

you are comparing might differ in more ways than just the explanatory variable. For example, smoking status was a confounding variable in the candy study, because smokers were less likely to eat candy than nonsmokers were. Thus, you cannot tell whether the increased longevity of candy lovers is due to the candy or to not smoking. This is an issue even if you are not trying to generalize the results to a larger population.

Most importantly, you have begun to consider two key questions to ask of statistical studies:

- To what population can you reasonably generalize the results of a study?
- Can you reasonably draw a cause-and-effect connection between the explanatory variable and the response variable?

The answer to the first question depends on how the sample was selected. The answer to the second question depends on whether or not the explanatory variable was assigned to the observational units.

Some useful definitions to remember and habits to develop from this topic include

- The **population** is the entire collection of observational units (people or objects) of interest; the **sample** is the part of the population on which you gather data.
- A **parameter** is a number that describes a population; a **statistic** is a number that describes a sample. Notice that these are numbers, not people or objects as population and sample are.
- An **explanatory variable** is one whose effect you study; a **response variable** is the outcome you record.
- **Sampling bias** is the systematic tendency of a sampling method to overrepresent some parts of the population and underrepresent others.
- An **observational study** is one in which researchers record information passively, without attempting to impose the explanatory variable on the observational units.
- A **confounding variable** is one whose effect on the response variable cannot be distinguished from the effect of the explanatory variable. To properly identify a confounding variable, discuss how it is related to both the explanatory and the response variables.
- You cannot draw cause-and-effect conclusions from observational studies because other factors (confounding variables) might differ between the groups.

In the next two topics, you will learn how to answer two key scope-of-conclusion questions, generalizability and causation. You will also learn how to design studies well in the first place, so that you can make generalizations and/or draw causal conclusions. More specifically, you will discover how to select a sample from the population of interest (Topic 4) and how to assign subjects to treatment groups (Topic 5). Both of these design considerations involve the concept of randomness, but always keep in mind that these are two separate issues.

● ● ● Homework Activities

Activity 3-6: Elvis Presley and Alf Landon

3-1, 3-6, 16-5

The question of whether Elvis Presley faked his death in 1977 has also been asked on an Internet site called misterpoll.com, where anyone can post a poll question and get responses from whoever sees the site and chooses to respond.

- a. Is this sampling method unbiased for estimating the proportion of adults who believe that Elvis faked his death in 1977? If not, identify the direction in which you expect the sample to be biased. Explain your answer.

Of the 2032 responses posted by August 1, 2006, 50% responded that Elvis had faked his death.

- b. Is this number a parameter or a statistic? Explain.
 c. Do you want to rethink your answer to part a in light of this sample result? Explain.
 d. Identify the sample size. Would taking a larger sample in the same manner help reduce bias? Explain.

Activity 3-7: Student Data

1-1, 1-5, 2-7, 2-8, 3-7, 7-8

Consider your class of students as a sample from the population of all students at your school.

- a. For each of the following variables, indicate whether or not your class would likely constitute a *representative* sample from this population with regard to that variable. Explain your answers.
- Grade point average
 - Hours slept last night
 - Number of siblings
 - Whether you prefer to call or to send a text/instant message to friends
 - Political viewpoint
 - Gender
 - Total income that you will receive over your lifetime
- b. Think of two additional variables for which you believe your class is a representative sample from this population. Explain your choices.
 c. Think of two additional variables for which you believe your class is not a representative sample from this population. Explain your choices.

Activity 3-8: Generation M

3-8, 4-14, 13-6, 16-1, 16-3, 16-7, 18-1, 21-11, 21-12

The generation of children raised in the 1990s and 2000s has been dubbed “Generation M” because of the impact of media on their lives. The Kaiser Family Foundation conducted an extensive study to quantify how much time teenagers spend with various types of media. They gathered data from what they describe as a “nationally representative sample” of 2032 teenagers. Identify each of the following as a parameter or a statistic:

- a. The proportion of all American teenagers who have a television in their bedrooms
- b. The proportion of these 2032 teenagers who have a television in their bedrooms
- c. The proportion of boys in this sample who have a television in their bedrooms
- d. The average number of hours per week the teenagers in this study spent reading
- e. The average number of hours per week of recreational computer use among all American teenagers
- f. The proportion in this study who rated their level of contentedness as high
- g. The proportion who rated their contentedness as high among all American teenagers

- h. The average amount of weekly media exposure among those surveyed who rate their level of contentedness as high

Activity 3-9: Community Ages

Suppose that you want to estimate the average age of the residents in your community.

- a. Is this number a parameter or a statistic? Explain.
- b. Suppose that you take a sample of people attending a local church on Sunday morning. Would you expect this sampling method to be biased for estimating the average age of community residents? If so, in which direction? Explain.
- c. Suppose that you take a sample of drivers as they arrive at a local daycare facility. Would you expect this sampling method to be biased for estimating the average age of community residents? If so, in which direction? Explain.

Activity 3-10: Penny Thoughts

2-1, 3-10, 16-23

On July 15, 2004, the Harris Poll released the results of a survey asking whether people favored or opposed abolishing the penny. Of a national sample of 2136 adults, 59% opposed abolishing the penny. Explain what is wrong with each of the following statements:

- a. The population is the 2136 adults contacted by the Harris Poll.
- b. The sample is the 59% who oppose abolishing the penny.
- c. The variable is the 59% who oppose abolishing the penny.
- d. The observational units in this survey are pennies.
- e. The parameter consists of all American adults.
- f. The statistic is the average number of adult Americans who oppose abolishing the penny.

Activity 3-11: Class Engagement

Suppose that you want to investigate the conjecture that statistics students at your school are more actively engaged in class than calculus students. You regularly attend classes and rate the level of student engagement based on the number of questions asked by students and the number of students who offer responses to instructor questions. You do this for Professor Newton's 8 AM calculus class and for Professor Fisher's 11 AM statistics class.

- a. If the statistics class has a substantially higher level of student engagement than the calculus class, can you attribute the difference to the subject matter? Explain, based on how the study was conducted.
- b. Identify two confounding variables in this study, being sure to identify them clearly as variables.

Activity 3-12: Web Addiction

The August 23, 1999, issue of the *Tampa Tribune* reported on a study involving data volunteered by 17,251 users of the abcnews.com Web site. Users of the site were asked to respond to questions including whether or not they used the Internet to escape problems, tried unsuccessfully to cut back their Internet use, and found themselves

preoccupied with the Internet even when not using it. Almost 6% of those responding confessed to some sort of addiction to the Internet.

- a. Identify the population and sample used in this study.
- b. The number 6% applies to the sample and so is a statistic. Identify, in words, the corresponding parameter of interest in this study.
- c. Do you think that 6% is a reasonable estimate of the parameter you defined in part b? If not, indicate whether you think this estimate is too high or too low. Explain.

Activity 3-13: Alternative Medicine

In a spring 1994 issue, *Self* magazine reported that 84% of its readers who responded to a mail-in poll indicated that they had used a form of alternative medicine, such as acupuncture, homeopathy, or herbal remedies. Comment on whether or not this sample result is representative of the population of all adult Americans. Do you suspect that the sampling method is biased? If so, is the sample result likely to overestimate or underestimate the proportion of all adult Americans who have used alternative medicine? Explain your answers.

Activity 3-14: Courtroom Cameras

An article appearing in the October 4, 1994, issue of *The Harrisburg Evening-News* reported that Judge Lance Ito (who was trying the O.J. Simpson murder case) had received 812 letters from around the country on the subject of whether or not to ban cameras from the courtroom. Of these 812 letters, 800 expressed the opinion that cameras should be banned.

- a. What proportion of this sample supports a ban on cameras in the courtroom? Is this number a parameter or a statistic?
- b. Do you think that this sample represents the population of all American adults? Comment on the sampling method.

Activity 3-15: Junior Golfer Survey

In the "Golf Plus" section of its August 21, 2006, issue, *Sports Illustrated* presented the results of its Junior Golfer Survey. Participants in the survey were the 72 golfers, 36 boys and 36 girls aged 13–18, who had played in a recent American Junior Golf Association Tournament.

- a. Would you consider this to be a representative sample from the population of all American teenagers? Explain why or why not.
- b. One of the questions asked was, "If you could vote for President, would you be more likely to vote for a Democrat or a Republican?" Would you expect this sampling procedure to be biased with regard to this variable? If so, do you think the bias will overestimate support for Democrats or Republicans? Explain.
- c. In response to the question posed in part b, 13 responded "Democrat," compared to 48 for "Republican," 4 for "neither," and 7 for "don't know." Construct an appropriate graph of these responses, and describe the distribution. Also, comment on whether you want to rethink your answer to part b in light of these data.

- d. Two other questions asked of these golfers were
- How many hours are you online during a typical week?
 - Do you have your own cell phone?

Would you expect this sampling procedure to be biased with regard to this variable? If so, indicate the direction of the bias, and explain your answers.

Activity 3-16: Accumulating Frequent Flyer Miles

Frequent flyers can accumulate their miles to receive free flights and other benefits. This practice has become so widespread that consumers can now accumulate miles through means other than flying, for example by using a credit card affiliated with an airline's frequent flyer program. On September 28, 2005, the Web site msnbc.com published an article about how consumers can redeem frequent flyer miles from airlines. Within the article was a link to participate in an online poll that asked, "Do you use a credit card to accumulate airline miles?" Of the 1935 responses that had been received when we checked, 83% had answered "yes."

- a. Identify the observational units and variable. Also, classify the variable as categorical or quantitative.
- b. Is 83% a parameter or a statistic? Explain.
- c. Comment on whether this sampling method is unbiased for estimating the proportion of all American adults who use a credit card to accumulate airline miles. If you think this sampling method is biased, indicate whether you think it will overestimate or underestimate the population proportion. Explain your answer.
- d. Identify the sample size, and comment on whether, and if so how, it affects your answer to part c.

Activity 3-17: Foreign Language Study

3-17, 5-11

Studies often show that students who study a foreign language in high school tend to perform better on the verbal portion of the SAT exam than students who do not study a foreign language.

- a. Do you think these are observational studies? Explain.
- b. Is it legitimate to conclude that foreign language study causes an improvement in students' verbal abilities? If so, explain. If not, identify a confounding variable and explain why it is confounded with the explanatory variable. *Hint:* Describe how the confounding variable provides an alternative explanation for the observed difference in SAT scores between the two groups.

Activity 3-18: Smoking and Lung Cancer

3-18, 3-19

Many early studies on the relationship between smoking and lung cancer in the 1950s found that smokers were about 13 times more likely to die from lung cancer than nonsmokers. Still, people argued against a cause-and-effect conclusion, citing numerous possible confounding variables. Suppose a student argues that these studies are not convincing evidence because the researchers did not record the diets of the individuals. What more does the student need to say in order to provide a complete explanation of why diet is a potentially confounding variable?

Activity 3-19: Smoking and Lung Cancer

3-18, 3-19

One of the early studies of the relationship between smoking and lung cancer was conducted by Hammond and Horn (1958). They and an extensive set of volunteers tracked 187,766 men over a 44-month period, noting their smoking habits and whether or not they died of lung cancer.

- Identify the explanatory and response variables in this study.
- Is this an observational study? Explain.
- Would you have any qualms about generalizing the results of this study to the population of all adults? Explain.

Activity 3-20: A Nurse Accused

1-6, 3-20, 6-10, 25-23

Recall from Activity 1-6 that nurse Kristin Gilbert was accused of murdering hospital patients. Hospital records showed that of 257 eight-hour shifts on which Gilbert worked, a patient died on 40 of those shifts (15.6%). But of 1384 eight-hour shifts on which Gilbert did not work, a patient died on only 34 of those shifts (2.5%).

- Identify the observational units, explanatory variable, and response variable.
- Is this an observational study? Explain.
- Based on the type of study, can you legitimately conclude that Gilbert caused the higher death rate on her shifts?
- Put yourself in the role of the defense attorney. Suggest a confounding variable that offers an alternative explanation (other than Gilbert's guilt) for the observed relationship between her shifts and a higher death rate. Also explain, as if to a jury, how this confounding variable provides an alternative explanation.

Activity 3-21: Buckle Up!

2-4, 3-21, 8-5

In Activity 2-4, you found that states with tougher seatbelt laws tended to have a higher proportion of residents who complied with the law to wear a seatbelt.

- Is this an observational study? Explain how you know.
- Can you conclude from these data that the tougher seatbelt laws cause a higher proportion of residents to comply?
- Even if no cause-and-effect conclusion can be drawn, do the data suggest a potential benefit of tougher seatbelt laws? Explain.

Activity 3-22: Yoga and Middle-Aged Weight Gain

Researchers studied 15,500 healthy middle-aged adults, asking them to complete extensive surveys detailing their physical activities and weight gain/loss between the ages of 45 and 55. They found that those who practiced yoga had substantially lower weight gain than those who did not practice yoga.

- Identify and classify the explanatory and response variables in this study.
- Is this an observational study? Explain how you can tell.
- Does this type of study allow for drawing a cause-and-effect conclusion between practicing yoga and gaining less weight? Explain.
- Suggest a confounding variable that provides an alternative explanation for the observed relationship between practicing yoga and gaining less weight. (State this

as a variable, and make a plausible argument for how this confounding variable is related to both the explanatory and response variables.)

Activity 3-23: Pet Therapy

3-23, 5-13

Suppose that you want to study whether having a pet is beneficial to patients recovering from a heart attack. You obtain records of such patients from a local hospital and select a sample for your study. For each patient, you record whether he/she has a pet, and then you follow the patients for five years and see which patients survive.

- Is this an observational study? Explain.
- Identify the explanatory and response variables. Also classify them as categorical (also binary) or quantitative.
- If you find that pet owners survive at a higher rate than non-pet owners, can you conclude that pet ownership leads to therapeutic benefits for heart attack patients? Explain.

Activity 3-24: Winter Heart Attacks

Studies conducted in New York City and Boston have noted that more heart attacks occur in December and January than in all other months. Some people have tried to conclude that holiday stress and overindulgence cause the increased risk of heart attack.

- Identify a confounding variable whose effect on heart attack rate might be confounded with that of the month variable, providing an alternative explanation for the increased risk of heart attack in December and January.

A more recent study in Los Angeles revealed a similar finding.

- Identify a potential confounding variable in the Boston and New York City studies that was eliminated from consideration in the Los Angeles study.
- Identify another confounding variable that still pertains to the Los Angeles study.
Hint: You might think of a variable that would be eliminated by conducting the study in the southern hemisphere.

Activity 3-25: Pursuit of Happiness

2-16, 3-25, 13-17, 25-1, 25-2, 25-4

Reconsider Activity 2-16. The Pew Research Center conducted a survey in February of 2006 in which American adults were asked how happy they were with their lives. Among those with a family income of less than \$30,000, only 24% responded that they were very happy. This percentage increased to 33% in the \$30–\$75,000 income group, 36% in the \$75–\$100,000 income group, and 49% in the greater than \$100,000 income group. Do these study results establish a causal connection between income and happiness? Explain based on statistical principles.

Activity 3-26: Televisions, Computers, and Achievement

Researchers compared achievement scores of 348 Chicago third-graders, based on whether or not the child had a television in his/her bedroom and whether or not there was a computer in the home (Hancox, Milne, and Poulton, 2005). Researchers found that those children with a television in their bedrooms scored substantially lower on the mathematics portion of the test than those without a television in their bedrooms. Also, those with a computer in their homes scored substantially higher on the language arts portion of the test than those without a computer.

- a. Identify the two explanatory variables and two response variables. Also classify each as categorical (also binary) or quantitative.
- b. Is this an observational study? Explain.
- c. Can you conclude that a bedroom television is harmful to a child's academic development? Can you conclude that a home computer is helpful to a child's academic development? Explain.
- d. Suggest a potentially confounding variable in this study, and explain why it is confounding.
- e. Identify the sample in this study.
- f. To what population would you feel comfortable generalizing the results of this study? Explain.

Activity 3-27: Parking Meter Reliability

In 1998, for her sixth-grade science project, Ellie Lammer selected 50 parking meters along Solano and Shattuck Avenues in Berkeley, California. The *Los Angeles Times* reported that she put in one hour's worth of coins in each meter and used three stopwatches to see how long the meters actually lasted. She found that only three meters provided the correct amount of time. Would you be willing to generalize Ellie's results to all Berkeley parking meters? To all California parking meters? Explain.

Activity 3-28: Night Lights and Nearsightedness

Studies have shown children who sleep with a night light are more likely to become nearsighted than those who sleep in a dark room (Quinn, Shin, Maguire, and Stone, 1999).

- a. Is it legitimate to conclude that sleeping with a night light causes a higher rate of nearsightedness? Explain.
- b. Suppose that a student is asked to identify a confounding variable and responds, "Genetics, because nearsighted parents tend to have nearsighted children." Explain why this argument is incomplete and then complete it.