ii. What proportion of emergency room visits in the US involve sports-related injuries?

iii. Is there a difference in the average waiting time to be seen by an emergency room physician between male and female patients?

iv. What proportion of US residents have visited an emergency room within the past year?

1.64 Interviewing the Film Crew on Hollywood Movies There were 970 movies made in Hollywood between 2007 and 2013. Suppose that, for a documentary about Hollywood film crews, a random sample of 5 of these movies will be selected for in-depth interviews with the crew members. Assuming the movies are numbered 1 to 970, use a random number generator or table to select a random sample of five movies by number. Indicate which numbers were selected. (If you want to know which movies you selected, check out the dataset **HollywoodMovies**.)

1.65 Sampling Some Hardee's Restaurants The Hardee's Restaurant chain has about 1900 quick-serve restaurants in 30 US states and 9 countries.[33] Suppose that a member of the Hardee's administration wishes to visit six of these restaurants, randomly selected, to gather some first-hand data. Suppose the restaurants are numbered 1 to 1900. Use a random-number generator or table to select the numbers for 6 of the restaurants to be in the sample.

[33] *hardees.com/company/franchise.*

1.3 EXPERIMENTS AND OBSERVATIONAL STUDIES

Association and Causation

Three neighbors in a small town in northern New York State enjoy living in a climate that has four distinct seasons: warm summers, cold winters, and moderate temperatures in the spring and fall. They also share an interest in using data to help make decisions about questions they encounter at home and at work.

- Living in the first house is a professor at the local college. She's been looking at recent heating bills and comparing them to data on average outside temperature. Not surprisingly, when the temperature is lower, her heating bills tend to be much higher. She wonders, "It's going to be an especially cold winter; should I budget for higher heating costs?"
- Her neighbor is the plant manager for a large manufacturing plant. He's also been looking at heating data and has noticed that when the building's heating plant is used, there are more employees missing work due to back pain or colds and flu. He wonders, "Could emissions from the heating system be having adverse health effects on the workers?"
- The third neighbor is the local highway superintendent. He is looking at data on the amount of salt spread on the local roads and the number of auto accidents. (In northern climates, salt is spread on roads to help melt snow and ice and improve traction.) The data clearly show that weeks when lots of salt is used also tend to have more accidents. He wonders, "Should we cut down on the amount of salt we spread on the roads so that we have fewer accidents?"

Each of these situations involves a relationship between two variables. In each scenario, variations in one of the variables tend to occur in some regular way with changes in the other variable: lower temperatures go along with higher heating costs, more employees have health issues when there is more activity at the heating plant, and more salt goes with more accidents. When this occurs, we say there is an *association* between the two variables.

Association

Two variables are **associated** if values of one variable tend to be related to the values of the other variable.

The three neighbors share a desirable habit of using data to help make decisions, but they are not all doing so wisely. While colder outside temperatures probably force the professor's furnace to burn more fuel, do you think that using less salt on icy roads will make them safer? The key point is that an association between two variables, even a very strong one, does not imply that there is a *cause and effect* relationship between the two variables.

Causation

Two variables are **causally associated** if changing the value of one variable influences the value of the other variable.

The distinction between association and causation is subtle, but important. In a causal relationship, manipulating one of the variables tends to cause a change in the other. For example, we put more pressure on the gas pedal and a car goes faster. When an association is not causal, changing one of the variables will not produce a predictable change in the other. Causation often implies a particular direction, so colder outside temperatures might cause a furnace to use more fuel to keep the professor's house warm, but if she increases her heating costs by buying more expensive fuel, we should not expect the outdoor temperatures to fall!

Recall from Section 1.1 that values of an explanatory variable might help predict values of a response variable. These terms help us make the direction of a causal relationship more clear: We say changing the explanatory variable tends to cause the response variable to change. A causal statement (or any association statement) means that the relationship holds as an overall trend—not necessarily in every case.

Example 1.22

For each sentence discussing two variables, state whether the sentence implies no association between the variables, association without implying causation, or association with causation. If there is causation, indicate which variable is the explanatory variable and which is the response variable.

(a) Studies show that taking a practice exam increases your score on an exam.

(b) Families with many cars tend to also own many television sets.

(c) Sales are the same even with different levels of spending on advertising.

(d) Taking a low-dose aspirin a day reduces the risk of heart attacks.

(e) Goldfish who live in large ponds are usually larger than goldfish who live in small ponds.

(f) Putting a goldfish into a larger pond will cause it to grow larger.

Solution

(a) This sentence implies that, in general, taking a practice exam *causes* an increase in the exam grade. This is association with causation. The explanatory variable is whether or not a practice exam was taken and the response variable is the score on the exam.

(b) This sentence implies association, since we are told that one variable (number of TVs) tends to be higher when the other (number of cars) is higher. However, it does not imply causation since we do not expect that buying another television set will somehow cause us to own more cars, or that buying another car will somehow cause us to own more television sets! This is association without causation.

(c) Because sales don't vary in any systematic way as advertising varies, there is no association.

(d) This sentence indicates association with causation. In this case, the sentence makes clear that a daily low-dose aspirin *causes* heart attack risk to go down. The explanatory variable is taking aspirin and the response variable is heart attack risk.

(e) This sentence implies association, but it only states that larger fish tend to be in larger ponds, so it does not imply causation.

(f) This sentence implies association with causation. The explanatory variable is the size of the pond and the response variable is the size of the goldfish.

Contrast the sentences in Example 1.22 parts (e) and (f). Both sentences are correct, but one implies causation (moving to a larger pond makes the fish grow bigger) and one does not (bigger fish just happen to reside in larger ponds). Recognizing the distinction is important, since implying causation incorrectly is one of the most common mistakes in statistics. Try to get in the habit of noticing when a sentence implies causation and when it doesn't.

Many decisions are made based on whether or not an association is causal. For example, in the 1950s, people began to recognize that there was an association between smoking and lung cancer, but there was a debate that lasted for decades over whether smoking *causes* lung cancer. It is now generally agreed that smoking causes lung cancer, and this has led to a substantial decline in smoking rates in the US. The fact that smoking causes lung cancer does not mean that everyone who smokes will get lung cancer, but it does mean that people who smoke are more likely to get it (in fact, 10 to 20 times more likely[34]). Other important causal questions, such as whether cell phones cause cancer or whether global warming is causing an increase in extreme weather events, remain topics of research and debate. One of the goals of this section is to help you determine when a study can, and cannot, establish causality.

Confounding Variables

Why are some variables associated even when they have no cause and effect relationship in either direction? As the next example illustrates, the reason is often the effect of other variables.

DATA 1.5

Vehicles and Life Expectancy

The US government collects data from many sources on a yearly basis. For example, Table 1.5 shows the number of vehicles (in millions) registered in the US[35] and the average life expectancy (in years) of babies born[36] in the US every four years from 1970 to 2010. A more complete dataset with values for each of the years from 1970 through 2013 is stored in **LifeExpectancyVehicles**. If we plot the points in Table 1.5, we obtain the graph in Figure 1.2. (This graph is an example of a scatterplot, which we discuss in Chapter 2.) As we see in the table and the graph, these two variables are very strongly associated; the more vehicles that are registered, the longer people are expected to live. ■

[34] *http://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm#1.*

[35] Vehicle registrations from US Census Bureau, *http://www.census.gov/compendia/statab/cats/transportation.html.*

[36] Centers for Disease Control and Prevention, National Center for Health Statistics, Health Data Interactive, *www.cdc.gov/nchs/hdi.htm*, Accessed June 2015.

Table 1.5 ***Vehicle registrations (millions) and life expectancy***

Year	Vehicles	Life Expectancy
1970	108.4	70.8
1974	129.9	72.0
1978	148.4	73.5
1982	159.6	74.5
1986	175.7	74.7
1990	188.8	75.4
1994	198.0	75.7
1998	211.6	76.7
2002	229.6	77.3
2006	244.2	77.7
2010	242.1	78.7

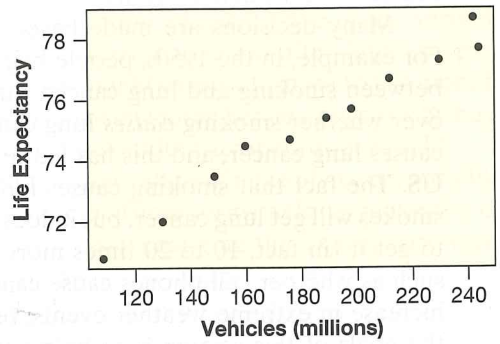


Figure 1.2 ***A strong association between vehicles and life expectancy***

There is a clear association between vehicle registrations and life expectancy. Is this a *causal* association? If so, which way might it go? Do people live longer because they have a car to drive? When people live longer, do they have time to buy more vehicles? Or is there something else driving this association?

Confounding Variable

A **confounding variable**, also known as a **confounding factor** or **lurking variable**,[37] is a third variable that is associated with both the explanatory variable and the response variable. A confounding variable can offer a plausible explanation for an association between two variables of interest.

Example 1.23

Describe a possible confounding variable in Data 1.5 about vehicle registrations and life expectancy.

Solution

One confounding variable is the year. As time goes along, the population grows so more vehicles are registered and improvements in medical care help people live longer. Both variables naturally tend to increase as the year increases and may have little direct relationship with each other. The years are an explanation for the association between vehicle registrations and life expectancy.

[37]Some statisticians distinguish between confounding variables and lurking variables. However, for simplicity in this book we treat them as synonymous.

When faced with a strong association such as that between vehicles and life expectancy, it can be tempting to immediately jump to conclusions of causality. However, it is important to stop and think about whether there are confounding variables which could be explaining the association instead.

Example 1.24

In 2008, the *Los Angeles Times* published a headline[38] that included "Hospitals... Riskier than a Casino in Event of Cardiac Arrest." The article, based on a study published in the *New England Journal of Medicine*,[39] states that the widespread availability of defibrillators and bystanders in public places like casinos leads to a higher survival rate than hospitals in the case of cardiac arrest.

(a) What are the primary variables of interest in this study? Which is the explanatory variable and which is the response variable?

(b) Give an example of one potential confounding variable in this study.

(c) If you are having a heart attack, would you go to a hospital or a casino?

Solution

(a) The two primary variables of interest are the place of cardiac arrest (explanatory) and whether or not the person survives (response).

(b) A confounding variable is the health of the person at the time of cardiac arrest. Older, frailer, sicker people are more likely to be in the hospital and also less likely to survive (not because they are in a hospital, but just because they are weaker to begin with). Someone in a casino is much more likely to be in better physical shape, and thus better able to survive a heart attack. Notice that the confounding variable (health of the person) influences *both* of the variables of interest: where the person might be and whether the person is likely to survive.

(c) If you are having a heart attack, you should go to a hospital! Even though casinos have a higher survival rate, this can be explained by the confounding variable, and we cannot conclude that being in a casino *causes* a higher survival rate. For a person of a given health status, it is probably safer to be in a hospital under the care of professionals.

Many seemingly surprising claims in the media (such as that hospitals are riskier than casinos) can be explained simply by the presence of a confounding variable. Knowing how and when to be on the lookout for confounding variables is essential for statistical literacy and for assessing any data-based claims.

Observational Studies vs Experiments

How can we establish (statistically) when an association represents a causal relationship? The key is in how the data are collected. If we want to study how the explanatory variable influences the response variable, we have to be able to control or specify the values of the explanatory variable to make sure it is not associated with any potential confounding variables.

Note that in data such as **LifeExpectancyVehicles** or the study of cardiac arrest we merely collect available data after the fact. We call data collected in this way, with no effort or ability to manipulate the variables of interest, an *observational study*.

[38]Maugh, T., "Study finds hospitals slow to defibrillate: Researchers say they're riskier than a casino in event of cardiac arrest," *Los Angeles Times*, January 3, 2008.

[39]Chan, P., Krumholz, H., Nichol, G., and Nallamothu, B., American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators, "Delayed Time to Defibrillation after In-Hospital Cardiac Arrest," *New England Journal of Medicine*, 2008; 358: 9–17.

With observational data we can never be certain that an apparent association is not due to some confounding variable, and thus the association is not evidence of a causal relationship.

The alternative is to intentionally control one or more of the explanatory variables when producing the data to see how the response variable changes. We call this method of data collection a *statistical experiment*. With a well-designed experiment, we can make conclusions about causation when we see a strong association, since the method for assigning the values of the explanatory variable(s) are not influenced by any confounding variables.

Observational Studies and Experiments

An **experiment** is a study in which the researcher actively controls one or more of the explanatory variables.

An **observational study** is a study in which the researcher does not actively control the value of any variable but simply observes the values as they naturally exist.

Example 1.25

Both studies below are designed to examine the effect of fertilizer on the yield of an apple orchard. Indicate whether each method of collecting the data is an experiment or an observational study.

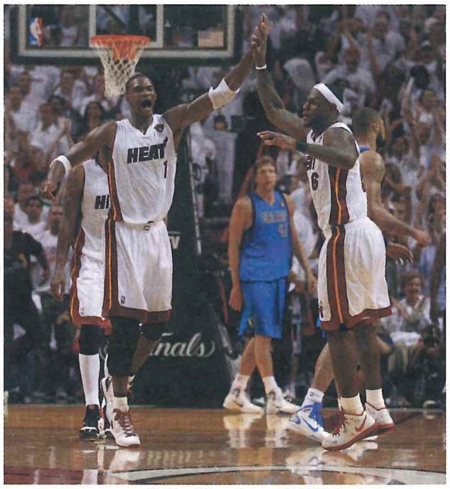
(a) Researchers find several different apple orchards and record the amount of fertilizer that had been used and the yield of the orchards.

(b) Researchers find several different apple orchards and assign different amounts of fertilizer to each orchard. They record the resulting yield from each.

Solution

(a) This is an observational study, since data were recorded after the fact and no variables were actively manipulated. Notice that there are many possible confounding variables that might be associated with both the amount of fertilizer used and the yield, such as the quality of soil.

(b) This is an experiment since the amount of fertilizer was *assigned* to different orchards. The researchers actively manipulated the assignment of the fertilizer variable, in order to determine the effect on the yield variable.



Al Diaz/Miami Herald/MCT via Getty Images

Do high fives help teams win?

Example 1.26

Basketball High Fives

In the 2011 NBA (National Basketball Association) finals, the Dallas Mavericks defeated the Miami Heat. One headline on NBC sports[40] stated, "Miami's real problem this series: Not enough high fives," citing a study[41] that found that teams exhibiting more "touching," such as high fives, early in the season had better performance later in the season. Is this study an experiment or an observational study? Does the study provide evidence that additional high fiving improves basketball performance?

Solution

The study is an observational study, because researchers did not manipulate or assign the number of high fives. The word "improves" implies causality, but because this was an observational study, confounding variables are likely and causality cannot be established. This study does not provide evidence that giving high fives improves basketball performance.

One possible confounding variable in Example 1.26 is how well a team gets along, which is likely to be associated both with the number of high fives and a team's performance. While we consider methods to account for some confounding variables later in this text, additional confounding variables may still exist. In an observational study, there is no way of guaranteeing that we haven't missed one.

Causation Caution

It is difficult to avoid confounding variables in observational studies. For this reason, observational studies can almost never be used to establish causality.

Randomized Experiments

In an experiment, the researcher controls the assignment of one or more variables. This power can allow the researcher to avoid confounding variables and identify causal relationships, if used correctly. But how can the researcher possibly avoid all potential confounding variables? The key is surprisingly simple: a *randomized experiment*. Just as randomness solved the problem of sampling bias, randomness can also solve the problem of confounding variables.

Randomized Experiment

In a **randomized experiment** the value of the explanatory variable for each unit is determined randomly, before the response variable is measured.

If a randomized experiment yields an association between the two variables, we can establish a causal relationship from the explanatory to the response variable.

[40] *http://probasketballtalk.nbcsports.com/2011/06/09/miami's-real-problem-this-series-not-enough-high-fives/.*

[41] Kraus, M., Huang, C., and Keltner, D., "Tactile communication, cooperation, and performance: An ethological study of the NBA," *Emotion*, 2010; 10(5): 745–749.

In a randomized experiment, we don't expect the explanatory variable to be associated with any other variables at the onset of the study, because its values were determined simply by random chance. If nothing is associated with the explanatory variable, then confounding variables do not exist! For this reason, if an association is found between the explanatory and response variables in a randomized experiment, we can conclude that it indeed is a causal association.

Recall from Section 1.2 that "random" does not mean haphazard. To ensure that the explanatory variable is determined by random chance alone, a formal randomization method (such as flipping a coin, dealing cards, drawing names out of a hat, or using technology) must be used.

Example 1.27

A college professor writes two final exams for her class of 50 students and would like to know if the two exams are similar in difficulty. On exam day, she gives Exam A to the first 25 students to enter the classroom and Exam B to the remaining 25 students.

(a) What is the explanatory variable? What is the response variable?

(b) Is this a randomized experiment? What might be a confounding variable?

Solution

(a) The explanatory variable is the exam the student took (A or B); the response variable is the exam score.

(b) No, this is not a randomized experiment. The exam students take is determined by when they enter the room, which is not random. Students that arrive especially early may be more motivated, and those that straggle in late may be less likely to perform well; time of arrival is a confounding variable.

Example 1.28

The following year the professor decides to do a truly randomized experiment. She prints the name of each of her students on an index card, shuffles the cards, and deals them into two piles. On exam day, she gives Exam A to the students with names dealt into one pile, and Exam B to the other pile. After grading, she observes that the students taking Exam B had much higher scores than those who took Exam A. Can we conclude that Exam B was easier?

Solution

Yes! This experiment provides evidence that Exam B was easier. Only random chance determined which student got which exam, so there is no reason to suspect confounding variables.

The key idea of Section 1.2 was that results from a sample can only be generalized to the population if the sampling units were selected *randomly* from the population. The key idea of this section is that causality can only be established if the values of the explanatory variable are *randomly* assigned to the units. Randomness is the essential ingredient in both cases, but the type of randomness should not be confused. In the first case we are randomly determining which units will be a part of the study. In the second case we are randomly determining which value of the explanatory variable will be assigned to each of the units already selected to be in our sample. Lack of randomness in either stage drastically influences the types of conclusions that can be made: Lack of randomness in sampling prevents generalizing to the population, lack of randomness in assigning the values of the explanatory variable prevents making causal conclusions. See Figure 1.3.

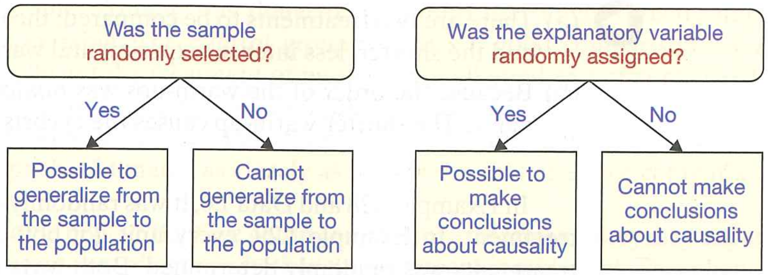


Figure 1.3 *Two fundamental questions about data collection*

DATA 1.6 **Physicians' Health Study**

Does anyone you know regularly take a low-dose aspirin? If so, it may be because of a randomized experiment conducted in the 1980s, the Physicians' Health Study.[42] The study recruited 22,071 male physicians and randomly assigned half of them to take an aspirin every other day for about five years and the other half to take a fake aspirin pill instead. They found that the physicians who took the real aspirin had 44% fewer heart attacks than those taking the fake aspirin. ■

The study in Data 1.6 is a randomized experiment because the researchers randomly determined which physicians received the real aspirin. The physicians themselves had no choice and in fact did not even know which pill they were taking. Because the physicians were split into two groups randomly, the only difference between the groups should be the aspirin. Therefore, we can conclude that the difference in heart attack rates must be *caused* by the aspirin. From this study we can conclude that regularly taking a low-dose aspirin reduces the risk of heart attack.

Many ideas of experimental design were originally developed for medical studies (such as Data 1.6) or agricultural experiments (like the fertilizer example of Example 1.25). For this reason, we often refer to values of the explanatory variable which the researcher controls as *treatments*. In Data 1.6, the treatments are the real aspirin and the fake aspirin.

Example 1.29

Warming Up

Warming up is a regular part of almost every athlete's pre-game routine, but the optimal amount to warm up is not always known. Cyclists typically have a very intense warm-up, and a study[43] in 2011 tests whether a shorter, less intense warm-up is better. Ten cyclists were recruited from the Calgary Track Cycling League and completed both a traditional intense warm-up and a shorter, less physically demanding, experimental warm-up. Each cyclist completed each warm-up at different times, and the order in which the warm-ups were performed was randomized. After each warm-up, performance was measured. The study found performance to be better after the shorter warm-up.

(a) What are the treatments?

(b) What conclusion can we draw from this study?

[42]The Steering Committee of the Physicians' Health Study Research Group, "Final report on the aspirin component of the ongoing Physicians' Health Study," *New England Journal of Medicine*, 1989; 321: 129–135.

[43]Tomaras, E. and Macintosh, B., "Less is More: Standard Warm-up Causes Fatigue and Less Warm-up Permits Greater Cycling Power Output," *Journal of Applied Physiology*, May 5, 2011.

Solution

(a) There are two treatments to be compared: the more intense traditional warm-up and the shorter, less intense, experimental warm-up.

(b) Because the order of the warm-ups was *randomized*, causal conclusions can be made. The shorter warm-up causes the cyclists to perform better.

In Example 1.28 and Data 1.6, it was randomly determined which units got which treatment. In Example 1.29, every unit got both treatments, but the *order* of the treatments was randomly determined. Both ways of randomization yield valid randomized experiments. The former is known as a *randomized comparative experiment* because two groups of units are compared. The latter is known as a *matched pairs experiment*, because each unit forms a pair of data values (one under each treatment), and comparisons are made within each pair. These are only two of many different ways to incorporate randomization into an experiment.

Two Types of Randomized Experiments

In a **randomized comparative experiment**, we randomly assign cases to different treatment groups and then compare results on the response variable(s).

In a **matched pairs experiment**, each case gets both treatments in random order (or cases get paired up in some other obvious way), and we examine individual differences in the response variable between the two treatments.

Example 1.30

Is the Dominant Hand Stronger?

We wish to run an experiment using 30 right-handed people to determine whether gripping strength in the dominant hand is greater than gripping strength in the other hand.

(a) Describe the experiment if we use a randomized comparative design.

(b) Describe the experiment if we use a matched pairs design.

(c) Which design makes more sense in this case?

Solution

(a) Using a randomized comparative design, we randomly divide the 30 people into two groups of 15 each. We measure gripping strength in the right hand for one of the groups and in the left hand for the other group, and compare results.

(b) In a matched pairs experiment, we measure the gripping strength in both hands for each of the 30 people. The data are "paired" because we compare the right- and left-handed gripping strength for each person, and examine the difference between the two values. We randomize the order in which participants use the hands: some (randomly determined) doing the right hand first and some the left hand first. Notice that all participants are doing both, unlike in the experiment described in part (a) with two distinct groups each assigned a different treatment.

(c) A matched pairs experiment makes sense here because hand-gripping strength can vary a great deal between different people and it makes sense to compare a person's right-hand strength to his or her own left-hand strength.

Control Groups, Placebos, and Blinding

The Physicians' Health Study illustrates many aspects of a well-designed experiment. The participants who did not take an aspirin pill are an example of a *control group*. Nothing was done to this group that might directly influence the response variable. The control group provides a good comparison for the group that actually got the treatment of interest. Not all good experiments need a control group. There is no control, for example, in Example 1.28 when testing to see if one exam is more difficult than the other. In all cases, however, procedures for handling the groups should match as closely as possible so that effective comparisons can be made.

If people believe they are getting an effective treatment, they may experience the desired effect regardless of whether the treatment is any good. This phenomenon is called the *placebo effect*. Although perhaps not intuitive, the placebo effect has been studied extensively and can be very powerful. A *placebo* is a fake pill or treatment, and placebos are often used to control for the placebo effect in experiments. The fake aspirin pill given to participants in the control group of the Physicians' Health Study is an example of a placebo.

Using a placebo is not helpful, however, if participants know they are not getting the real treatment. This is one of the reasons that *blinding* is so important. In a *single-blind* experiment, the participants are not told which group they are in. In a *double-blind* experiment, the participants are not told which group they are in *and* the people interacting with the participants and recording the results of the response variable also do not know who is in which group. The Physicians' Health Study was double-blind: The people taking the pills did not know whether they were taking an aspirin or a placebo and the doctors treating them and determining who had heart attacks also did not know.

DATA 1.7

Sham Knee Surgery

For people suffering from arthritis of the knee, arthroscopic surgery has been one possible treatment. In the mid-1990s, a study[44] was conducted in which 10 men with arthritic knees were scheduled for surgery. They were all treated exactly the same except for one key difference: only some of them actually had the surgery! Once each patient was in the operating room and anesthetized, the surgeon looked at a randomly generated code indicating whether he should do the full surgery or just make three small incisions in the knee and stitch up the patient to leave a scar. All patients received the same post-operative care, rehabilitation, and were later evaluated by staff who didn't know which treatment they had. The result? The men who got the sham knee surgery and the men who got the real knee surgery showed similar and indistinguishable levels of improvement. ■

Example 1.31

Discuss the experiment in Data 1.7. How is randomization used? Is there a placebo? Is the study double-blind? Why did the doctors make incisions in the knees of the men not getting the surgery?

Solution

Randomization was used to divide the men into groups, determining who got the real surgery and who didn't. The placebo was the fake surgery. Because the placebo surgery should match the real surgery as much as possible, those in the placebo group still received incisions and stitches. The men needed similar scars so that both the patients and the staff giving follow-up care were blind as to who actually had surgery done inside their knee. This made the study double-blind.

[44]Talbot, M., "The Placebo Prescription," *The New York Times*, January 9, 2000.

You may wonder whether data from only 10 patients are sufficient to make strong conclusions about the best treatment plan for arthritic knees. That would be a valid concern. In general, we would like to *replicate* each treatment on as many experimental units as is feasible. In many situations a small pilot study, such as the one described in Data 1.7, is used for initial guidance before undertaking a larger, more expensive experiment. In the case of the placebo knee surgery, a follow-up study with 180 patients produced similar results[45] – indicating that full knee surgery may not be needed for patients with this condition.

Example 1.32

Does an injection of caffeine help rats learn a maze faster? Design an experiment to investigate this question. Incorporate elements of a well-designed experiment.

Solution

We take the rats that are available for the study and *randomly* divide them into two groups. One group will get a shot of caffeine, while the other group will get a shot of saline solution (placebo). We have the rats run the maze and record their times. Don't tell the rats which group they are in! Ideally, all people who come in contact with the rats (the people giving the shots, the people recording the maze times, and so on) should not know which rats are in which group. This makes the study double-blind. Only the statistician analyzing the data will know which rats are in which group. (We describe here a randomized comparative experiment. A matched pairs experiment would also work, and in that case we would also use a placebo and blinding.)

Realities of Randomized Experiments

Randomization should always be used in designing an experiment. Blinding and the use of a placebo treatment should be used when appropriate and possible. However, there are often ethical considerations that preclude the use of an experiment in any form. For example, imagine designing an experiment to determine whether cell phones cause cancer or whether air pollution leads to adverse health consequences. It would not be appropriate to require people to wear a cell phone on their head for large amounts of time to see if they have higher cancer rates! Similarly, it would not be appropriate to require some people to live in areas with more polluted air. In situations such as these, observational studies can at least help us determine associations.

SECTION LEARNING GOALS

You should now have the understanding and skills to:

- Recognize that not every association implies causation
- Identify potential confounding variables in a study
- Distinguish between an observational study and a randomized experiment
- Recognize that only randomized experiments can lead to claims of causation
- Explain how and why placebos and blinding are used in experiments
- Distinguish between a randomized comparative experiment and a matched pairs experiment
- Design and implement a randomized experiment

[45] Moseley, J., et al., "A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee," *The New England Journal of Medicine*, 2002; 347: 81–88.

Exercises for Section 1.3

SKILL BUILDER 1

In Exercises 1.66 to 1.71, we give a headline that recently appeared online or in print. State whether the claim is one of association and causation, association only, or neither association nor causation.

1.66 Daily exercise improves mental performance.

1.67 Among college students, no link found between number of friends on social networking websites and size of the university.

1.68 Cell phone radiation leads to deaths in honey bees.

1.69 Wealthy people are more likely than other folks to lie, cheat, and steal.

1.70 Cat owners tend to be more educated than dog owners.

1.71 Want to lose weight? Eat more fiber!

SKILL BUILDER 2

Exercises 1.72 to 1.77 describe an association between two variables. Give a confounding variable that may help to account for this association.

1.72 More ice cream sales have been linked to more deaths by drowning.

1.73 The total amount of beef consumed and the total amount of pork consumed worldwide are closely related over the past 100 years.

1.74 People who own a yacht are more likely to buy a sports car.

1.75 Sales of toboggans tend to be higher when sales of mittens are higher.

1.76 Air pollution is higher in places with a higher proportion of paved ground relative to grassy ground.

1.77 People with shorter hair tend to be taller.

SKILL BUILDER 3

In Exercises 1.78 to 1.81, we describe data collection methods to answer a question of interest. Are we describing an experiment or an observational study?

1.78 To examine whether eating brown rice affects metabolism, we ask a random sample of people whether they eat brown rice and we also measure their metabolism rate.

1.79 To examine whether playing music in a store increases the amount customers spend, we randomly assign some stores to play music and some to stay silent and compare the average amount spent by customers.

1.80 To examine whether planting trees reduces air pollution, we find a sample of city blocks with similar levels of air pollution and we then plant trees in half of the blocks in the sample. After waiting an appropriate amount of time, we measure air pollution levels.

1.81 To examine whether farm-grown salmon contain more omega-3 oils if water is more acidic, we collect samples of salmon and water from multiple fish farms to see if the two variables are related.

REVISITING QUESTIONS FROM SECTION 1.1

Exercises 1.82 to 1.84 refer to questions of interest asked in Section 1.1 in which we describe data collection methods. Indicate whether the data come from an experiment or an observational study.

1.82 "Is there a sprinting gene?" Introduced in Example 1.5 on page 9.

1.83 "Do metal tags on penguins harm them?" Introduced in Data 1.3 on page 10.

1.84 "Are there human pheromones?" Introduced on page 11. Three studies are described; indicate whether each of them is an experiment or an observational study.

1.85 Salt on Roads and Accidents Three situations are described at the start of this section, on page 29. In the third bullet, we describe an association between the amount of salt spread on the roads and the number of accidents. Describe a possible confounding variable and explain how it fits the definition of a confounding variable.

1.86 Height and Reading Ability In elementary school (grades 1 to 6), there is a strong association between a child's height and the child's reading ability. Taller children tend to be able to read at a higher level. However, there is a very significant confounding variable that is influencing both height and reading ability. What is it?

1.87 Music Volume and Beer Consumption In 2008, a study[46] was conducted measuring the impact that music volume has on beer consumption. The researchers went into bars, controlled the music

[46]Gueguen, N., Jacob, C., Le Guellec, H., Morineau, T., and Lourel, M., "Sound Level of Environmental Music and Drinking Behavior: A Field Experiment with Beer Drinkers," *Alcoholism: Clinical and Experimental Research*, 2008; 32: 1795–1798.

volume, and measured how much beer was consumed. The article states that "the sound level of the environmental music was manipulated according to a randomization scheme." It was found that louder music corresponds to more beer consumption. Does this provide evidence that louder music causes people to drink more beer? Why or why not?

1.88 Nuts and Cholesterol Several studies have been performed to examine the relationship between nut consumption and cholesterol levels. Here we consider two such studies. In Study 1,[47] participants were assigned into two groups: one group was given nuts to eat each day, and the other group was told to consume a diet without nuts. In Study 2,[48] participants were free to follow their own diet, and reported how many nuts they consumed. Cholesterol levels were measured for all participants, and both studies found that nut consumption was associated with lower levels of LDL ("bad") cholesterol. Based on the information above, which study do you think provides better evidence that nut consumption reduces LDL cholesterol? Explain your answer.

1.89 Antibiotics in Infancy and Obesity in Adults "Antibiotics in infancy may cause obesity in adults," claims a recent headline.[49] A study in mice randomly assigned infant mice to either be given antibiotics or not, and the mice given antibiotics were more likely to be obese as adults. A separate study in humans found that children who had been given antibiotics before they were a year old (for example, for an ear infection) were more likely to be obese as adults. (Researchers believe the effect may be due to changes in the gut microbiome.) Based on these studies, is the headline an appropriate conclusion to make:

(a) For mice?

(b) For humans?

1.90 Do Online Cat Videos Improve Mood? Exercise 1.59 on page 28 introduced a study on cat videos, in which people who clicked on the link were asked questions regarding their mood before and after the most recent time they watched a cat video. Overall, participants reported that after watching a cat video they had significantly more energy, fewer negative emotions, and more positive emotions. Can we conclude from this study that watching cat videos increases energy and improves emotional state?

1.91 Green Spaces Make Kids Smarter A recent article[50] claims that "Green Spaces Make Kids Smarter." The study described in the article involved 2,623 schoolchildren in Barcelona. The researchers measured the amount of greenery around the children's schools, and then measured the children's working memories and attention spans. The children who had more vegetation around their schools did better on the memory and attention tests.

(a) What are the cases in this study?

(b) What is the explanatory variable?

(c) What is the response variable?

(d) Does the headline imply causation?

(e) Is the study an experiment or an observational study?

(f) Is it appropriate to conclude causation in this case?

(g) Suggest a possible confounding variable, and explain why it meets the requirements of a confounding variable.

1.92 Infections Can Lower IQ A headline in June 2015 proclaims "Infections can lower IQ."[51] The headline is based on a study in which scientists gave an IQ test to Danish men at age 19. They also analyzed the hospital records of the men and found that 35% of them had been in a hospital with an infection such as an STI or a urinary tract infection. The average IQ score was lower for the men who had an infection than for the men who hadn't.

(a) What are the cases in this study?

(b) What is the explanatory variable? Is it categorical or quantitative?

(c) What is the response variable? Is it categorical or quantitative?

(d) Does the headline imply causation?

(e) Is the study an experiment or an observational study?

(f) Is it appropriate to conclude causation in this case?

1.93 Sitting Is the New Smoking A 2014 headline reads "Sitting Is the New Smoking: Ways a

[47] Morgan, W.A., and Clayshulte, B.J., "Pecans lower low density lipoprotein cholesterol in people with normal lipid levels." *Journal of the American Dietetic Association*, 200; 100(3), 312–318.

[48] Li, T.Y., Brennan, A.M., Wedick, N.M., Mantzoros, C., Rifai, N., and Hu, F.B. "Regular consumption of nuts is associated with a lower risk of cardiovascular disease in women with type 2 diabetes." *The Journal of Nutrition*, 2009; 139(7), 1333–1338.

[49] Saey, T.H., "Antibiotics in infancy may cause obesity in adults," *Science News*, September 20, 2014.

[50] Khazan, O., "Green Spaces Make Kids Smarter," *The Atlantic*, June 16, 2016.

[51] "Infections can lower IQ," *The Week*, June 12, 2015, p. 18.

Sedentary Lifestyle is Killing You,"[52] and explains the mounting evidence for ways in which sitting is bad for you. A more recent large 2015 study[53] contributed to this evidence by following 69,260 men and 77,462 women and finding that for women, those who spent more leisure time sitting were significantly more likely to get cancer.

(a) What are the explanatory and response variables for the 2015 study?

(b) Is the 2015 study an observational study or a randomized experiment?

(c) Can we conclude from the 2015 study that spending more leisure time sitting causes cancer in women? Why or why not?

(d) Can we conclude from the 2015 study that spending more leisure time sitting does not cause cancer in women?

1.94 Late Night Eating It is well-known that lack of sleep impairs concentration and alertness, and this might be due partly to late night food consumption. A 2015 study[54] took 44 people aged 21 to 50 and gave them unlimited access to food and drink during the day, but allowed them only 4 hours of sleep per night for three consecutive nights. On the fourth night, all participants again had to stay up until 4 am, but this time participants were randomized into two groups; one group was only given access to water from 10 pm until their bedtime at 4 am while the other group still had unlimited access to food and drink for all hours. The group forced to fast from 10 pm on performed significantly better on tests of reaction time and had fewer attention lapses than the group with access to late night food.

(a) What are the explanatory and response variables?

(b) Is this an observational study or a randomized experiment?

(c) Can we conclude that eating late at night worsens some of the typical effects of sleep deprivation (reaction time and attention lapses)?

(d) Are there likely to be confounding variables? Why or why not?

1.95 To Spoon or Not to Spoon? Does cuddling after sex help boost sexual and relationship satisfaction? A study[55] involving 335 participants involved in romantic relationships found that people who reported more time spent on cuddling and affection after sex were more satisfied with their sex lives and relationships. This fact held true for both men and women. The average amount of time spent cuddling after sex was 15 minutes, and time spent on after sex affection was more strongly associated with sexual and relationship satisfaction than time spent on either foreplay or sex itself.

(a) Is this an observational study or a randomized experiment?

(b) Can we conclude that spending more time on affection after sex increases sexual and relationship satisfaction?

(c) A headline for an article[56] describing this study was titled "To Spoon or Not to Spoon? After-Sex Affection Boosts Sexual and Relationship Satisfaction." Does the study support this title?

(d) The title of the scientific article in which the study was originally published is "Post sex affectionate exchanges promote sexual and relationship satisfaction." Does the study support this title?

1.96 Does Early Language Reduce Tantrums? A recent headline reads "Early Language Skills Reduce Preschool Tantrums, Study Finds,"[57] and the article offers a potential explanation for this: "Verbalizing their frustrations may help little ones cope." The article refers to a study that recorded the language skill level and the number of tantrums of a sample of preschoolers.

(a) Is this an observational study or a randomized experiment?

(b) Can we conclude that "Early Language Skills Reduce Preschool Tantrums"? Why or why not?

(c) Give a potential confounding variable.

[52]"Sitting Is the New Smoking: Ways a Sedentary Lifestyle Is Killing You," *http://www.huffingtonpost.com/the-active-times/sitting-is-the-new-smokin_b_5890006.html*, September 29, 2014, Accessed July 17, 2015.

[53]Patel, A.V., et al., "Leisure-time spent sitting and site-specific cancer incidence in a large US cohort," *Cancer Epidemiology, Biomarkers & Prevention*, June 30, 2015, doi:10.1158/1055-9965.EPI-15-0237.

[54]University of Pennsylvania School of Medicine. "Eating less during late night hours may stave off some effects of sleep deprivation." *ScienceDaily*, June 4, 2015 *www.sciencedaily.com/releases/2015/06/150604141905.htm.*

[55]Muise, A., Giang, E., and Impett, E.A., "Post sex affectionate exchanges promote sexual and relationship satisfaction," *Archives of Sexual Behavior,* October 2014; 43(7): 1391–1402.

[56]"To Spoon or Not to Spoon? After-Sex Affection Boosts Sexual and Relationship Satisfaction," *http://www.scienceofrelationships.com/home/2014/5/16/to-spoon-or-not-to-spoon-after-sex-affection-boosts-sexual-a.html*, Accessed July 17, 2015.

[57]"Early Language Skills Reduce Preschool Tantrums, Study Finds," *US News and World Report*, *http://health.usnews.com/health-news/news/articles/2012/12/20/early-language-skills-reduce-preschool-tantrums-study-finds*, 20 December 2012, Accessed July 17, 2015.

1.97 Sleep and Recognition of Facial Expressions The ability to recognize and interpret facial expressions is key to successful human interaction. Could this ability be compromised by sleep deprivation? A 2015 study[58] took 18 healthy young adult volunteers and exposed them to 70 images of facial expressions, ranging from friendly to threatening. They were each shown images both after a full night of sleep and after sleep deprivation (24 hours of being awake), and whether each individual got a full night of sleep or was kept awake first was randomly determined. The study found that people were much worse at recognizing facial expressions after they had been kept awake.

(a) What are the explanatory and response variables?

(b) Is this an observational study or a randomized experiment? If it is a randomized experiment, is it a randomized comparative experiment or a matched pairs experiment?

(c) Can we conclude that missing a night of sleep hinders the ability to recognize facial expressions? Why or why not?

(d) In addition, for the people who had slept, the study found a strong positive association between quality of Rapid Eye Movement (REM) sleep and ability to recognize facial expressions. Can we conclude that better quality of REM sleep improves ability to recognize facial expressions? Why or why not? (*Hint:* What is the explanatory variable in this case? Was it randomly assigned?)

1.98 Diet Cola and Weight Gain in Humans A study[59] found that American senior citizens who report drinking diet soda regularly experience a greater increase in weight and waist circumference than those who do not drink diet soda regularly.

(a) From these results, can we conclude that drinking diet soda causes weight gain? Explain why or why not.

(b) Consider the results of this study on senior citizens, and the randomized experiment on rats introduced in Exercise 1.60 on page 28, which showed a similar association. Discuss what these two studies together might imply about the likelihood that diet cola causes weight gain in humans.

1.99 Does Red Increase Men's Attraction to Women? A study[60] examined the impact of the color red on how attractive men perceive women to be. In the study, men were randomly divided into two groups and were asked to rate the attractiveness of women on a scale of 1 (not at all attractive) to 9 (extremely attractive). One group of men were shown pictures of women on a white background and the other group were shown the same pictures of women on a red background. The men who saw women on the red background rated them as more attractive. All participants and those showing the pictures and collecting the data were not aware of the purpose of the study.

(a) Is this an experiment or an observational study? Explain.

(b) What is the explanatory variable and what is the response variable? Identify each as categorical or quantitative.

(c) How was randomization used in this experiment? How was blinding used?

(d) Can we conclude that using a red background color instead of white increases men's attractiveness rating of women's pictures?

1.100 Urban Brains and Rural Brains A study published in 2010 showed that city dwellers have a 21% higher risk of developing anxiety disorders and a 39% higher risk of developing mood disorders than those who live in the country. A follow-up study published in 2011 used brain scans of city dwellers and country dwellers as they took a difficult math test.[61] To increase the stress of the participants, those conducting the study tried to humiliate the participants by telling them how poorly they were doing on the test. The brain scans showed very different levels of activity in stress centers of the brain, with the urban dwellers having greater brain activity than rural dwellers in areas that react to stress.

(a) Is the 2010 study an experiment or an observational study?

(b) Can we conclude from the 2010 study that living in a city increases a person's likelihood of developing an anxiety disorder or mood disorder?

[58] Goldstein-Piekarski, A., et al., "Sleep Deprivation Impairs the Human Central and Peripheral Nervous System Discrimination of Social Threat," *The Journal of Neuroscience*, July 15, 2015; 35(28): 10135–10145; doi: 10.1523/JNEUROSCI.5254-14.2015

[59] Fowler, S.P., Williams, K., and Hazuda, H.P. "Diet Soda Intake Is Associated with LongTerm Increases in Waist Circumference in a Bioethnic Cohort of Older Adults: The San Antonio Longitudinal Study of Aging." *Journal of the American Geriatrics Society*, 2015; 63(4), 708–715.

[60] Elliot, A. and Niesta, D., "Romantic Red: Red Enhances Men's Attraction to Women," *Journal of Personality and Social Psychology*, 2008; 95(5): 1150–1164.

[61] "A New York state of mind," *The Economist*, June 25, 2011, p. 94.

(c) Is the 2011 study an experiment or an observational study?

(d) In the 2011 study, what is the explanatory variable and what is the response variable? Indicate whether each is categorical or quantitative.

(e) Can we conclude from the 2011 study that living in a city increases activity in stress centers of the brain when a person is under stress?

1.101 Split the Bill? When the time comes for a group of people eating together at a restaurant to pay their bill, sometimes they might agree to split the costs equally and other times will pay individually. If this decision were made in advance, would it affect what they order? Suppose that you'd like to do an experiment to address this question. The variables you will record are the type of *payment* (split or individual), *sex* of each person, number of *items* ordered, and the *cost* of each person's order. Identify which of these variables should be treated as *explanatory* and which as *response*. For each explanatory variable, indicate whether or not it should be randomly assigned.

1.102 Be Sure to Get Your Beauty Sleep! New research[62] supports the idea that people who get a good night's sleep look more attractive. In the study, 23 subjects ages 18 to 31 were photographed twice, once after a good night's sleep and once after being kept awake for 31 hours. Hair, make-up, clothing, and lighting were the same for both photographs. Observers then rated the photographs for attractiveness, and the average rating under the two conditions was compared. The researchers report in the *British Medical Journal* that "Our findings show that sleep-deprived people appear less attractive compared with when they are well rested."

(a) What is the explanatory variable? What is the response variable?

(b) Is this an experiment or an observational study? If it is an experiment, is it a randomized comparative design or a matched pairs design?

(c) Can we conclude that sleep deprivation *causes* people to look less attractive? Why or why not?

1.103 Do Antidepressants Work? Following the steps given, design a randomized comparative experiment to test whether fluoxetine (the active ingredient in Prozac pills) is effective at reducing depression. The participants are 50 people suffering from depression and the response variable is the change on a standard questionnaire measuring level of depression.

(a) Describe how randomization will be used in the design.

(b) Describe how a placebo will be used.

(c) Describe how to make the experiment double-blind.

1.104 Do Children Need Sleep to Grow? About 60% of a child's growth hormone is secreted during sleep, so it is believed that a lack of sleep in children might stunt growth.[63]

(a) What is the explanatory variable and what is the response variable in this association?

(b) Describe a randomized comparative experiment to test this association.

(c) Explain why it is difficult (and unethical) to get objective verification of this possible causal relationship.

1.105 Carbo Loading It is commonly accepted that athletes should "carbo load," that is, eat lots of carbohydrates, the day before an event requiring physical endurance. Is there any truth to this? Suppose you want to design an experiment to find out for yourself: "Does carbo loading actually improve athletic performance the following day?" You recruit 50 athletes to participate in your study.

(a) How would you design a randomized comparative experiment?

(b) How would you design a matched pairs experiment?

(c) Which design do you think is better for this situation? Why?

1.106 Alcohol and Reaction Time Does alcohol increase reaction time? Design a randomized experiment to address this question using the method described in each case. Assume the participants are 40 college seniors and the response variable is time to react to an image on a screen after drinking either alcohol or water. Be sure to explain how randomization is used in each case.

(a) A randomized comparative experiment with two groups getting two separate treatments

(b) A matched pairs experiment

1.107 Causation and Confounding Causation does not necessarily mean that there is no confounding variable. Give an example of an association between two variables that have a causal relationship AND have a confounding variable.

[62]Stein, R., "Beauty sleep no myth, study finds," *Washington Post*, *washingtonpost.com*, Accessed December 15, 2010.

[63]Rochman, B., "Please, Please, Go to Sleep," *Time Magazine*, March 26, 2012, p. 46.