Every freshman at Star Wars Academy is issued the same laptop when they start school. Since laptops are continually evolving, the following year's freshmen are all issued the same, but newer, laptops than the previous year's class. Each class from freshman to senior year has the same computer, but the computers vary across class year.

Students are allowed to keep their laptops after they graduate. Unfortunately, some of the laptop batteries fail before graduation. Consider a study to estimate the average lifetime of laptop batteries for Star Wars students. One hundred laptops are randomly sampled and the battery lifetime for each is recorded.

(a) What is the population of interest?

- A. All laptop batteries of Star Wars Academy students
- B. Laptop batteries for the sample of 100 students
- C. The laptop battery of a single student
- D. All laptop batteries of this type sold worldwide

(b) What is the parameter of interest?

- A. The lifetime of a laptop battery for a given student
- B. The number of laptop batteries belonging to Star Wars Academy students that fail within four years
- C. The average lifetime of the laptop batteries belonging to the sample of 100 students
- D. The average lifetime of all laptop batteries belonging to Star Wars Academy students

(c) True or False. Using a simple random sample is the most appropriate sampling method for obtaining 100 laptop batteries.

True

False

Exercise 2

At the end of every academic year, ABC college (with 2,000 students) has an on-campus Spelling Bee. The school is trying to determine what proportion of students choose to participate in the Spelling Bee. A simple random sample of 500 college students is selected from all students registered at ABC College. Of these 500 students, 100 of them choose to participate in the annual on-campus Spelling Bee.

(a) True or False. The proportion of all students at ABC College who would participate in the Spelling Bee is 20%.

True

False

(b) True or False. The proportion of all students at ABC College who would participate in the Spelling Bee is likely to be close to 20%, but not equal to 20%.

True

False

Exercise 3

In 2012, CBS News reported that “worrying about a math test can quite literally hurt.” Before we discuss the validity of this claim, read the short article found on the following website:

<http://www.cbsnews.com/news/fear-of-math-can-cause-real-pain/>

The article begins “Worrying about a math test can quite literally hurt.” This claim alone implies that anyone who worries about a math test could feel pain. Is this claim justified by the results of the article? Explain your answer. **Hint:** Consider the population to which the results generalize; that is, of what population is this sample representative? The sample is representative of what group of people?

Note: A stratified sample is not a simple random sample since it is not possible to obtain every subgroup of size n from the original population. For example, if you sample less than 100% from a given strata, then not every subset of size n can be part of the sample.

Exercise 4

Lorem Ipsum is dummy text that is used in printing. It has been an industry standard since the 1500's. For more information, see lipsum.com. A set of 200 words – 20 rows, each containing 10 words – from this filler text is printed below. We are interested in estimating the average number of letters per word, so let's take 3 simple random samples (SRS), each of size 5.

Minitab 19

- 1 Choose **Calc > Set Base**.
- 2 In **Set base of random digit generator to**, enter 1129. [Using a generator ensures that all students arrive at the same sample.] Click **OK**.
- 3 Choose **Calc > Random Data > Integer**.
- 4 In **Number of rows of data to generate**, enter 5.
- 5 In **Store in column(s)**, enter C1-C3.
- 6 In **Minimum value**, enter 1.
- 7 In **Maximum value**, enter 200.
- 8 Click **OK**.

Minitab 19 (Mac)

- 1 Choose **Calc > Set Base**.
- 2 In **Set base of random digit generator to**, enter 1129. [Using a generator ensures that all students arrive at the same sample.] Click **OK**.
- 3 Choose **Calc > Random Data**.
- 4 In **Number of rows of data to generate**, enter 5.
- 5 In **Store in column(s)**, enter C1-C3.
- 6 From **Distribution**, select **Integer**.
- 7 In **Minimum value**, enter 1.
- 8 In **Maximum value**, enter 200.
- 9 Click **OK**.

Minitab Express

- 1 Open the generate random data dialog box.
 - Mac: **Data > Generate Random Data**
 - PC: **DATA > Random Data**
- 2 In **Number of columns to generate**, enter 3.
- 3 In **Number of rows in each column**, enter 5.
- 4 From **Distribution**, select **Integer**.
- 5 In **Minimum value**, enter 1.
- 6 In **Maximum value**, enter 200.
- 7 In **Base for random number generator**, enter 1129.
- 8 Click **OK**.

Label the columns as follows:

| C1 | C2 | C3 | C4 | C5 | C6 |
|------|------|------|---------|---------|---------|
| SRS1 | SRS2 | SRS3 | Length1 | Length2 | Length3 |

Each number in columns C1-C3 represents a word in the passage below. Look up that word and record the corresponding number of letters in columns C4-C6.

Note: The numbers at the beginning of the line tell you which word begins that line. For example, 51 tells you that *et* is word number 51 and *libero* is word number 53.

| | |
|-----|------------------------------------------------------------------------------------|
| 1 | Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut aliquam |
| 11 | nunc vitae neque aliquam dictum. Suspendisse et odio felis. Vivamus |
| 21 | posuere imperdiet pretium. Donec vel lacus luctus massa blandit gravida. |
| 31 | Nunc in nisi non elit cursus blandit. Nam vulputate elit |
| 41 | ut tortor porttitor rhoncus. Morbi posuere mauris eget lectus lobortis |
| 51 | et suscipit libero egestas. Fusce est felis, lobortis vitae vestibulum |
| 61 | at, vestibulum et dolor. Integer iaculis neque et est semper |
| 71 | eu sollicitudin turpis posuere. Cum sociis natoque penatibus et magnis |
| 81 | dis parturient montes, nascetur ridiculus mus. Donec bibendum vestibulum tellus, |
| 91 | vitae dignissim lorem vehicula non. Vestibulum vitae elit a velit |
| 101 | volutpat auctor. Ut non erat massa, sed luctus orci. |
| 111 | aptent taciti sociosqu ad litora torquent per conubia nostra, per |
| 121 | inceptos himenaeos. Suspendisse potenti. Proin gravida interdum est, vel consequat |
| 131 | lacus aliquet ac. Vivamus id metus tellus. Ut nec mattis |
| 141 | erat. Quisque sed nulla nisl. In mi arcu, semper ac |
| 151 | tincidunt sit amet, faucibus id mi. Vestibulum sapien lectus, molestie |
| 161 | sed mollis sit amet, lacinia at purus. Duis iaculis lacus |
| 171 | non augue facilisis ornare. Mauris varius enim vel nisl lacinia |
| 181 | eu rutrum nulla suscipit. Morbi arcu nunc, fringilla vel facilisis |
| 191 | eget, ornare et odio. Etiam magna risus, varius in imperdiet |

(a) Compute the average word length for your three random samples.

Minitab 19 (Mac and PC)

- 1 Choose **Stat > Basic Statistics > Display Descriptive Statistics**.
- 2 In **Variables**, enter *Length1 Length2 Length3*.
- 3 Click **Statistics** and check **Mean**. Click **OK** in each dialog box.

Minitab Express

- 1 Open the descriptive statistics dialog box.
 - Mac: **Statistics > Summary Statistics > Descriptive Statistics**
 - PC: **STATISTICS > Describe > Descriptive Statistics**
- 2 From the drop-down list, select **Data are in more than one column**.
- 3 In **Variables**, enter *Length1 Length2 Length3*.
- 4 Click the **Statistics** tab, and then select **Mean**.
- 5 Click **OK**.

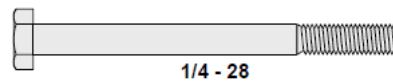
(b) Suppose we were to repeat this exercise by drawing samples of size 25 (you do not have to do this). Which averages would you trust more: those from part **(a)** or these new ones? Explain your answer.

Exercise 5

Quality engineers at an industrial fastener company, Tighten Up, are responsible for assessing product quality. One product of interest is a certain type of bolt produced in the tens of thousands and sold to auto manufacturers and their suppliers across the country. To assess the quality of the current shipment of 10,000 bolts being sent to My Motors, inspectors obtain 150 bolts from the shipment and subject them to various quality tests and measurements to determine thread pitch, which is the distance from one thread groove to the next, measured from crest to crest.

What is the population for the current shipment to My Motors?

- A. Tens of thousands of bolts
- B. Sample of 150 bolts
- C. Any shipment of 10,000 bolts
- D. Current shipment of 10,000 bolts



Exercise 6

For each of the following, select the best description of the sampling strategy employed in the given situation. If you are uncertain of your answer, provide an explanation for your best guess.

(a) Body Mass Index (BMI) is a useful measure of overall fitness in adults. In an effort to assess the overall fitness level of residents in a certain community, a researcher selects a sample of the first 20 citizens entering the local gym wearing white socks on a given day.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(b) There are 11,200 students enrolled at Cavalier College. The registrar draws a sample of 100 students to question about the online course registration process. He obtains an alphabetical listing of all 11,200 students, numbered from 1 to 11,200, and uses a random number generator to generate 100 random integers between 1 and 11,200. He then surveys the 100 students corresponding to those numbers.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(c) A pollster is interested in gauging how a given state's Governor would perform in his home state if he ran for President. The pollster randomly selects a sample of voters from a list of all eligible voting citizens in that state. Because nearly 75% of the state's voters claim a party affiliation (e.g. Democrat, Republican), the pollster decides to include approximately 25 non-affiliated voters for every 75 affiliated voters in his sample.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(d) A professor wants to know how his students performed on an exam from a total population of 100 students. Since he's eager to calculate sample statistics for the scores, he uses the first 20 exams that are turned in and begins his computations.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(e) Citrus trees are usually grown in orderly arrangements of rows to facilitate automated farming and harvesting practices. Suppose 1000 trees are grown down the sides of a small hill and are laid out uniformly about the hill in 40 rows of 25 trees each. To test the crop weight of fruit from these trees, a sample of 32 trees will be selected. Because growing conditions are different on the four sides of the hill, the hill should be divided into four quadrants, and 8 trees should be selected from each quadrant.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(f) A clinical psychologist is interested in investigating the relationship between excessive drinking and depression. In an effort to collect data about individuals' drinking habits and emotional state, she distributes a questionnaire to students in the library on Saturday night. To encourage responses, she offers chocolate bars to individuals who complete the survey.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(g) University personnel want to determine the proportion of students who choose to live on-campus versus off-campus. They examine a sample of 60 records. Since the percentage of freshmen at the university is 35%, sophomores 25%, juniors 15%, and seniors 25%, they separate the school records by these four class ranks. Finally, they randomly select 21 freshmen, 15 sophomore, 9 junior, and 15 senior records.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(h) A Starbucks coffee shop wants to identify the most popular beverages. On a Saturday morning, the manager surveys the first 20 customers who order drinks.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(i) On the 35th anniversary of the (alleged) death of Elvis Presley, a record company sponsored a national call-in survey. Listeners at thousands of radio stations were asked to call a 1-900 number (at a charge of \$2.50) to voice an opinion as to whether Elvis was dead. It turned out that 56% of the callers felt that Elvis was still alive.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

(j) A professor at Cape Fear College is interested in studying the effects of math anxiety on student math exam scores. The professor is currently teaching an advanced calculus course to college freshmen and a few sophomores. She decides to use her class as the study sample and passes out surveys during class.

Simple Random Sampling

Stratified Sampling

Non-random Sampling

Exercise 7

A researcher is studying the age of U.S. freshmen. She randomly samples 250 students from all U.S. colleges. The average age of the college students in her sample is an example of a:

A. Statistic

B. Parameter

C. Median

D. Population

E. None of these

Exercise 8

In order to capture the opinions of an entire college campus as to whether a certain person or group should be TIME magazine's "Person of the Year," an electronic survey is sent to the entire campus. The response rate is 10%.

(a) Give one example of why the population the "entire college campus" is ambiguous. That is, what is at least one questionable aspect of this definition of the population?

(b) Non-response bias may be affecting the survey results. Which one of these methods of sampling would most likely result in non-response bias?

A. Simple Random Sample

B. Stratified Sample

C. Voluntary Sample

Exercise 9

Email spam filters are based on statistical analysis. Consider a simple spam filter that obtains a sample of size n words from an email. It then compares the sample to a list of questionable words. If more than 75% of the sample appears in the list, the email is determined to be spam. Below are two sampling implementations.

(a) Sampling Method 1: Put all the words in a "bin" and randomly select n words. Or, number all of the words in the email and then use a random number generator to select words. Every word or every group of words of size n has an equally likely chance of being selected.

True or False. The above sampling method is an example of simple random sampling.

True or False

(b) Sampling Method 2: Separate the words in the email into two "bins", or strata, based on word length. Consider small words to consist of 3 or fewer letters and big words to consist of 4 or more letters. Pick a simple random sample from each bin corresponding to the proportion of small and big words. For example, if 40% of the words in the email are small, then randomly choose $0.4 \cdot n$ of the small words and $0.6 \cdot n$ of the big words.

True or False. The above sampling method is an example of stratified sampling.

True or False

Exercise 10

Ann Landers, who wrote a popular daily newspaper advice column, once asked her readers “If you had it to do over again, would you have children?” Of the more than 10,000 readers who responded, 70% said no. What does this show?

- A. The survey results are meaningless because we are not provided with information about her readership.
- B. The survey results would have been meaningful if she had picked a random sample of size 1,000 (from the 10,000 respondents) in which to draw a conclusion.
- C. The survey results are meaningless due to voluntary response bias.
- D. The survey results would have been meaningful if she had used control (people with only one child) and treatment (people with more than one child) groups.
- E. The survey results are meaningful because the readers who responded were parents who had children and the sample size was large enough for us to draw a valid conclusion.

Exercise 11

Every NBA basketball team has 12 players. A sample of players is to be chosen as follows. Each team will be asked to place all 12 players’ names into a hat and randomly draw 2 names. The 2 names from each team will be combined to make up the sample. Will this method result in a simple random sample of basketball players?

Assume that every National Basketball Association (NBA) team has its maximum allowed players of 15. Currently there are 30 NBA teams, resulting in 450 total NBA players. We want to make an All-Star team composed of 2 players from each team.

The All-Star team of 60 players (2 from each team) is to be chosen as follows:

- Each team will be asked to place all 15 players’ names into a hat and randomly draw 2 names.
- The 2 names from each team will be combined to make up the All-Star team.

Will this selection method result in a simple random sample of NBA players on the All-Star team?

- A. Yes. Each team is equally represented, and each player has the same chance of being selected for the All-Star team.
- B. Yes. This is an example of stratified sampling, which is a special case of simple random sampling.
- C. No. Not every possible combination of NBA players of the 450 total player has the same chance of being selected.

Exercise 12

To survey the opinions of baseball bleacher fans at Wrigley Field, a surveyor plans to select every one-hundredth fan who enters the bleachers one afternoon. Will this result in a simple random sample of fans who sit in the stadium's bleachers?

- A. Yes, because each bleacher fan has the same chance of being selected.
- B. Yes, but only if there is a single entrance to the bleachers.
- C. Yes, because the 99 out of 100 bleacher fans who are not selected will form a control group.
- D. Yes, because this is a special case of simple random sampling.
- E. No, because not every sample of any given size n has an equal chance of being selected.

Exercise 13

Read the following sampling design scenarios. What fault do they all have in common?

I. The Wall Street Journal plans to make a prediction for a presidential election based on a survey of its readers.

II. A radio talk show asks people to phone in their views on whether the United States should pay off its huge debt to the United Nations.

III. A police detective, interested in determining the extent of drug use by teenagers, randomly selects a sample of high school students and interviews each one about any illegal drug use by the student during the past year.

A. All of the designs exhibit improper use of stratification.

B. All of the designs have errors that can lead to strong bias.

C. All of the designs confuse association (or correlation) with cause and effect. For example, reading the WSJ does not cause you to vote in a certain way (e.g. conservative).

Exercise 14

A state auditor is given an assignment to choose and audit 26 companies. She lists all companies whose name begins with A, assigns each a number, and uses a random number generator to select one of the numbers, and thus one company. She proceeds to use the same procedure for each letter of the alphabet and then combines the 26 results into a group for auditing. Which of the following statements are true?

I. Her procedure makes use of chance.

II. Her procedure results in a simple random sample.

III. Each company has an equal probability of being selected.

A. I only B. I and II C. I and III D. I, II, and III E. None of the above

Exercise 15

A researcher planning a survey of heads of households in a particular state has census lists for each of the 23 counties in that state. He plans to obtain a random sample from each of the counties rather than grouping all census lists together and obtaining a sample from the entire group. Which of the following statements about the resulting stratified sample are true?

I. It is not a simple random sample.

II. It will help to avoid samples that do not represent the population.

III. It provides comparative information (e.g. mean salary of County X vs County Y) that a simple random sample would not provide.

A. I only B. I and II C. I and III D. I, II, and III E. None of the above

Exercise 16

To conduct a survey of long-distance calling patterns, a researcher opens a telephone book to a random page, closes his eyes, puts his finger down on the page, and then reads off the next 50 names. Which of the following statements about this scenario are true?

- I. The survey design incorporates chance.
- II. The procedure results in a simple random sample.
- III. The procedure could easily result in selection bias.

A. I only B. I and II C. I and III D. I, II, and III E. None of the above