Collecting Data

11.3

Essential Question What are some considerations when undertaking a statistical study?

The goal of any statistical study is to collect data and then use the data to make a decision. Any decision you make using the results of a statistical study is only as reliable as the process used to obtain the data. If the process is flawed, then the resulting decision is questionable.

What You Will Learn

▶ Identify types of sampling methods in statistical studies.

▶ Recognize bias in sampling.

▶ Analyze methods of collecting data.

▶ Recognize bias in survey questions.

Analyzing Sampling Techniques

Determine whether each sample is representative of the population.. Explain your reasoning.

a. To determine the number of hours people exercise during a week, researchers use random-digit dialing and call 1500 people.

b. To determine how many text messages high school students send in a week, researchers post a survey on a website and receive 750 responses.

c. To determine how much money college students spend on clothes each semester, a researcher surveys 450 college students as they leave the university library.

d. To determine the quality of service customers receive, an airline sends an e-mail survey to each customer after the completion of a flight.

a. no; Random digit dialing will not necessarily give a random sample of any population. Many people will not answer the phone, possibly even because they are exercising, and some people do not have a phone, especially if they are younger in age.

b. no; Members of the population who do not have access to the Internet cannot be surveyed.

c. no; Students who are not in the library cannot be chosen for the survey.

d. no; Only those customers who choose to respond are surveyed.

Analyzing Survey Questions

Determine whether each survey question is biased. Explain your reasoning. If so, suggest an unbiased rewording of the question.

a. Does eating nutritious, whole-grain foods improve your health?

b. Do you ever attempt the dangerous activity of texting while driving?

c. How many hours do you sleep each night?

d. How can the mayor of your city improve his or her public image?

a. yes; It encourages a yes response; Sample answer: What effect does eating whole-grain food have on health?

b. yes; It encourages a no response; Sample answer: Do you ever text while driving?

b. no; It is a simple question to which an accurate response can be given.

b. d. yes; It implies the mayor needs to improve his or her public image; Sample answer: How would you describe the mayor's public image?
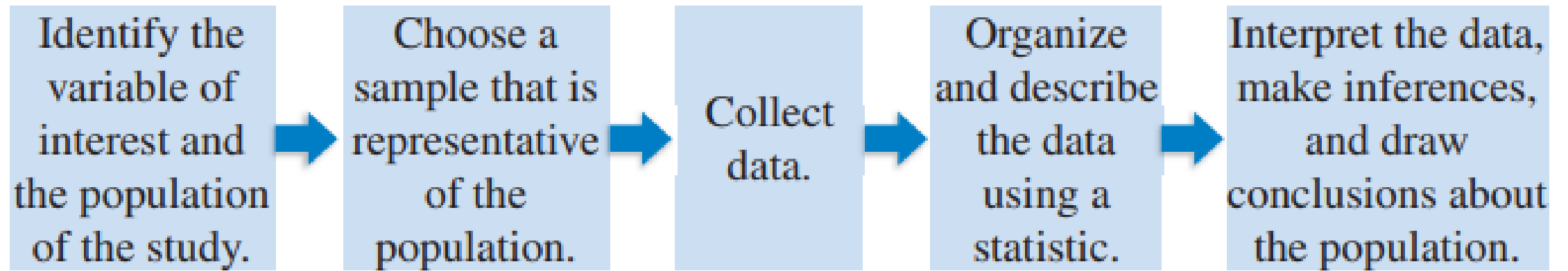
Analyzing Survey Randomness and Truthfulness

Discuss each potential problem in obtaining a random survey of a population. Include suggestions for overcoming the problem

a. The people selected might not be a random sample of the population.

b. The people selected might not be willing to participate in the survey.

c. The people selected might not be truthful when answering the question.

d. The people selected might not understand the survey question.

a. Sample answer: It is possible to select a group that would not be representative of the population as a whole; Choose a random sample.

b. Sample answer: The people choosing not to respond may represent a response different from those choosing to respond; Offer some type of reward to encourage participation

c. Sample answer: The results would not be representative of the true opinions of the population; Provide a method of completing the survey anonymously.

d. Sample answer: The people selected may not respond in a way that represents their true opinion; Rewrite the survey question so it is clear.

Identifying Sampling Methods in Statistical Studies

The steps in a typical statistical study are shown below.

Identify the variable of interest and the population of the study. ➡ Choose a sample that is representative of the population. ➡ Collect data. ➡ Organize and describe the data using a statistic. ➡ Interpret the data, make inferences, and draw conclusions about the population.

There are many different ways of sampling a population, but a *random sample* is preferred because it is most likely to be representative of a population. In a ==random sample==, each combination of members of a given sample size, n, of a population has an equal chance of being selected.

What are some ways of random sampling?

1. Number everyone in the population and use a random number generator
2. Picking names out of a hat

Ways to select a sample

PICK FROM THE FULL POPULATION

- **Self-selected**
  - o Members choose/volunteer to participate
  - o Example: online survey

- **Systematic**
  - o Use a rule/process to select members
  - o Example: select every other person

BREAK POPULATION INTO GROUPS, PICK FROM GROUPS

- **Convenience**
  - o Pick members that are easy to reach
  - o Example: select all my neighbors
- **Stratified**
  - o Pick key attributes, group by those attributes
  - o Groups are different from each other, each with different attribute
  - o Individuals w/in each group are similar/have same attribute
  - o Example: group by gender
- **Cluster**
  - o Divide population into groups, each group represents the population
  - o Each group is similar, looks like the others ... groups are homogeneous
  - o Individuals w/in each group are different, have different attributes
  - o Example: group by street from a larger community,

Core Concept

Types of Samples

For a **self-selected sample**, members of a population can volunteer to be in the sample.

For a **systematic sample**, a rule is used to select members of a population. For instance, selecting every other person.

For a **stratified sample**, a population is divided into smaller groups that share a similar characteristic. A sample is then randomly selected from each group.

For a **cluster sample**, a population is divided into groups, called *clusters*. All of the members in one or more of the clusters are selected.

For a **convenience sample**, only members of a population who are easy to reach are selected.

HOW DOES STRATIFIED DIFFER FROM CLUSTER SAMPLING?

Stratified Sample:
- Population broken into groups
- Groups are different from each other
  - Groups are diverse
- Individual member of a group are all the same in some sense
  - Individuals are homogeneous

	Groups	Individuals
Stratified	Different from each other	Homogeneous
Cluster	Homogeneous	Different from each other

Cluster Sample:
- Population broken into groups
- Groups all have the same make-up, are similar/same
  - Groups are homogeneous
- Individual members of a group are diverse
  - Individuals are diverse

EXAMPLE 1 **Identifying Types of Samples**

You want to determine whether students in your school like the new design of the school's website. Identify the type of sample described.

a. You list all of the students alphabetically and choose every sixth student.

b. You mail questionnaires and use only the questionnaires that are returned.

c. You ask all of the students in your algebra class.

d. You randomly select two students from each classroom, each a diff subject

SOLUTION

a. You are using a rule to select students. So, the sample is a *systematic* sample.

b. The students can choose whether to respond. So, the sample is a
self-selected sample.

c. You are selecting students who are readily available. So, the sample is a
convenience sample.

d. The students are divided into similar groups by their classrooms, and two students
are selected at random from each group. So, the sample is a *stratified* sample.

Recognizing Bias in Sampling

A **bias** is an error that results in a misrepresentation of a population.

In order to obtain reliable information and draw accurate conclusions about a population, it is important to select an **unbiased sample**. An **unbiased sample is representative** of the population that you want information about.

A sample that **over-represents** or **under-represents** part of the population is a **biased sample**. When a sample is biased, the data are invalid.

A random sample can help reduce the possibility of a biased sample.

Identifying Bias in Samples

Identify the type of sample and explain why the sample is biased.

a) A news organization asks its viewers to participate in an online poll about bullying.

SOLUTION

a. The viewers can choose whether to participate in the poll. So, the sample is a **self-selected** sample. The sample is biased because people who go online and respond to the poll most likely have a strong opinion on the subject of bullying.

b) A computer science teacher wants to know how students at a school most often access the Internet. The teacher asks students in one of the computer science classes.

SOLUTION

b. The teacher selects students who are readily available. So, the sample is a **convenience** sample. The sample is biased because other students in the school do not have an opportunity to be chosen.

Selecting an Unbiased Sample

You are a member of your school's yearbook committee. You want to poll members of the senior class to find out what the theme of the yearbook should be. There are 246 students in the senior class. Describe a method for selecting a random sample of 50 seniors to poll.

STUDY TIP

When you obtain a duplicate integer during the generation, ignore it and generate a new, unique integer as a replacement.

ONE POSSIBLE SOLUTION

Step 1 Make a list of all 246 seniors. Assign each senior a different integer from 1 to 246.

Step 2 Generate 50 unique random integers from 1 to 246 using the randInt feature of a graphing calculator.

Step 3 Choose the 50 students who correspond to the 50 integers you generated in Step 2.

The manager of a concert hall wants to know how often people in the community attend concerts. The manager asks 45 people standing in line for a rock concert how many concerts they attend per year. Identify the type of sample the manager is using and explain why the sample is biased.

SOLUTION:
convenience sample; People in the community who do not attend rock concerts do not have an opportunity to be chosen

Homework Textbook page 614; Exercises 5 – 12, 14 - 19