Minitab ▶®

Sampling

A data set is not provided for these exercises.

Exercise 1

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

Solution: Parameters are calculated using population data. Statistics are computed using samples.

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

True	**False**

Minitab ▶®

Solution: Since the laptops and battery lifetimes may differ between class ranks, we want to ensure that laptops from each class are represented in the sample. Therefore, a stratified sampling scheme is most appropriate.

Exercise 2

A simple random sample of 500 college students is selected from all students registered at ABC College, and of these 500 students, 100 of them participate in the annual on-campus Spelling Bee.

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

<div align="center">

True   **False**

</div>

Solution: The sample may be representative of the population, but without conducting a census, it is not possible to state that the statistic exactly equals the parameter of interest.

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

<div align="center">

True   False

</div>

Solution: Since a random sample was taken, the statistic is likely to be close to the parameter of interest because the sample should be representative of the population.

Exercise 3

In 2012, CBS News reported that "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.

Solution: The patients that felt pain were math-anxious people. Therefore, the results can only generalize to people who have math anxiety, not anyone who happens to worry about one exam in particular. Therefore, the claim is too extreme. The study results suggest that those people who suffer from math anxiety could feel pain about an upcoming exam.

Exercise 4

(a) Take 3 simple random samples (SRS), each of size 5 and compute the average word length for your three random samples.

Solution: Your Minitab worksheet and descriptive statistics should look as follows:

↓	C1	C2	C3	C4	C5	C6
	SRS1	SRS2	SRS3	Length1	Length2	Length3
1	16	15	172	11	6	5
2	38	69	47	3	3	6
3	9	125	32	2	5	2
4	146	63	7	2	2	10
5	138	197	168	2	5	4

Descriptive Statistics: Length1, Length2, Length3

```
Variable    Mean
Length1     4.00
Length2     4.200
Length3     5.40
```

(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.

Solution: Larger sample sizes should result in a more representative sample, with smaller variability. Therefore, we should trust the ones from a larger sample, assuming it was collected appropriately.

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 samples 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 or not Elvis is really 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   Yes – measurement on a sample, $\bar{x}$ = average age

B. Parameter   No – measurement on the entire population

C. Median   No – the average age is not a median

D. Population   No – the population is college students in IN

E. None of these

Exercise 8

In order to capture the opinions of an entire college campus as to whether or not Steve Jobs should have been TIME magazine's "person of the year," an electronic survey is sent to the entire campus.

(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?

Solution: What does the "entire college campus" entail? Is staff included? Are alumni included? Is cafeteria staff (possibly employees of a food service and not the campus) included? In other words, "entire college campus" is not clearly defined.

(b) What type of sampling method may be creating bias in this case?

A. Simple Random Sample

B. Stratified Sample

C. Voluntary Sample

D. Convenience Sample

E. Haphazard Sample

Solution: Voluntary Sample – It's likely only people with a strong opinion will respond. This creates non-response bias.

Exercise 9

Consider a simple spam filter that obtains a sample of size *n* words from an email. 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*n of the small words and 0.6*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 is meaningless because of voluntary response bias.

Solution: *Voluntary response bias means only people with strong opinions about the topic will respond. It provides meaningless information for that reason.*

B. No meaningful conclusion is possible without knowing something more about the characteristics of her readers.

There is no meaningful conclusion because of voluntary response bias. It doesn't make a difference what the characteristic of her readers is if the sample data is collected through voluntary response.

C. The survey would have been more meaningful if she had picked a random sample of the 10,000 readers who responded.

The sample is biased to begin with and sampling from it to make decisions is still meaningless.

D. The survey would have been more meaningful if she had used a control group.

It's not possible to have a control group when the responses are voluntary.

E. This was a legitimate sample, randomly drawn from her readers and of sufficient size to allow the conclusion that most of her readers who are parents would have second thoughts about having children

The survey results weren't randomly drawn; they came from voluntary responders.

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?

A. Yes, because each player has the same chance of being selected.

Every possible sample of size n does not have the same probability of being chosen. For example, it is impossible for all of the team members of the Cleveland Cavaliers to be chosen.

B. Yes, because each team is equally represented.

Even though each team IS equally represented, this method will not produce a SRS since not every possible sample of size n has the same probability of being chosen.

C. Yes, because this is an example of stratified sampling, which is a special case of simple random sampling.

This is an example of stratified sampling, but stratified sampling is not a special case of simple random sampling.

D. No, because the teams are not chosen randomly.

Although the teams are not chosen randomly, what makes this not a SRS is the fact that every sample of size n does not have the same probability of being chosen.

E. No, because not every possible combination of players has the same chance of being selected.

Solution: *Correct! For example, it is impossible for all of the team members of the Cleveland Cavaliers to be chosen.*

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.

Each bleacher fan does not have the same chance of being selected if every one-hundredth fan is chosen. Sampling in the given manner will not result in a SRS.

B. Yes, but only if there is a single entrance to the bleachers.

Each bleacher fan does not have the same chance of being selected if every one-hundredth fan is chosen, even if all of the fans are coming through the same entrance.

C. Yes, because the 99 out of 100 bleacher fans who are not selected will form a control group.

There are no control groups with surveys.

D. Yes, because this is an example of systematic sampling, which is a special case of simple random sampling.

This is an example of systematic sampling, but this type of sampling is not a SRS.

E. No, because not every sample of the intended size has an equal chance of being selected.

Solution: *Correct!*

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.

This has voluntary response bias, so only the readers with strong opinions will care enough to respond.

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.

This has voluntary response bias, so only the readers with strong opinions will care enough to respond.

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.

This is an example of response bias, since no student will admit face-to-face to a cop that he or she is using illegal drugs.

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

None of the designs use stratification.

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

Solution: *Correct – each has some type of response bias.*

C. All of the designs confuse association with cause and effect.

Answer doesn't make sense.

D. None of the designs satisfactorily control for sampling error.

Although a true answer, it's not the major fault of the designs. The errors in the designs can definitely lead to bias.

E. None of the designs make use of chance in selecting a sample.

Although a true answer, it's not the major fault of the designs. The errors in the designs can definitely lead to bias.

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.

True. She does use a random number generator to pick one of the companies starting with a given letter.

II. Her procedure results in a simple random sample.

Not true. Not every sample of size n has an equally likely chance of being chosen. For example, all companies starting with the letter A cannot be chosen in the same sample.

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

Not true. If there are 20 companies whose name starts with the letter A, then any one of these companies has a 1/20 chance of being chosen. If there are 50 companies whose name starts with the letter B, then any one of these companies has a 1/50 chance of being chosen.

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.

This is true since a stratified sample is not a type of simple random sample.

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

True. This is one of the advantages of using a stratified sample compared to a SRS.

III. It provides comparative information that a simple random sample would not provide.

True. Perhaps some counties are poorer than others. Now each geographic region has a voice in the sample.

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.

True – at least the phone book is opened to a random page.

II. The procedure results in a simple random sample.

Not at all! Not every sample of size n has an equally likely chance of being chosen. In fact, once the researcher puts his finger down on the page, the next 50 names are selected. Any names in any other part of the phone book do not have a chance of being selected.

III. The procedure could easily result in selection bias.

Definitely! If a family and all of its relatives live in the same area, then they may all be listed one after the other in the phone book by last names. Perhaps the family and its relatives have the same opinions on long-distance calling. Also, some people are not listed in the phone book.

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