3 Descriptive Designs—Observing Behavior

Alexander Macfarlane/Axiom Photographic/Design Pics/Superstock

Learning Outcomes

By the end of this chapter, you should be able to:

- Explain the distinguishing features of qualitative research.
- Distinguish the key features, pros, and cons of case studies.
- Distinguish the key features, pros, and cons of archival research.
- Distinguish the key features, pros, and cons of observational research.
- Outline best practices for describing data, both graphically and numerically.

In the fall of 2009, Phoebe Prince and her family relocated from Ireland to South Hadley, Massachusetts. Phoebe was immediately singled out by bullies at her new high school and subjected to physical threats, insults about her Irish heritage, and harassing posts on her Facebook page. This relentless bullying continued until January of 2010, ending only because Phoebe elected to take her own life in order to escape her tormentors ("Report of plea deal," 2011). Tragic stories like this one are all too common, and it should come as no surprise that the Centers for Disease Control and Prevention (2012) has identified bullying as a serious problem facing our nation's children and adolescents.

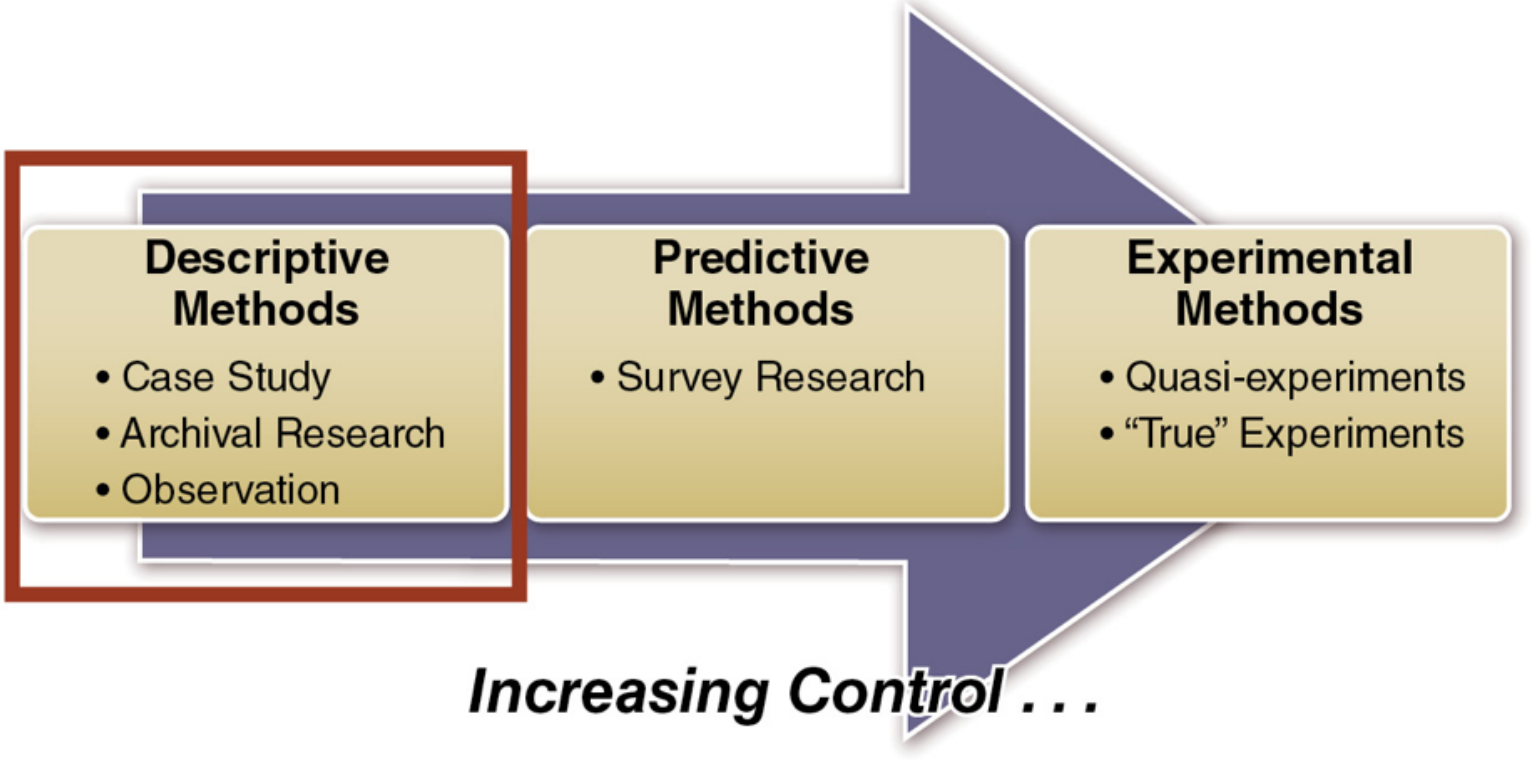
Scientific research on bullying began in Norway in the late 1970s in response to a wave of teen suicides. Work begun by psychologist Dan Olweus—and since continued by many others—has documented both the frequency and the consequences of bullying in the school system. Thus, we know that approximately one third of children are victims of bullying at some point during development, with between 5% and 10% bullied on a regular basis (Griffin & Gross, 2004; Nansel et al., 2001). Victimization by bullies has been linked with a wide range of emotional and behavioral problems, including depression, anxiety, self-reported health problems, and an increased risk of both violent behavior and suicide (for a detailed review, see Griffin & Gross, 2004). Recent research even suggests that bullying during adolescence may have a lasting impact on the body's physiological stress response (Hamilton, Newman, Delville, & Delville, 2008).

Nevertheless, most of this research has a common limitation: It has studied the phenomenon of bullying using self-report survey measures. That is, researchers typically ask students and teachers to describe the extent of bullying in the schools. In many studies, researchers will also have students fill out a collection of survey measures, describing both bullying experiences and psychological functioning in their own words. These studies are conducted rigorously, and the measures they use certainly meet the criteria of reliability and validity discussed in Chapter 2 (2.2). However, as Wendy Craig, Professor of Psychology at Queen's University, and Debra Pepler, a Distinguished Professor at York University, suggested in a 1997 article, this questionnaire approach cannot capture the full context of bullying behaviors. As we have already discussed, self-report measures are fully dependent on people's ability and willingness to answer honestly and accurately. It is easy to imagine scenarios in which reports of bullying experiences might be downplayed out of fear, or perhaps misremembered simply due to the stress of the experience itself.

To address this limitation, Craig and Pepler (1997) decided to observe bullying behaviors as they occurred naturally on the playground. Among other things, the researchers found that acts of bullying occurred approximately every 7 minutes, lasted only about 38 seconds, and tended to occur within 120 feet of the school building. They also found that peers intervened to try to stop the bullying more than twice as often as adults did (11% versus 4%, respectively). These findings add significantly to scientific understanding of when and how bullying occurs. For our purposes, the most notable thing about them is that none of the findings could have been documented without directly observing and recording bullying behaviors on the playground. By using this technique, the researchers were able to gain a more thorough understanding of the phenomenon of bullying and, as a result, to provide real-world advice to teachers and parents.

One recurring theme in this book is that it is absolutely critical for researchers to pick the right research design to address their hypothesis. The next three chapters will discuss three specific categories of research designs, proceeding in order of increasing control over elements of the design (see Figure 3.1). This chapter focuses on descriptive research designs, in which the primary goal is to describe attitudes or behavior. We will begin by contrasting qualitative and quantitative approaches to description. We will then discuss three approaches to descriptive designs—studying single cases, mining existing archives, and observing behavior directly—covering the basic concept and the pros and cons of each. Finally, the chapter concludes with a discussion of guidelines for presenting descriptive data in both graphical and numeric form.

Figure 3.1: Descriptive designs on the continuum of control


Descriptive Methods
• Case Study
• Archival Research
• Observation
Predictive Methods
• Survey Research
Experimental Methods
• Quasi-experiments
• "True" Experiments
Increasing Control . . .

3.1 Qualitative and Quantitative Methods

Chapter 1 explained that researchers generally take one of two broad approaches to answering their research questions. *Quantitative* research is a systematic and empirical approach that attempts to generalize results to other contexts, whereas *qualitative* research is a more descriptive approach that attempts to gain a deep understanding of particular cases and contexts. Before we discuss specific examples of descriptive designs, it is important to understand that these can represent either quantitative *or* qualitative perspectives. This section contrasts the two approaches in more detail.

Chapter 1 used the analogy of studying traffic patterns to contrast qualitative and quantitative methods—a qualitative researcher would likely study a single busy intersection in detail. This example illustrates a key point about this approach: Qualitative researchers are focused on interpreting and making sense out of what they observe rather than trying to simplify and quantify these observations. In general, qualitative research involves a collection of interviews and observations made in a natural setting. Regardless of the overall approach (qualitative or quantitative), data collection in the real world results in less control and structure than it does in a laboratory setting. But whereas quantitative researchers might view reduced control as a threat to reliability and validity, qualitative researchers view it as a strength of the study. Conducting observations in a natural setting makes it possible to capture people's natural and unfiltered responses.

As an example, consider two studies of the ways people respond to traumatic events. In a 1993 paper, psychologists James Pennebaker and Kent Harber took a quantitative approach to examining the community-wide impact of the 1989 Loma Prieta earthquake (near San Francisco). These researchers conducted phone surveys of 789 area residents, asking people to indicate, using a 10-point scale, how often they "thought about" and "talked about" the earthquake during the three-month period after its occurrence. In analyzing these data, Pennebaker and Harber discovered that people tend to stop talking about traumatic events about two weeks after they occurred but keep thinking about the event for approximately four more weeks. That is, the event is still on people's minds, but they decide to stop discussing it with other people. In a follow-up study using the 1991 Gulf War, the same researchers found that this conflict leads to an increased risk of illness, measured via an increase in visits to the doctor (Pennebaker & Harber, 1993). The goal of the study was to gather data in a controlled manner and test a set of hypotheses about community responses to trauma.

Contrast Pennebaker and Harber's approach with the more qualitative one taken by the developmental psychologist Paul Miller and colleagues (2012), who used a qualitative approach to study the ways that parents model coping behavior for their children. These researchers conducted semistructured interviews of 24 parents whose families had been evacuated following the 2007 wildfires in San Diego County and an additional 32 parents whose families had been evacuated following a 2008 series of deadly tornadoes in Tennessee. Because of a lack of prior research on how parents teach their children to cope with trauma, Miller and colleagues approached their interviews with the goal of "documenting and describing" (p. 8) these processes. That is, rather than attempt to impose structure and test a strict hypothesis, the researchers focused on learning from these interviews and letting the interviewees' perspectives drive the acquisition of knowledge.

Qualitative and quantitative methods also differ quite strikingly in how they approach analyses of the data. Because all quantitative methods have the goal of discovering findings that can be generalized—that apply across different contexts—all quantitative studies must translate phenomena into numerical values and conduct statistical analyses. So, for example, Pennebaker and Harber's (1993) study of coping with trauma measured the concrete value of "visits to the doctor," and then compared changes in this number over time. In contrast, because qualitative methods have the goal of learning and interpreting phenomena from the ground up, qualitative studies focus on discovering the underlying meaning of phenomena in their own right. So, for example, Miller and colleagues' 2012 study of coping focused on "documenting and describing" the ways that

Lisafx/iStock/Thinkstock

Paul Miller's research, which involved a series of

parents teach children to cope and learning from a critical evaluation of the interview content. At risk of oversimplifying: Quantitative methods gloss over some of the richness of experience in order to discover knowledge that can be generalized, while qualitative methods sacrifice the ability to generalize in order to capture the richness of experience.

semi-structured, qualitative interviews, attempted to document and describe a phenomenon rather than test a theory.

As one final example of this contrast, consider the way that each approach would analyze the content of an interview. Interviewing people can be a very effective way to understand their experiences and can form the basis for many of the descriptive designs we cover in this chapter. A qualitative researcher would likely conduct a smaller number of interviews (perhaps only one, for a case study), due to the time required for analysis. The researcher would read each interview in depth and then start to identify themes that appeared across the entire set. These themes would serve as the basis for understanding people's experiences. (For an excellent deep dive into different theoretical approaches to interview analysis, see Smith [2008].) By comparison, a quantitative researcher would conduct a larger number of interviews, because quantitative text analysis can be very fast. Rather than read each interview, the researcher could input the text of these interviews into a software program, which could count and categorize the overall sentiment of the language people used. These counts and categories would then serve as the basis for quantifying people's experiences on a larger scale.

The following three sections examine three specific examples of descriptive designs—case studies, archival research, and observational research. Because each of these methods has the goal of describing attitudes, feelings, and behaviors, each one can be used from either a quantitative or a qualitative perspective. In other words, qualitative and quantitative researchers use many of the same general methods but do so with different goals. To illustrate this flexibility, each section concludes with a paragraph that contrasts qualitative and quantitative uses of the particular method.

3.2 Case Studies

At the 1996 meeting of the American Psychological Association, James Pennebaker—now chair of the psychology department at the University of Texas—delivered an invited address, describing his research on the benefits of therapeutic writing. Rather than follow the normal approach to an academic conference presentation, showing graphs and statistical tests to support his arguments, Pennebaker told a story. In the mid-1980s, when Pennebaker's lab was starting to study the effects of structured writing on physical and psychological health, one study participant was an American soldier who had served in the Vietnam War. Like many others, this soldier experienced difficulty adjusting to what had happened during the war and consequent trouble reintegrating into "normal" life. In Pennebaker's study, he was asked to simply spend 15 minutes per day, over the course of a week, writing about a traumatic experience—in this case, his tour of duty in Vietnam. At the end of this week, as might have been expected, this veteran felt awful, revisiting unpleasant memories over a decade old. But over the next few weeks, amazing things started to happen: The soldier slept better; he made fewer visits to his doctor; he even reconnected with his wife after a long separation.

Pennebaker's presentation was a **case study**, which provides a detailed, in-depth analysis of one person over a period of time. Although this case study was collected as part of a larger quantitative experiment, case studies are usually conducted in a therapeutic setting and involve a series of interviews. An interviewer will typically study the subject in detail, recording everything from direct quotes and observations to his or her own interpretations. We encountered this technique briefly in Chapter 2 (2.1), in discussing Oliver Sacks's case studies of individuals learning to live with neurological impairments.

Pros and Cons of Case Studies

Case studies in psychology are a form of qualitative research and represent the lowest point on our continuum of control. Because they involve one person at a time, without a control group, case studies are often unsystematic. That is, the participants are chosen, rather than selected randomly, because they tell a compelling story or because they represent an unusual set of circumstances. Studying these individuals allows for a great deal of exploration, which can often inspire future research. However, it is nearly impossible to generalize from one case study to the larger population. In addition, because the case study includes both direct observation and the researcher's interpretation, a researcher's biases run the risk of influencing the interpretations. For example, Pennebaker's personal investment in demonstrating that writing has health benefits could have led to more positive interpretations of the Vietnam veteran's outcomes. However, in this particular case study, the benefits of writing mirror those seen in hundreds of controlled experimental studies involving thousands of people, so we can feel confident in the conclusions from the single case.

Case studies have two distinct advantages over other forms of research. First is the simple fact that anecdotes are persuasive. Despite Pennebaker's nontraditional approach to a scientific talk, the audience came away utterly convinced of the benefits of therapeutic writing. And although Oliver Sacks studied one neurological patient at a time, the collection of stories in his books sheds very convincing light on the ability of humans to adapt to their circumstances. Second, case studies provide a useful way to study rare populations and individuals with rare conditions. For example, from a scientific point of view, the ideal might be to gather a random sample of individuals living with severe memory impairment due to alcohol abuse and conduct some sort of controlled study in a laboratory environment. This approach could allow us to make causal statements about the results, as Chapter 5 (5.4) will discuss. But from a practical point of view, such a study would be nearly impossible to conduct, making case studies such as Sacks's interviews with William Thompson the best strategy for understanding this condition in depth.

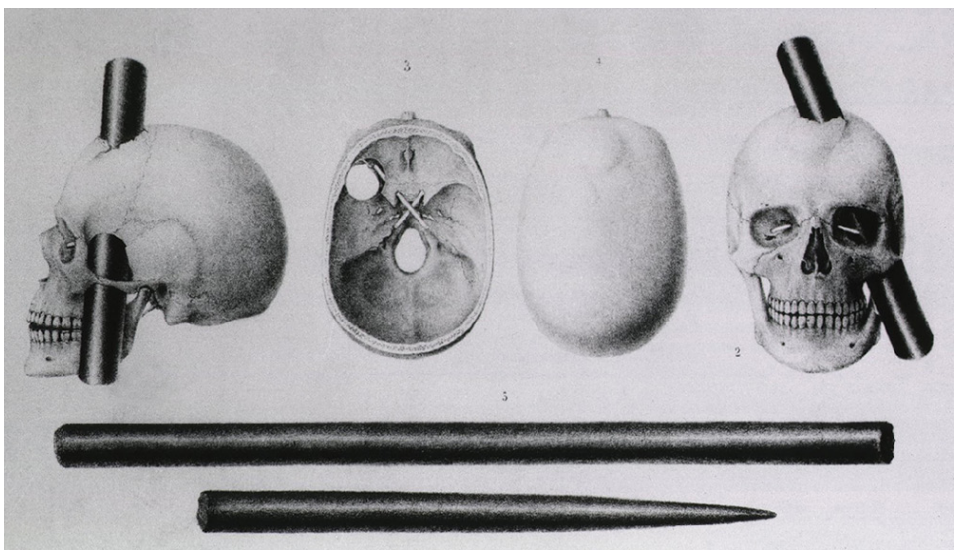
Examples of Case Studies

Throughout the history of psychology, case studies have been used to address a number of important questions and to provide a starting point for controlled quantitative studies. For example, in developing his theories of cognitive development, the Swiss psychologist Jean Piaget first studied the way that his own children developed and changed their thinking styles. Piaget proposed that children would progress through a series of four stages in the way that they

approached the world—sensorimotor, preoperational, concrete operational, and formal operational—with each stage involving more sophisticated cognitive skills than the previous stage. By observing his own children, Piaget noticed preliminary support for this theory and later was able to conduct more controlled research with larger populations.

Perhaps one of the most famous case studies in psychology is that of Phineas Gage, a 19th-century railroad worker who suffered severe brain damage. In September of 1848, Gage was working with a team to blast large sections of rock to make way for new rail lines. After a large hole was drilled into a section of rock, Gage's job was to pack the hole with gunpowder, sand, and a fuse and then tamp it down with a long cylindrical iron rod (known as a "tamping rod"). On this particular occasion, it seems Gage forgot to pack in the sand. So, when the iron rod struck gunpowder, the powder exploded, sending the 3-foot long iron rod through his face, behind his left eye, and out the top of his head. Against all odds, Gage survived this incident with relatively few physical side effects. However, everyone around him noticed that his personality had changed—Gage became more impulsive, violent, and argumentative. Gage's physician, John Harlow, reported the details of this case in an 1868 article. The following passage offers a strong example of the rich detail that is often characteristic of case studies:

> He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires. A child in his intellectual capacity and manifestations, he has the animal passions of a strong man. Previous to his injury, although untrained in the schools, he possessed a well-balanced mind, and was looked upon by those who knew him as a shrewd, smart businessman, very energetic and persistent in executing all his plans of operation. In this regard his mind was radically changed, so decidedly that his friends and acquaintances said he was "no longer Gage." (pp. 339–342)



Everett Collection

Various views show an iron rod embedded in Phineas Gage's (1823–1860) skull.

Gage's transformation ultimately inspired a large body of work in psychology and neuroscience that attempts to understand the connections between brain areas and personality. The area of his brain destroyed by the tamping rod is known as the frontal lobe, now understood to play a critical role in impulse control, planning, and other high-level thought processes. Gage's story is a perfect illustration of the pros and cons of case studies: On the one hand, it is difficult to determine exactly how much the brain injury affected his behavior because he is only one person. On the other hand, Gage's tragedy inspired researchers to think about the connections among mind, brain, and personality. As a result, we now have a vast—and still growing—understanding of the brain. The story illustrates a key point about case studies: Although individual cases provide only limited knowledge about people in general, these cases often lead researchers to conduct additional work that does lead to generalizable knowledge.

Qualitative versus Quantitative Approaches

Case studies tend to be qualitative more often than not: The goal of this method is to study a particular case in depth, as a way to learn more about a rare phenomenon. In both Pennebaker's study of the Vietnam veteran and Harlow's study of Phineas Gage, the researcher approached the interview process as a way to gather information and learn from the bottom up about the interviewee's experience. However, a case study can certainly represent quantitative research. This is often the case when researchers conduct a series of case studies, learning from the first one or the first few and then developing hypotheses to test on future cases. For example, a researcher could use the case of Phineas Gage as a starting point for hypotheses about frontal lobe injury, perhaps predicting that other cases would show similar changes in personality. Another way in which case studies can add a quantitative element is for researchers to conduct analyses within a single subject. For example, a researcher could study a patient with brain damage for several years following an injury, tracking the association between deterioration of brain regions with changes in personality and emotional responses. At the end of

the day, though, these examples would still suffer the primary drawback of case studies: Because they examine a single individual, researchers find it difficult to generalize findings.

Research: Thinking Critically

Analyzing Acupuncture

Follow the link below to a press release from the Peninsula College of Medicine and Dentistry. This short article reviews recent research from the college, suggesting that acupuncture treatment might be of benefit to patients suffering from "unexplained" symptoms. As you read the article, consider what you have learned so far about the research process, and then respond to the questions below.

http://www.sciencedaily.com/releases/2011/05/110530080513.htm
(http://www.sciencedaily.com/releases/2011/05/110530080513.htm)

Think about it:

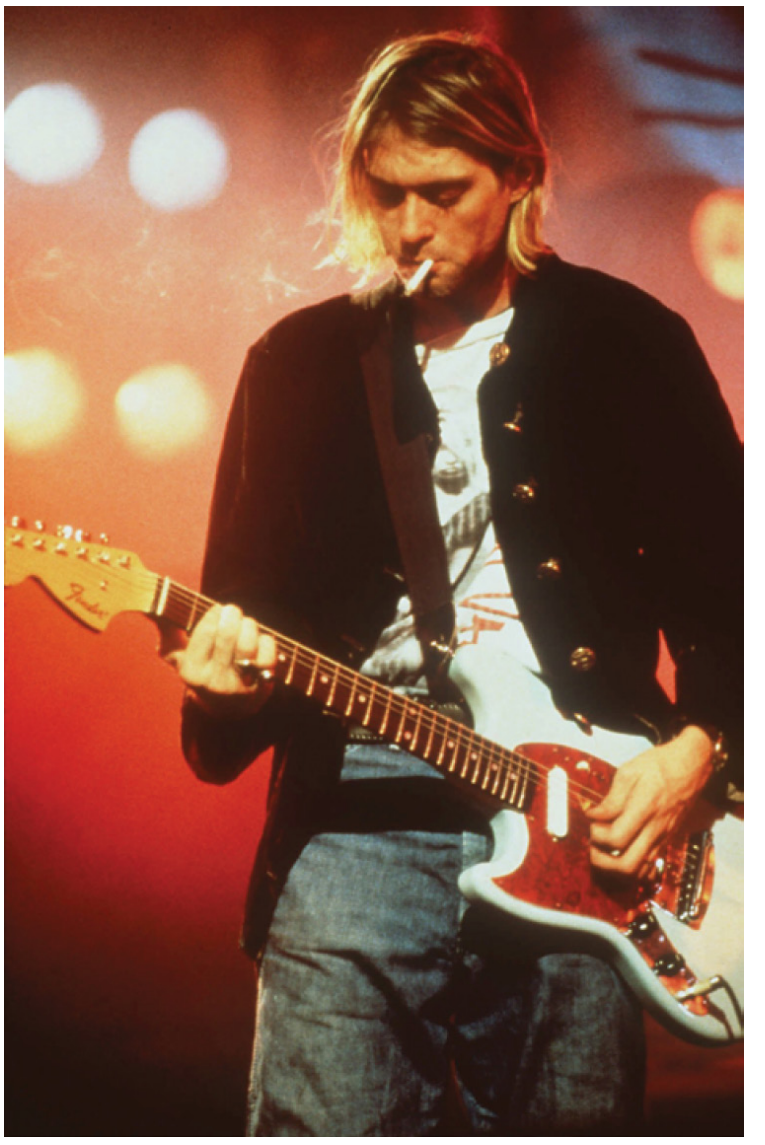
1. In this study, researchers interviewed acupuncture patients using open-ended questions and recorded their verbal responses, which is a common qualitative research technique. What advantages does this approach have over administering a quantitative questionnaire with multiple-choice items?
2. What are some advantages of adding a qualitative element to a controlled medical trial like this?
3. What would be some disadvantages of relying exclusively on this approach?

3.3 Archival Research

Slightly further along the continuum of control is **archival research**, which involves drawing conclusions by analyzing existing sources of data, including both public and private records. Sociologist David Phillips (1977) hypothesized that media coverage of suicides would lead to "copycat" suicides. He tested this hypothesis by gathering archival data from two sources: front-page newspaper articles devoted to high-profile suicides and the number of fatalities in the 11-day period following coverage of the suicide. By examining these patterns of data, Phillips found support for his hypothesis. Specifically, fatalities appeared to peak three days after coverage of a suicide, and a greater degree of publicity was associated with a greater peak in fatalities.

Pros and Cons of Archival Research

It is difficult to imagine a better way to test Phillips's hypothesis about copycat suicides. A researcher could never randomly assign people to learn about suicides and then wait to see whether they killed themselves. Nor could someone interview people right before they commit suicide to determine whether they were inspired by media coverage. Archival research provides a test of the hypothesis by examining data that already exist and, thereby, avoids most of the ethical and practical problems of other research designs. One key element of archival research is that it neatly sidesteps issues of **participant reactivity**, or the tendency of people to behave differently when they are aware of being observed. Any time research is conducted in a laboratory, participants know they are part of a study and may not behave in a completely natural manner. In contrast, archival data involves making use of records of people's natural behaviors. The subjects of Phillips's study of copycat suicides were individuals who decided to kill themselves, who had no awareness that they would be part of a research study.



AP Photo

Copycat suicides often peak 3 days after media coverage of a high profile suicide, such as when Nirvana's Kurt Cobain killed himself in 1994.

Archival research is also an excellent strategy for examining trends and changes over time. For example, much of the evidence for global warming comes from observing upward trends in recorded temperatures around the globe. To gather this evidence, researchers dig into existing archives of weather patterns and conduct statistical tests of the changes over time. Psychologists and other social scientists also make use of this approach to examine population-level changes in everything from suicide rates to voting patterns over time. These comparisons can sometimes involve a blend of archival and current data. For example, a great deal of social-psychology research has been dedicated to understanding people's stereotypes about other groups. In a classic series of studies known as the "Princeton Trilogy," researchers documented the stereotypes held by Princeton students for 25 years (1933 to 1969). Social psychologist Stephanie Madon and her colleagues (2001) collected a new round of data but also conducted a new analysis of the previous archival data. These new analyses suggested that, over time, people have become more willing stereotype other groups, even as the stereotypes themselves have become less negative.

One final advantage of archival research is that once a researcher gains access to the relevant archives, it requires relatively few resources. The typical laboratory experiment involves one participant at a time, sometimes requiring the dedicated attention of more than one research assistant for an hour or more. After researchers assemble data from the archives, though, conducting statistical analyses is a relatively simple matter. In a 2001 article, the psychologists Shannon Stirman and James Pennebaker used a text-analysis computer program to compare the language of poets who committed suicide (e.g., Sylvia Plath) with the language of similar poets who had not committed suicide (e.g. Denise Levertov). In total, these researchers examined 300 poems from 20 poets, half of whom had committed suicide. Consistent with Durkheim's theory

of suicide as a form of "social disengagement," Stirman and Pennebaker (2001) found that suicidal poets used more self-references and fewer references to other people in their poems. The impressive part of the study is this: Once they had assembled their archive of poems, their computer program took only seconds to analyze the language and generate a statistical profile of each poet.

Overall, however, archival research is still relatively low on the continuum of control. Researchers have to accept the archival data in whatever form they exist, with no control over the way they were collected. For instance, in Stephanie Madon's (2001) reanalysis of the "Princeton Trilogy" data, she had to trust that the original researchers had collected the data in a reasonable and unbiased way. In addition, because archival data often represent natural behavior, it can be difficult to categorize and organize responses in a meaningful and quantitative way. The upshot is that archival research often requires some creativity on the researcher's part—such as analyzing poetry using a text-analysis program. In many cases, as we discuss next, the process of analyzing archives involves developing a coding strategy for extracting the most relevant information.

Content Analysis—Analyzing Archives

In most examples so far, the data come in a straightforward, ready-to-analyze form. That is, it is relatively simple to count the number of suicides, track the average temperature, or compare responses to questionnaires about stereotyping over time. In other cases, the data can come as a sloppy, disorganized mass of information. How does someone who wants to analyze literature, media images, or changes in race relations on television accomplish the analysis? These types of data can yield incredibly useful information, provided the researcher can develop a strategy for extracting it.

Mark Frank and Tom Gilovich—both psychologists at Cornell University—were interested in whether cultural associations with the color black affected behavior. In virtually all cultures, the term "black" is associated with evil—the bad guys wear black hats; people have a "black day" when things turn sour; and some are excluded from social groups by being "blacklisted" or "blackballed." These associations appear to be independent of any culture-specific prejudices regarding race or skin color. Frank and Gilovich (1988) wondered whether "a cue as subtle as the color of a person's clothing" (p. 74) would influence aggressive behavior. To test this hypothesis, they examined aggressive behaviors in professional football and hockey games, comparing teams whose uniforms were black to teams who wore other colors. Imagine for a moment being a researcher for this study. Professional sporting events contain a wealth of behaviors and events. How would information about the relationship between uniform color and aggressive behavior be extracted?

Frank and Gilovich (1988) solved this problem by examining public records of penalty yards (football) and penalty minutes (hockey) because these represent instances of punishment for excessively aggressive behavior, as recognized by the referees. In addition, in both sports, the size of the penalty increases according to the degree of aggression. These penalty records were obtained from the central offices of both leagues, covering the period from 1970 to 1986. Consistent with the researchers' hypothesis, teams with black uniforms were "uncommonly aggressive" (p. 76). Most strikingly, two NHL hockey teams changed their uniforms to black during the period under study and showed a marked increase in penalty minutes with the new uniforms. One equally compelling alternative explanation is that, rather than the teams *acting* more aggressive in black uniforms, referees *perceived* them to be more aggressive while wearing black uniforms. Both explanations are consistent with the idea that cultural associations can affect behavior.

Even this analysis, however, is relatively straightforward because it involved data that were already in quantitative form (penalty yards and minutes). In many cases, the starting point is a jumbled mess of human behavior. In a pair of journal articles, psychologist Russell Weigel and colleagues (1980; 1995) examined the portrayal of race relations on prime-time television. To do so, they had to make several critical decisions about what to analyze and how to quantify it. The process of systematically extracting and analyzing the contents of a collection of information is known as **content analysis**. In essence, content analysis involves developing a plan to code and record specific behaviors and events in a consistent way. We can break this plan down into a three-step process.

Step 1—Identify Relevant Archives

Before we develop our coding scheme, we have to start by finding the most appropriate source of data. Sometimes the choice is fairly obvious: To compare temperature trends, the most relevant archives will be weather records. To track changes in stereotyping over time, the most relevant archive is questionnaire data assessing people's attitudes. In other cases, this decision involves careful consideration of both the research question and practical concerns. Frank and Gilovich decided to study penalties in professional sports because these data were both readily available (from the central league offices) and highly relevant to their hypothesis about aggression and uniform color.

Because these penalty records were publicly available, the researchers were able to access them easily. But if the research question involved sensitive or personal information—such as hospital records or personal correspondence—researchers would need to obtain permission from a responsible party. Say we wanted to analyze the love letters written by soldiers serving overseas and then try to predict relationship stability. Given the personal, even intimate nature of these letters, we would need permission from each person involved before proceeding with the study. However researchers manage to obtain access to private records, protecting the privacy and anonymity of the people involved is paramount. This would mean, for example, using pseudonyms and/or removing names and other identifiers from published excerpts of personal letters.

Step 2—Sample From the Archives

In Weigel's research on race relations, the most obvious choice of archives comprised snippets of both television programming and commercials. Yet this decision was only the first step of the process. Should the researchers examine every second of every program ever aired on television? Naturally not; instead, their approach was to take a smaller sample of television programming. Chapter 4 (4.3) will discuss sampling in more detail, but the basic process involves taking a smaller, representative collection of the broader population to conserve resources. Weigel and colleagues (1980) decided to sample one week's worth of prime-time programming from 1978, assembling videotapes of everything broadcast by the three major networks at the time (CBS, NBC, and ABC). The research team narrowed its sample by eliminating news, sports, and documentary programming because the hypotheses centered on portrayals of fictional characters of different races.

Step 3—Code and Analyze the Archives

Content analysis' third and most involved step is to develop a system for coding and analyzing the archival data. Even a sample of one week's worth of prime-time programming contains a near-infinite amount of information. In the race-relations studies, Weigel et al. elected to code four key variables: (1) the "total human appearance time," or time during which people were onscreen; (2) the "Black appearance time," in which Black characters appeared onscreen; (3) the "cross-racial appearance time," in which characters of two races were onscreen at the same time; and (4) the "cross-racial interaction time," in which cross-racial characters interacted. In the original (1980) paper, these authors reported that Black characters were shown only 9% of the time, and cross-racial interactions only 2% of the time. Fortunately, by the time of their 1995 follow-up study, the rate of Black appearances had doubled, and the rate of cross-racial interactions had more than tripled. However, depressingly little change occurred in some of the qualitative dimensions that they measured, including the degree of emotional connection between characters of different races.

This study also highlights the variety of options for coding complex behaviors. The four key ratings of "appearance time" consist of simply recording the amount of time that each person or group is onscreen. In addition, the researchers assessed several abstract qualities of interaction using judges' ratings. The degree of emotional connection, for instance, was measured by having judges rate the "extent to which cross-racial interactions were characterized by conditions promoting mutual respect and understanding" (Weigel et al., 1980, p. 888). As Chapter 2 (2.2) explained, any time researchers use judges' ratings, it is important to collect ratings from more than one rater and to make sure they agree in their assessments.

A researcher's goal is to find a systematic way to record the observations most relevant to the hypothesis. This is particularly true for quantitative research, where the key is to start with clear operational definitions that capture the variables of interest. This involves both deciding the most appropriate variables and the best way to measure these variables. For example, if someone who analyzes written communication might decide to compare words, sentences,

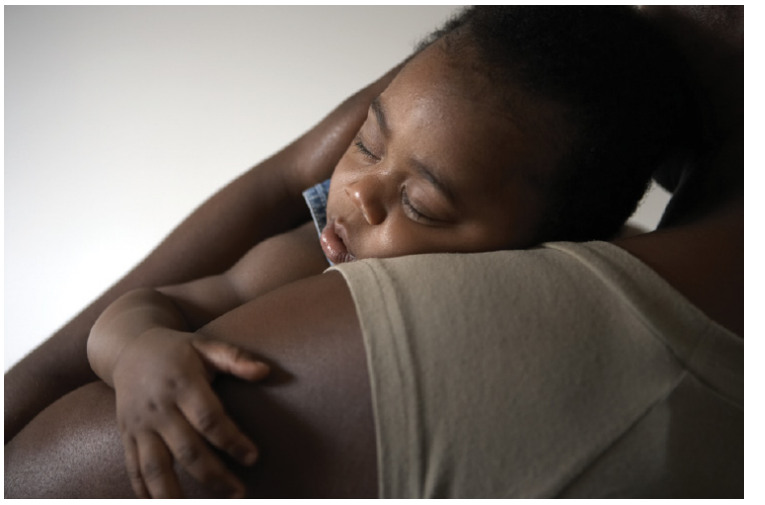
characters, or themes across the sample. A study of newspaper coverage might code the amount of space or number of stories dedicated to a topic, while a study of television news might code the amount of airtime given to different positions. The best strategy in each case will be the one that best represents the variables of interest.

Qualitative versus Quantitative Approaches

Archival research can represent either qualitative or quantitative research, depending on the researcher's approach to the archives. Most of the examples in this section represent the quantitative approach: Frank and Gilovich (1988) counted penalties to test their hypothesis about aggression, and Stirman and Pennebaker (2001) counted words to test their hypothesis about suicide. However, the race-relations work by Weigel and colleagues (1980; 1995) represents a nice mix of qualitative and quantitative research. In the initial 1980 study, their primary goal was to document the portrayal of race relations on prime-time television, learning from the ground up (i.e., qualitative). In the 1995 follow-up study, though, the researchers primarily wanted to determine whether these portrayals had changed over a 15-year period. That is, they tested the hypothesis that race relations were portrayed in a more positive light (i.e., quantitative). Another way in which archival research can be qualitative is to study open-ended narratives, without attempting to impose structure upon them. This approach is commonly used to study free-flowing text, such as personal correspondence or letters to the editor in a newspaper. A researcher approaching these from a qualitative perspective would attempt to learn from these narratives, without attempting to impose structure via the use of content analyses.

3.4 Observational Research

Moving further along the continuum of control, we come to the descriptive design with the greatest amount of researcher control. **Observational research** involves studies that directly observe behavior and record these observations in an objective and systematic way. Your previous psychology courses may have explored the concept of attachment theory, which argues that an infant's bond with his or her primary caregiver has implications for later social and emotional development. Mary Ainsworth, a Canadian developmental psychologist, and John Bowlby, a British psychologist and psychiatrist, articulated this theory in the 1960s. They argued that children can form either "secure" or a variety of "insecure" attachments with their caregivers (Ainsworth & Bell, 1970; Bowlby, 1963).



Rayes/Photodisc/Thinkstock

Observational research can be used to measure an infant's attachment to a caregiver.

To assess these classifications, Ainsworth and Bell developed an observational technique called the "strange situation." Mothers would arrive at their laboratory with their children for a series of structured interactions, including having the mother play with the infant, leave him alone with a stranger, and then return to the room after a brief absence. The researchers were most interested in coding the ways in which the infant responded to these various episodes (eight in total). One group of infants, for example, was curious when the mother left but then returned to playing with toys, trusting that she would return. Another group showed immediate distress when the mother left and clung to her nervously upon her return. Based on these and other behavioral observations, Ainsworth and colleagues classified these groups of infants as "securely" and "insecurely" attached to their mothers, respectively.

Research: Making an Impact

Harry Harlow

In the 1950s, U.S. psychologist Harry Harlow conducted a landmark series of studies on the mother–infant bond using rhesus monkeys. Although contemporary standards would consider his research unethical, the results of his work revealed the importance of affection, attachment, and love on healthy childhood development.

Prior to Harlow's findings, it was believed that infants attached to their mothers as a part of a drive to fulfill exclusively biological needs, in this case obtaining food and water and avoiding pain (Herman, 2007; van der Horst & van der Veer, 2008). In an effort to clarify the reasons that infants so clearly need maternal care, Harlow removed rhesus monkeys from their natural mothers several hours after birth, giving the young monkeys a choice between two surrogate "mothers." Both mothers were made of wire, but one was bare and one was covered in terry cloth. Although the wire mother provided food via an attached bottle, the monkeys preferred the softer, terry-cloth mother, even though the latter provided no food (Harlow & Zimmerman, 1958; Herman, 2007).

Further research with the terry-cloth mothers contributed to the understanding of healthy attachment and childhood development (van der Horst & van der Veer, 2008). When the young monkeys were given the option to explore a room with their terry-cloth mothers and had the cloth mothers in the room with them, they used the mothers as a safe base. Similarly, when exposed to novel stimuli such as a loud noise, the monkeys would seek comfort from the cloth-covered surrogate (Harlow & Zimmerman, 1958). However, when the monkeys were left in the room without their cloth mothers, they reacted poorly—freezing up, crouching, crying, and screaming.

A control group of monkeys who were never exposed to either their real mothers or one of the surrogates revealed stunted forms of attachment and affection. They were left incapable of forming lasting emotional attachments with

other monkeys (Herman, 2007). Based on this research, Harlow discovered the importance of proper emotional attachment, stressing the importance of physical and emotional bonding between infants and mothers (Harlow & Zimmerman, 1958; Herman, 2007).

Harlow's influential research led to improved understanding of maternal bonding and child development (Herman, 2007). His research paved the way for improvements in infant and child care and in helping children cope with separation from their mothers (Bretherton, 1992; Du Plessis, 2009). In addition, Harlow's work contributed to the improved treatment of children in orphanages, hospitals, day care centers, and schools (Herman, 2007; van der Horst & van der Veer, 2008).

Pros and Cons of Observational Research

Observational designs are well suited to a wide range of research questions, provided the questions can be addressed through *directly observable* behaviors and events. For example, researchers can observe parent–child interactions, or nonverbal cues to emotion, or even crowd behavior. However, if they are interested in studying thought processes—such as how close mothers feel to their children—then observation will not suffice. This point harkens back to the discussion of behavioral measures in Chapter 2 (2.2): In exchange for giving up access to internal processes, researchers gain access to unfiltered behavioral responses.

To capture these unfiltered behaviors, it is vital for the researcher to be as unobtrusive as possible. As we have already discussed, people have a tendency to change their behavior when they are being observed. In the bullying study by Craig and Pepler (1997) discussed at the beginning of this chapter, the researchers used video cameras to record children's behavior unobtrusively. Imagine how (artificially) low the occurrence of bullying might be if the playground had been surrounded by researchers with clipboards!

If researchers conduct an observational study in a laboratory setting, they have no way to hide the fact that people are being observed, but the use of one-way mirrors and video recordings can help people to become comfortable with the setting. Researchers who conduct an observational study out in the real world have even more possibilities for blending into the background, including using observers who are literally hidden. For example, someone hypothesizes that people are more likely to pick up garbage when the weather is nicer. Rather than station an observer with a clipboard by the trash can, the researcher could place someone out of sight behind a tree, or perhaps sitting on a park bench pretending to read a magazine. In both cases, people would be less conscious of being observed and therefore more likely to behave naturally.

One extremely clever strategy for blending in comes from a study by the social psychologist Muzafer Sherif et al. (1954), involving observations of cooperative and competitive behaviors among boys at a summer camp. For Sherif, it was particularly important to make observations in this context without the boys realizing they were part of a research study. Sherif took on the role of camp janitor, which allowed him to be a presence in nearly all of the camp activities. The boys never paid enough attention to the "janitor" to realize his omnipresence—or his discreet note-taking. The brilliance of this idea is that it takes advantage of the fact that people tend to blend into the background once we become used to their presence.

Types of Observational Research

Several variations of observational research exist, according to the amount of control that a researcher has over the data collection process. **Structured observation** involves creating a standard situation in a controlled setting and then observing participants' responses to a predetermined set of events. The "strange situation" studies of parent–child attachment (discussed above) are a good example of structured observation—mothers and infants are subjected to a series of eight structured episodes, and researchers systematically observe and record the infants' reactions. Even though these types of studies are conducted in a laboratory, they differ from experimental studies in an important way: Rather than

systematically manipulate a variable to make comparisons, researchers present the same set of conditions to all participants.

Another example of structured observation comes from the research of John Gottman, a psychologist at the University of Washington. For nearly three decades, Gottman and his colleagues have conducted research on the interaction styles of married couples. Couples who take part in this research are invited for a three-hour session in a laboratory that closely resembles a living room. Gottman's goal is to make couples feel reasonably comfortable and natural in the setting to get them talking as they might do at home. After allowing them to settle in, Gottman adds the structured element by asking the couple to discuss an "ongoing issue or problem" in their marriage. The researchers then sit back to watch the sparks fly, recording everything from verbal and nonverbal communication to measures of heart rate and blood pressure. Gottman has observed and tracked so many couples over the decades that he is able to predict, with remarkable accuracy, which couples will divorce in the 18 months following the lab visit (Gottman & Levenson, 1992).

Naturalistic observation, meanwhile, involves observing and systematically recording behavior in the real world. This can be conducted in two broad ways—with or without intervention on the part of the researcher. *Intervention* in this context means that the researcher manipulates some aspect of the environment and then observes people's responses. For example, a researcher might leave a shopping cart just a few feet away from the cart-return area and track whether people move the cart. (Given the number of carts that are abandoned just inches away from their proper destination, someone must be doing this research all the time.) Recall an example from Chapter 1 (the discussion of ethical dilemmas in section 1.5) in which Harari et al. (1995) used naturalistic observation to study whether people would help in emergency situations. In brief, these researchers staged what appeared to be an attempted rape in a public park and then observed whether groups or individual males were more likely to rush to the victim's aid.

The ABC network has developed a hit reality show that mimics this type of research. The show, *What Would You Do?*, sets up provocative situations in public settings and videotapes people's reactions. An unwitting participant in one of these episodes might witness a customer stealing tips from a restaurant table, or a son berating his father for being gay, or a man proposing to his girlfriend who minutes earlier had been kissing another man at the bar. Of course, these observation "studies" are more interested in shock value than data collection (or Institutional Review Board [IRB] approval; see Section 1.5), but the overall approach can be a useful strategy to assess people's reactions to various situations. In fact, some of the scenarios on the show are based on classic studies in social psychology, such as the well-documented phenomenon that people are reluctant to take responsibility for helping in emergencies.

Alternatively, naturalistic studies can involve simply recording ongoing behavior without any attempt by the researchers to intervene or influence the situation. In these cases, the goal is to observe and record behavior in a completely natural setting. For example, researchers might station themselves at a liquor store and observe the numbers of men and women who buy beer versus wine. Or, they might observe the numbers of people who give money to the Salvation Army bell-ringers during the holiday season. A researcher can use this approach to compare different conditions, provided the differences occur naturally. That is, researchers could observe whether people donate more money to the Salvation Army on sunny or snowy days, or compare donation rates when the bell ringers are different genders or races. Do people give more money when the bell-ringer is an attractive female? Or do they give more to someone who looks needier? These are all research questions that could be addressed using a well-designed naturalistic observation study.

Finally, **participant observation** involves having the researcher(s) conduct observations while engaging in the same activities as the participants. The goal is to interact with these participants to gain better access and insight into their behaviors. In one famous example, the psychologist David Rosenhan (1973) was interested in the experience of people hospitalized for mental illness. To study these experiences, he had eight perfectly sane people gain admission to different mental hospitals. These fake patients were instructed to give accurate life histories to a doctor but lie about one diagnostic symptom. They all claimed to hear an occasional voice saying the words "empty," "hollow," and "thud." Such auditory hallucinations are a symptom of schizophrenia, and Rosenhan chose these words to vaguely suggest an existential crisis.

Once admitted, these "patients" behaved in a normal and cooperative manner, with instructions to convince hospital staff that they were healthy enough to be released. In the meantime, they observed life in the hospital and took notes on their

experiences—a behavior that many doctors interpreted as "paranoid note-taking." The main finding of this study was that hospital staff tended to view all patient behaviors through the lens of their initial diagnoses. Despite immediately acting "normally," these fake patients were hospitalized an average of 19 days (with a range from 7 to 52) before being released. All but one was diagnosed with "schizophrenia in remission" upon release. Rosenhan's other striking finding was that treatment was generally depersonalized, with staff spending little time with individual patients.

RENARD/BSIP/Superstock

Psychologists David Rosenhan's study of staff and patients in a mental hospital found that patients tended to be treated based on their diagnosis, not on their actual behavior.

In another example of participant observation, Festinger, Riecken, and Schachter (1956) decided to join a doomsday cult to test their new theory of cognitive dissonance. Briefly, this theory argues that people are motivated to maintain a sense of consistency among their various thoughts and behaviors. So, for example, a person who smokes a cigarette despite being aware of the health risks might rationalize smoking by convincing herself that lung-cancer risk is really just genetic. In this case, Festinger and colleagues stumbled upon the case of a woman named Mrs. Keach, who was predicting the end of the world, via alien invasion, at 11 p.m. on a specific date six months in the future. What would happen, they wondered, when this prophecy failed to come true? (One can only imagine how shocked they would have been had the prophecy turned out to be correct.)

To answer this question, the researchers pretended to be new converts and joined the cult, living among the members and observing them as they made their preparations for doomsday. Sure enough, the day came, and 11 p.m. came and went without the world ending. Mrs. Keach first declared that she had forgotten to account for a time-zone difference, but as sunrise started to approach, the group members became restless. Finally, after a short absence to communicate with the aliens, Mrs. Keach returned with some good news: The aliens were so impressed with the devotion of the group that they decided to postpone their invasion. The group members rejoiced, rallying around this brilliant piece of rationalizing, and quickly began a new campaign to recruit new members.

As these examples illustrate, participant observation can provide access to amazing and one-of-a-kind data, including insights into group members' thoughts and feelings. This approach also provides access to groups that might be reluctant to allow outside observers. However, the participant approach has two clear disadvantages over other types of observation. The first problem is ethical; data are collected from individuals who do not have the opportunity to give informed consent. Indeed, the whole point of the technique is to observe people without their knowledge. Before an IRB can approve this kind of study, researchers must show an extremely compelling reason to ignore informed consent, as well as extremely rigorous measures to protect identities. The second problem is methodological; the approach provides ample opportunity for the objectivity of observations to be compromised by the close contact between researcher and participant. Because the researchers are a part of the group, they can change the dynamics in subtle ways, possibly leading the group to confirm their hypothesis. In addition, the group can shape the researchers' interpretations in subtle ways, leading them to miss important details.

Another spin on participant observation is called **ethnography**, or the scientific study of the customs of people and cultures. This is very much a qualitative method that focuses on observing people in the real world and learning about a culture from the perspective of the person being studied—that is, learning from the ground up rather than testing hypotheses. Ethnography is used primarily in other social-science fields, such as anthropology. In one famous example, the cultural anthropologist Margaret Mead (1928) used this approach to shed light on differences in social norms around adolescence between American and Samoan societies. Mead's conclusions were based on interviews she conducted over a six-month period, observing and living alongside a group of 68 young women. Mead concluded from these interviews that Samoan children and adolescents are largely ignored until they reach the age of 16 and become full members of society. Among her more provocative claims was the idea that Samoan adolescents were much more liberal in their sexual attitudes and behaviors than American adolescents.

Mead's work has been the subject of criticism by a handful other anthropologists, one of whom has even suggested that Mead was taken in by an elaborate joke played by the group of young girls. Still others have come to Mead's rescue and challenged the critics' interpretations. The nature of this debate between Mead's critics and her supporters highlights a distinctive characteristic of qualitative methods: "Winning" the argument is based on challenging interpretations of the original interviews and observations. In contrast, disagreements around quantitative methods are generally based on examining statistical results from hypothesis testing. While quantitative methods may lose much of the richness of people's experiences, they do offer an arguably more objective way of settling theoretical disputes.

Steps in Observational Research

One of the major strengths of observational research is its high degree of **ecological validity**; that is, the research can be conducted in situations that closely resemble the real world. Think of the chapter examples so far—married couples observed in a living-room-like laboratory; doomsday cults observed from within; bullying behaviors on the school playground. In every case, people's behaviors are observed in the natural environment or something very close to it. However, this ecological validity comes at a price; the real world is a jumble of information, some relevant, some not so much. The challenge for researchers, then, is to decide on a system that provides the best test of their hypothesis, one that can sort out the signal from the noise. This section discusses a three-step process for conducting observational research. The key point to note right away is that most of this process involves making decisions ahead of time so that the process of data collection is smooth, simple, and systematic.

Step 1—Develop a Hypothesis

For research to be systematic, it is important to impose structure by having a clear research question, and, in the case of quantitative research, a clear hypothesis as well. Other chapters have covered hypotheses in detail, but the main points bear repeating: A hypothesis must be testable and falsifiable, meaning that it must be framed in such a way that it can be addressed through empirical data and might be disconfirmed by these data. In the example involving Salvation Army donations, we predicted that people might donate more money to an attractive bell-ringer. This hypothesis could easily be tested empirically and could just as easily be disconfirmed by the right set of data—say, if attractive bell-ringers brought in the *fewest* donations.

This particular example also highlights an additional important feature of observational hypotheses; namely, they must be based on observable behaviors. That is, we can safely make predictions about the amount of money people will donate because we can directly observe it. We are, nonetheless, unable to make predictions in this context about the *reasons* for donations. We would have no way to observe, say, that people donate more to attractive bell-ringers because they are trying to impress them. In sum, one limitation of observing behavior in the real world is that it prevents researchers from delving into the cognitive and motivational reasons behind the behaviors.

Step 2—Decide What and How to Sample

Once a researcher has developed a hypothesis that is testable, falsifiable, and observable, the next step is to decide what kind of information to gather from the environment to test this hypothesis. The simple fact is that the world is too complex to sample everything. Imagine that someone wanted to observe the dinner rush at a restaurant. A nearly infinite list of possibilities for observation presents itself: What time does the restaurant get crowded? How often do people send their food back to the kitchen? What are the most popular dishes? How often do people get in arguments with the wait staff? To simplify the process of observing behavior, the researcher will need to take a **sample**, or a smaller portion of the population, that is relevant to the hypothesis. That is, rather than observing "dinner at the restaurant," the researcher's goal is to narrow his or her focus to something as specific as "the number of people waiting in line for a table at 6 p.m. versus 9 p.m."

The choice of what and how to sample will ultimately depend on the best fit for the hypothesis. The context of observational research offers three strategies for sampling behaviors and events. The first strategy, **time sampling**, involves

comparing behaviors during different time intervals. For example, to test the hypothesis that football teams make more mistakes when they start to get tired, researchers could count the number of penalties in the first five minutes and the last five minutes of the game. This data would allow researchers to compare mistakes at one time interval with mistakes at another time interval. In the case of Festinger's (1956) study of a doomsday cult, time sampling was used to compare how the group members behaved before and after their prophecy failed to come true.

The second strategy, **individual sampling**, involves collecting data by observing one person at a time to test hypotheses about individual behaviors. Many of the examples already discussed involve individual sampling: Ainsworth and colleagues (1970) tested their hypotheses about attachment behaviors by observing individual infants, while Gottman (1992) tests his hypotheses about romantic relationships by observing one married couple at a time. These types of data allow researchers to examine behavior at the individual level and test hypotheses about the kinds of things people do—from the way they argue with their spouses to whether they wear team colors to a football game.

The third strategy, **event sampling**, involves observing and recording behaviors that occur throughout an event. For example, we could track the number of fights that break out during an event such as a football game, or the number of times people leave the restaurant without paying the check. This strategy allows for testing hypotheses about the types of behaviors that occur in a particular environment or setting. For instance, a researcher might compare the number of fights that break out in a professional football versus a professional hockey game. Or, the next time we host a party, we could count the number of wine bottles versus beer bottles that end up in the recycling bin. The distinguishing feature of this strategy is its focus on *occurrence* of behaviors more than on the *individuals* performing these behaviors.

Steve Mason/Photodisc/Thinkstock

The dinner scene at a busy restaurant offers a wide variety of behaviors to observe. In order to simplify the observation process, researchers should narrow the focus by taking a sample.

Step 3—Record and Code Behavior

Having formulated a hypothesis and decided on the best sampling strategy, researchers must perform one final and critical step before beginning data collection. Namely, they have to develop good operational definitions of the variables by translating the underlying concepts into *measurable* variables. Gottman's research turns the concept of marital interactions into a range of measurable variables, such as the number of dismissive comments and passive-aggressive sighing—all things that can be observed and counted objectively. Rosenhan's 1973 study involving fake schizophrenic patients turned the concept of patient experience into measureable variables such as the amount of time staff members spent with each patient—again, something very straightforward to observe.

It is vital that researchers decide up front what kinds and categories of behavior they will be observing and recording. In the last section, we narrowed down our observation of dinner at the restaurant to the number of people in line at 6 p.m. versus the number of people in line at 9 p.m. But how can we be sure of an accurate count? What if two people are waiting by the door while the other two members of the group are sitting at the bar? Are those at the bar waiting for a table or simply having drinks? One possibility might be to count the number of individuals who walk through the door in different time periods, although our count could be inflated by those who give up on waiting or who only enter to sneak in and out of the restroom.

In short, observing behavior in the real world can be messy. The best way to deal with this mess is to develop a clear and consistent categorization scheme and stick with it. That is, in testing a hypothesis about the most crowded time at a restaurant, researchers would choose one method of counting people and use it for the duration of the study. In part, this

choice of a method is a judgment call, but researchers' judgment should be informed by three criteria. First, they should consider *practical issues*, such as whether their categories can be directly observed. A researcher can observe the number of people who leave the restaurant but cannot observe whether they got impatient. Second, they should consider *theoretical issues*, such as how well the categories represent the underlying theory. Why did researchers decide to study the most crowded time at the restaurant? Perhaps this particular restaurant is in a new, up-and-coming neighborhood, and they expect the restaurant to become crowded over the course of the evening. The time would also lead researchers to include people sitting both at tables and at the bar—because this crowd may come to the restaurant with the sole intention of staying at the bar. Finally, researchers should consider *previous research* in choosing their categories. Have other researchers studied dining patterns in restaurants? What kinds of behaviors did they observe? If these categories make sense for the project, researchers may feel free to re-use them—no need to reinvent the wheel.

Last but not least, a researcher should take a step back and evaluate both the validity and the reliability of the coding system. (See Section 2.2 for a review of these terms.) Validity in this case means making sure the categories capture the underlying variables in the hypothesis (i.e., construct validity; see Section 2.2). For example, in Gottman's studies of marital interactions, some of the most important variables are the emotions expressed by both partners. One way to observe emotions would be to count the number of times a person smiles. However, we would have to think carefully about the validity of this measure, because smiling could indicate either genuine happiness or condescension. As a general rule, the better and more specific researchers' operational definitions, the more valid their measures will be (Chapter 2).

Reliability in this context means making sure data are collected in a consistent way. If research involves more than one observer using the same system, their data should look roughly the same (i.e., interrater reliability). This reliability is accomplished in part by making the observation task simple and straightforward—for example, having trained assistants use a checklist to record behaviors rather than depending on open-ended notes. The other key to improving reliability is careful training of the observers, giving them detailed instructions and ample opportunities to practice the rating system.

Observation Examples

To explain how all of this comes together, we will explore a pair of examples, from research question to data collection.

Example 1—Theater Restroom Usage

First, imagine, for the sake of this example, that someone is interested in whether people are more likely to use the restroom before or after watching a movie. Such a research question could provide valuable information for theater owners in planning employee schedules (i.e., when are bathrooms most likely to need cleaning). Thus, studying patterns of human behavior results in valuable applied knowledge.

The first step is to develop a specific, testable, and observable hypothesis. In this case, we might predict that people are more likely to use the restroom after the movie, as a result of consuming those 64-ounce sodas during the movie. Just for fun, we will also compare the restroom usage of men and women. Perhaps men are more likely to wait until after the movie, whereas women are just as likely to go before as after? This pattern of data might look something like the percentages in Table 3.1. That is, men make 80% of their restroom visits after the movie and 20% before the movie, while women make about 50% of their restroom visits at each time.

Table 3.1: Hypothesized restroom visits

Gender	Men	Women
Before movie	20%	50%
After movie	80%	50%
Total	100%	100%

The next step is to decide on the best sampling strategy to test this hypothesis. Of the three sampling strategies discussed—individual, event, and time—which one seems most relevant here? The best option would probably be time sampling because the hypothesis involves comparing the number of restroom visitors in two time periods (before versus after the movie). So, in this case, we would need to define a time interval for collecting data. We could limit our observations to the 10 minutes before the previews begin and the 10 minutes after the credits end. The potential problem here, of course, is that some people might use either the previews or the end credits as a chance to use the restroom. Another complication arises in trying to determine which movie people are watching; in a giant multiplex theater, movies start just as others are finishing. One possible solution, then, would be to narrow the sample to movie theaters that show only one movie at a time and to define the sampling times based on the actual movie start- and end-times.

Having determined a sampling strategy, the next step is to identify the types of behaviors we want to record. This particular hypothesis poses a challenge because it deals with a rather private behavior. To faithfully record people "using the restroom," we would need to station researchers in both men's and women's restrooms to verify that people actually, well, "use" the restroom while they are in it. However, this strategy poses the potential downside that the researcher's presence (standing in the corner of the restroom) will affect people's behavior. Another, less intrusive option would be to stand outside the restroom and simply count "the number of people who enter." The downside to *that*, of course, is that we technically do not know why people are going into the restroom. But sometimes research involves making these sorts of compromises—in this case, we chose to sacrifice a bit of precision in favor of a less-intrusive measurement. This compromise would also serve to reduce ethical issues with observing people in the restroom.

So, in sum, we started with the hypothesis that men are more likely to use the restroom after a movie, while women use the restroom equally before and after. We then decided that the best sampling strategy would be to identify a movie theater showing only one movie and to sample from the 10-minute periods before and after the actual movie's running time. Finally, we decided that the best strategy for recording behavior would be to station observers outside the restrooms and count the number of people who enter. Now, say we conduct these observations every evening for one week and collect the data in Table 3.2.

Table 3.2: Findings from observing restroom visits

Gender	Men	Women
Before movie	75 (25%)	300 (60%)
After movie	225 (75%)	200 (40%)
Total	300 (100%)	500 (100%)

Notice that more women (N = 500) than men (N = 300) attended the movie theater during our week of sampling. The real test of our hypothesis, however, comes from examining the percentages within gender groups. That is, of the 300 men who went into the restroom, what percentage of them did so before the movie and what percentage of them did so after the movie? In this dataset, women used the restroom with relatively equal frequency before (60%) and after (40%) the movie. Men, in contrast, were three times as likely to use the restroom after (75%) than before (25%) the movie. In other words, our hypothesis appears to be confirmed by examining these percentages.

Example 2—Cell Phone Usage While Driving

Imagine that we are interested in patterns of cell phone usage among drivers. Several recent studies have reported that drivers using cell phones are as impaired as drunk drivers, making this an important public safety issue. Thus, if we could understand the contexts in which people are most likely to use cell phones, it would provide valuable information for developing guidelines for safe and legal use of these devices. So, this study might count the number of drivers using cell phones in two settings: while navigating rush-hour traffic and while moving on the freeway.

The first step is to develop a specific, testable, and observable hypothesis. In this case, we might predict that people are more likely to use cell phones when they are bored in the car. So, we hypothesize that we will see more drivers using cell

phones while stuck in rush-hour traffic than while moving on the freeway.

The next step is to decide on the best sampling strategy to test this hypothesis. Of the three sampling strategies discussed—individual, event, and time—which one seems most relevant here? The best option would probably be individual sampling because we are interested in the cell phone usage of individual drivers. That is, for each individual car we see during the observation period, we want to know whether the driver is using a cell phone. One strategy for collecting these observations would be to station observers along a fast-moving stretch of freeway, as well as along a stretch of road that is clogged during rush hour. These observers would keep a record of each passing car and note whether the driver is on the phone.

After selecting a sampling strategy, we next must decide the types of behaviors to record. One challenge this study presents is how broadly to define cell phone usage. Should we include both talking and text messaging? Given our interest in distraction and public safety, we probably want to include text messaging. Several states have recently banned this practice while driving, often in response to tragic accidents. Because we will be observing moving vehicles, the most reliable approach might be to simply note whether drivers have a cell phone in their hand. As with the restroom study, we sacrifice a little bit of precision (i.e., knowing what the driver is using the cell phone for) to capture behaviors that are easier to record.

To sum up, we started with the hypothesis that drivers would be more likely to use cell phones when stuck in traffic. We then decided that the best sampling strategy would be to station observers along two stretches of road who would note whether drivers were using cell phones. Finally, we decided that the cell phone usage would be defined as each driver holding a cell phone. Now, suppose we conducted these observations over a 24-hour period and collected the data in Table 3.3.

Table 3.3: Findings from observing cell phone usage

	Rush Hour	Highway
Cell Phone	30 (30%)	200 (67%)
No Cell Phone	70 (70%)	100 (33%)
Total	100 (100%)	300 (100%)

The results show that more cars passed by on the highway ($N = 300$) than on the street during the rush-hour stretch ($N = 100$). The real test of our hypothesis, though, comes from examining the percentages within each stretch. That is, of the 100 people observed during rush hour and the 300 observed on the highway, what percentage was using cell phones? In this data set, 30% of those in rush hour were using cell phones, compared with 67% of those on the highway. In other words, the data did not confirm our hypothesis. Drivers in rush hour were less than half as likely to be using cell phones. The next step in this research program would be to speculate on the reasons the data contradicted the hypothesis.

Qualitative versus Quantitative Approaches

The general method of observation lends itself equally well to qualitative and quantitative approaches, although some types of observation fit one approach better than the other. For example, structured observation tends to focus on hypothesis testing and quantification of responses. In Mary Ainsworth's (1970) "strange situation" research (described previously), the primary goal was to expose children to a predetermined script of events and to test hypotheses about how children with secure and insecure attachments would respond to these events. In contrast, naturalistic observation—and, to a greater extent, participant observation—tends to focus on learning from events as they occur naturally. In Leon Festinger's "doomsday cult" study, the researchers joined the group to observe the ways members reacted when their prophecy failed to come true. Margaret Mead (1928) spent several months living with Samoan adolescents to understand social norms around coming of age.

Research: Thinking Critically

"Irritable Heart" Syndrome in Civil War Veterans

Follow the link below to an article by science writer and editor K. Kris Hirst. In this article, Hirst reviews compelling research from health psychologist Roxanne Cohen Silver and her colleagues at the University of California, Irvine. Cohen Silver and her colleagues reviewed the service records of 15,027 Civil War veterans, finding an astounding rate of mental illness—long before post-traumatic stress disorder was recognized. As you read the article, consider what you have learned so far about the research process, and then respond to the questions below.

http://psychology.about.com/od/ptsd/a/irritableheart.htm (http://psychology.about.com/od/ptsd/a/irritableheart.htm)

Think about it:

1. What hypotheses are the researchers testing in this study?
2. How did the researchers quantify trauma experienced by Civil War soldiers? Do you think this is a valid way to operationalize trauma? Explain why or why not.
3. Would this research be best described as case studies, archival research, or natural observation? Does the study involve elements of more than one type? Explain.

3.5 Describing Your Data

Before we move on from descriptive research designs, this last section discusses the process of presenting descriptive data in both graphical and numeric form. No matter how the researcher presents data, a good description is accurate, concise, and easy to understand. In other words, researchers have to represent the data accurately and in the most efficient way possible so that their audience can understand it. Another, more eloquent way to think of these principles is to take the advice of Edward Tufte, a statistician and expert in the display of visual information. Tufte (2001) suggests that when people view visual displays, they should spend time on *content-reasoning* rather than *design-decoding*. The sole purpose of designing visual presentations is to communicate information. So, the audience should spend time thinking about the information being presented, not trying to puzzle through the display itself. The following sections explain guidelines for accomplishing this goal in both numeric and visual form.

Table 3.4 presents hypothetical data from a sample of 20 participants. In this example, we have asked people to report their gender and ethnicity, as well as answer questions about their overall life satisfaction and daily stress. Each row in this table represents one participant in the study, and each column represents one of the variables for which data were collected. This chapter focuses on ways to describe the sample characteristics. Later chapters will return to these principles in discussing graphs that display the relationship between two or more variables.

Table 3.4: Raw data from a sample of 20 individuals

Subject ID	Gender	Ethnicity	Life satisfaction	Daily stress
1	Male	White	40	10
2	Male	White	47	9
3	Female	Asian	29	8
4	Male	White	32	9
5	Female	Hispanic	25	3
6	Female	Hispanic	35	3
7	Female	White	28	8
8	Male	Hispanic	40	9
9	Male	Asian	37	10
10	Female	African-American	30	10
11	Male	White	43	8
12	Male	Asian	40	4
13	Male	White	48	7
14	Female	African-American	30	4
15	Female	White	37	7
16	Male	Hispanic	40	1
17	Female	White	36	1
18	Male	African-American	45	8
19	Female	White	42	8
20	Female	African-American	38	7

Numeric Descriptions

Because psychology is a scientific discipline, it often expresses preference for presenting data in number form. These numbers provide a metric that can be used to compare findings from one study to another, to evaluate the overall consistency of whatever phenomenon is being studied. Following is a brief overview of some common numeric descriptors for data.

Frequency Tables

Often, a good first step in approaching a data set is to obtain a sense of the frequencies for demographic variables—in this example, gender and ethnicity. The **frequency tables** shown in Table 3.5 are designed to present the number and percentage of the sample that fall into each of a set of categories. As this pair of tables shows, the sample consisted of an equal number of men and women (i.e., 50% for each gender). The majority of participants were White (45%), with the remainder divided almost equally between African-American (20%), Asian (15%), and Hispanic (20%) ethnicities.

Table 3.5: Frequency table summarizing ethnicity and sex distribution

Gender	**Frequency**	**Percentage**
Female	10	50.0
Male	10	50.0
Total	20	100.0
Ethnicity	**Frequency**	**Percentage**
African-American	4	20.0
Asian	3	15.0
Hispanic	4	20.0
White	9	45.0
Total	20	100.0

Researchers can gain a lot of information from numerical summaries of data. In fact, numeric descriptors form the starting point for doing inferential statistics and testing hypotheses. A statistics course explores these statistics in detail, but for now it is important to understand that two numeric descriptors can provide a wealth of information about a data set: measures of central tendency and measures of dispersion.

Measures of Central Tendency

The first number we need to describe our data is a measure of **central tendency**, which represents the most typical case in our data set. Central tendency is a single number that provides an overall sense of all the numbers. Think of what happens when colors are mixed: Adding yellow to blue creates green, so green gives us an overall sense of the combination of the two colors. In the same way, think of a household where one parent has a high salary, another has a moderate salary, and a teenager makes minimum wage. Taking the average of all three gives us an overall sense of the income for the entire household.

Central tendency can be represented by these three indices:

The **mean** is the mathematical average of a data set, calculated by adding up all the scores in the data set and then dividing this total by the number of scores in the data set. Because we are adding and dividing our scores, the mean can only be calculated using interval or ratio data (see Chapter 2 for a review of the four scales of measurement).

The **median**, another measure of central tendency, represents the number in the middle of a dataset, with 50% of scores both above and below it. The median is identified by placing the list of values in ascending numeric order, then selecting the number in the middle. This measure of central tendency can be used for ordinal, interval, or ratio data because it does not require mathematical manipulation to obtain.

The final measure of central tendency, the **mode**, represents the most frequent score in a data set and is obtained either by visual inspection of the values or by consulting a frequency table like in the one in Table 3.5. Because the mode represents a simple frequency count, it can be used with any of the four scales of measurement. In addition, it is the only measure of central tendency that is valid for use with nominal data—that is, those that do not have a numerical value—since the numbers assigned to these data are arbitrary.

One important takeaway is that the scale of measurement largely dictates the choice between measures of central tendency—nominal scales can only use the mode, and only interval or ratio scales can use the mean. (For a review of these scales of measurement, see Chapter 2, Section 2.3.) The other piece of the puzzle is to consider which measure best represents the data. Remember that the central tendency is a way to represent the "typical" case using a single number, so the goal is to settle on the most representative number. The examples in Table 3.6 illustrate this process.

Table 3.6: Comparing the mean, median, and mode

Data	Mean	Median	Mode	Discussion
1,2,3,4,5,11,11	5.29	4	11	• Both the mean and the median seem to represent the data fairly well. • The mean is a slightly better choice because it hints at the higher scores. • The mode is not representative—two people seem to have higher scores than everyone else.
1,1,1,5,10,10,100	18.29	5	1	• The mean is inflated by the atypical score of 100 and therefore does not represent the data accurately. • The mode is also not representative because it ignores the higher values. • In this case, the median is the most representative value to describe this dataset.

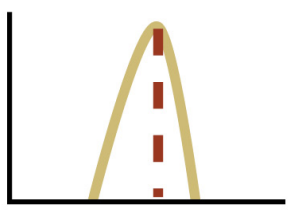
Measures of Dispersion

The second measure used to describe a dataset is a measure of **dispersion**, or the spread of scores around the central tendency—also referred to as measures of "variability." Measures of dispersion tell us just how typical the typical score is. If the dispersion is low, then scores are clustered tightly around the central tendency; if dispersion is higher, then the scores stretch out farther from the central tendency. Figure 3.2 presents a conceptual illustration of dispersion. The graph on the left has a low amount of dispersion because the scores (i.e., the yellow curve) cluster tightly around the average value (i.e., the red dotted line). The graph on the right shows a high amount of dispersion because the scores (yellow curve) spread out widely from the average value (red dotted line). The graph on the right might represent the earlier example of household income: The average income represents all three family members, but between the high-earning parent and the minimum-wage-earning teenager is a fairly large spread.

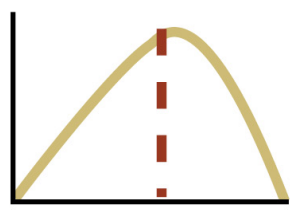
Figure 3.2: Two distributions with a low versus high amount of dispersion

One of the most straightforward measures of dispersion is the **range**, which is the difference between the highest and lowest scores. In Table 3.6, the range of the first dataset would be found by simply subtracting the lowest value (1) from the highest value (11), to get a range of 10.

The range is useful in giving a general idea of the spread of scores, although it does not say much about how tightly these scores cluster around the mean.



Low Amount of Dispersion Around the Mean (red dotted line)



High Amount of Dispersion Around the Mean (red dotted line)

The most common measures of dispersion are the **variance** and standard deviation, both of which represent the average difference between the mean and each individual score. The variance is calculated by subtracting each score from the mean to obtain a **deviation score**, squaring and summing these individual deviation scores, and then dividing by the sample size. The more scores are spread out around the mean, the higher the sum of these deviation scores will be, and therefore the higher the variance will be. Another common measure, the standard deviation (*SD*), is calculated as the square root of the variance.

Once we know the central tendency and the dispersion of variables, we have a good sense of what the sample looks like. These numbers also provide a valuable part in calculating the inferential statistics that we ultimately use to test our hypotheses.

Standard Scores

So far we have discussed ways to describe a particular sample in numeric terms. What do we do when we want to compare results from different samples—or from studies using different scales? Say we want to compare the anxiety levels of two people; unfortunately, in this example, these people were measured using different anxiety scales:

> Joe scored 25 on the ABC Anxiety Scale, which has a mean of 15 and a standard deviation of 2.
>
> Deb scored 40 on the XYZ Anxiety Scale, which has a mean of 30 and a standard deviation of 10.

At first glance, Deb's anxiety score appears higher, but note that the scales have different properties: The ABC scale has an average score of 15, while the XYZ scale has a higher average score of 30. The dispersion of these scales is also different; scores on the ABC scale cluster more tightly around the mean (i.e., the standard deviation is 2 compared to 10 on the XYZ scale).

The solution for comparing these scores is to convert both of them to **standard scores** (often expressed as ***z* scores**), which represent the distance of each score from the sample mean, expressed in standard deviation units. Standard scores let researchers translate raw scores into distributions with a predefined mean and standard deviation for easier interpretation. For example, scores on IQ tests are converted (i.e., *standardized*) onto a scale that has a mean of 100 and a standard deviation of 15. This tells us that a person with an IQ score of 100 is right at the average for the population, while someone with a score of 130 is two standard deviations above average.

The formula for a *z* score is worth examining in greater detail, as a way to understand the broader concept. Memorizing or using the formula in this research methods course is not required. The formula for a *z* score is:

$$z = (x - M)/SD$$

This formula subtracts the mean (*M*) from the individual score (*x*) and then divides this difference by the standard deviation of the sample (*SD*). To compare Joe's score with Deb's score, we simply substitute the appropriate numbers, using the mean and standard deviation from the scale that each one completed. This enables us to place scores from very different distributions on the same scale, making them easier to compare with one another. So, in this case:

$$\text{Joe: } z = (x - M)/SD = (25 - 15)/2 = 10/2 = 5$$

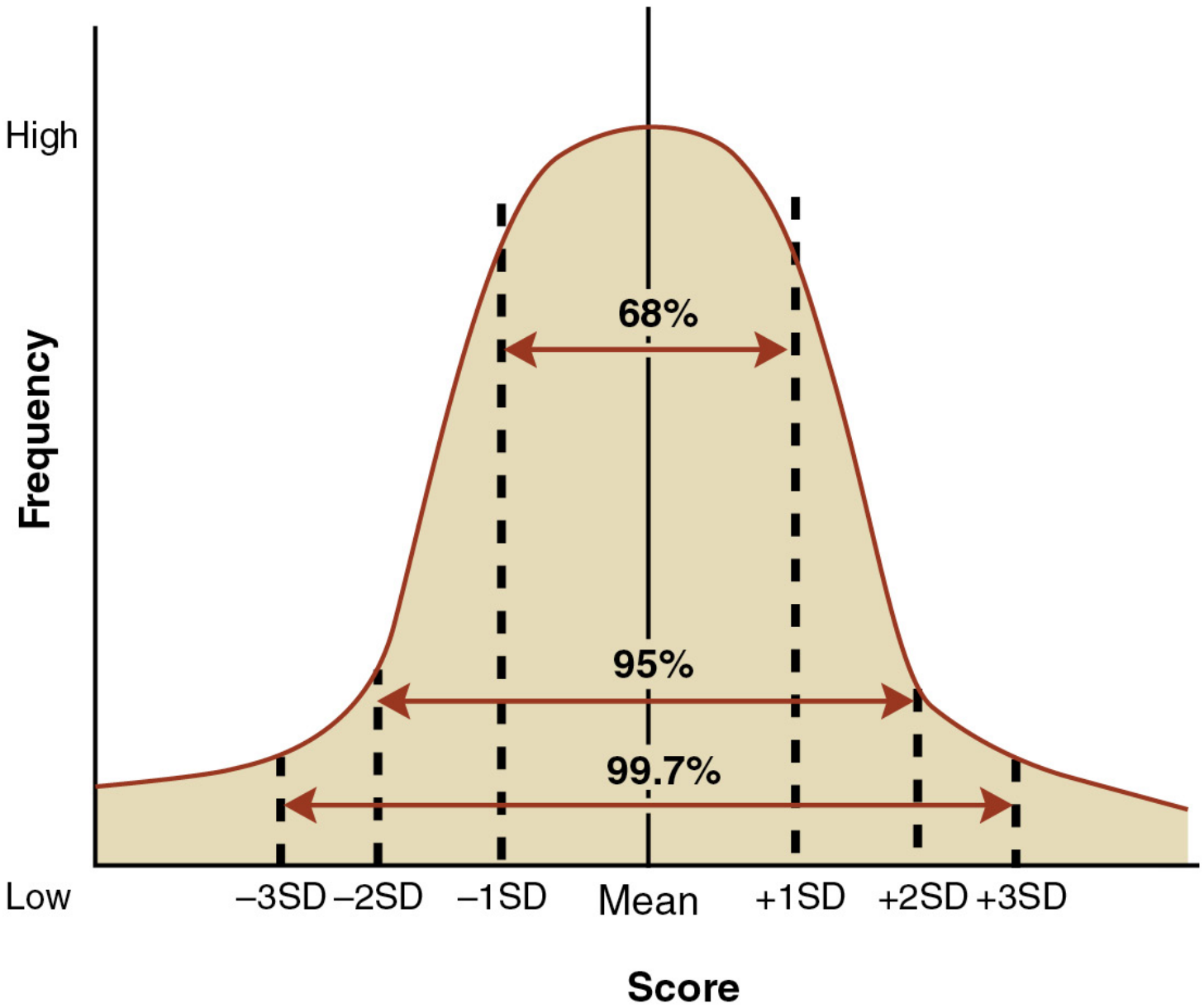
$$\text{Deb: } z = (x - M)/SD = (40 - 30)/10 = 10/10 = 1$$

The resulting scores represent each person's score in standard deviation terms: Joe is 5 standard deviations above the mean of the ABC scale, while Deb is only 1 standard deviation above the mean of the XYZ scale. Or, in plain English, Joe is considerably more anxious than Deb.

To understand just *how* anxious Joe is, it is helpful to know a bit about why this technique works. Anyone who has taken a statistics class will have encountered the concept of the **normal distribution** (or "bell curve"), a symmetric distribution with an equal number of scores on either side of the mean, as Figure 3.3 illustrates.

It turns out that many variables in the social and behavioral sciences fit this normal distribution, provided the sample sizes are large enough. A normal distribution is useful because it has a consistent set of properties, such as having the same value for mean, median, and mode. In addition, if the distribution is normal, each standard deviation cuts off a known percentage of the curve, as illustrated in Figure 3.3. That is, 68% of scores will fall within ±1 standard deviation of the mean; 95% of scores will fall within ± two standard deviations; and 99.7% of scores will fall within ± three standard deviations.

Figure 3.3: Standard deviations and the normal distribution



These percentages allow us to understand individual data points in even more useful ways, because we can easily move back and forth between *z* scores, percentages, and standard deviations. Take the example of Joe and Deb's anxiety scores: Deb has a *z* score of 1, which means her anxiety is 1 standard deviation above the mean. Furthermore, as we can see by consulting the normal distribution (Figure 3.3), her anxiety level is higher than 84% of the population. Joe has a *z* score of 5, which means his anxiety is 5 standard deviations above the mean. This also means that his anxiety is higher than 99.999% of the population. (For a handy online calculator that converts between *z* scores and percentages, see: **http://www.measuringusability.com/pcalcz.php (http://www.measuringusability.com/pcalcz.php)** .)

Discussions of intelligence test scores also commonly use the relationship between *z* scores and percentiles. Tests that

purport to measure IQ are converted to a scale that has a mean of 100 and a standard deviation of 15. Because IQ is normally distributed, we can move easily back and forth between *z* scores and percentages. For example, someone who has an IQ test score of 130 falls 2 standard deviations above the mean and falls in the upper 2.5% of the population. A person with an IQ test score of 70 is 2 standard deviations below the mean and thus falls in the bottom 2.5% of the population.

Ultimately, the use of standard scores allows us to take data that have been collected on different scales—perhaps in different laboratories and different countries—and place them on the same metric for comparison. As we have discussed in several contexts, science is all about the accumulation of knowledge one study at a time. The best support for an idea comes when data from different researchers, using different measures to capture the same concept, back the idea. The ability to convert these different measures back to the same metric is an invaluable tool for researchers who want to compare research results.

Visual Descriptions

Displaying data in visual form is often one of the most effective ways to communicate findings—as the cliché goes, a picture is worth a thousand words. What sort of visual should a researcher use? The choice of graphs is guided by two criteria: the scale of measurement and the best fit for the results. This section introduces some of the most common visual displays, based on hypothetical data used in Table 3.4.

Displaying Frequencies

One common type of graph is the **bar graph**, which also summarizes the frequency of data by category. Figure 3.4a depicts a bar graph, showing four categories of ethnicity along the horizontal axis and the number of people falling into each category indicated by the height of the bars. So, for example, this sample contains nine White participants and four Hispanic participants. Notice that these bar graphs contain exactly the same information as the frequency table in Table 3.5. When reporting results in a paper, a researcher would, of course, use only one of these methods. More often than not, graphical displays are the most effective way to communicate information.

Figure 3.4a: Bar graph displaying frequency by ethnicity

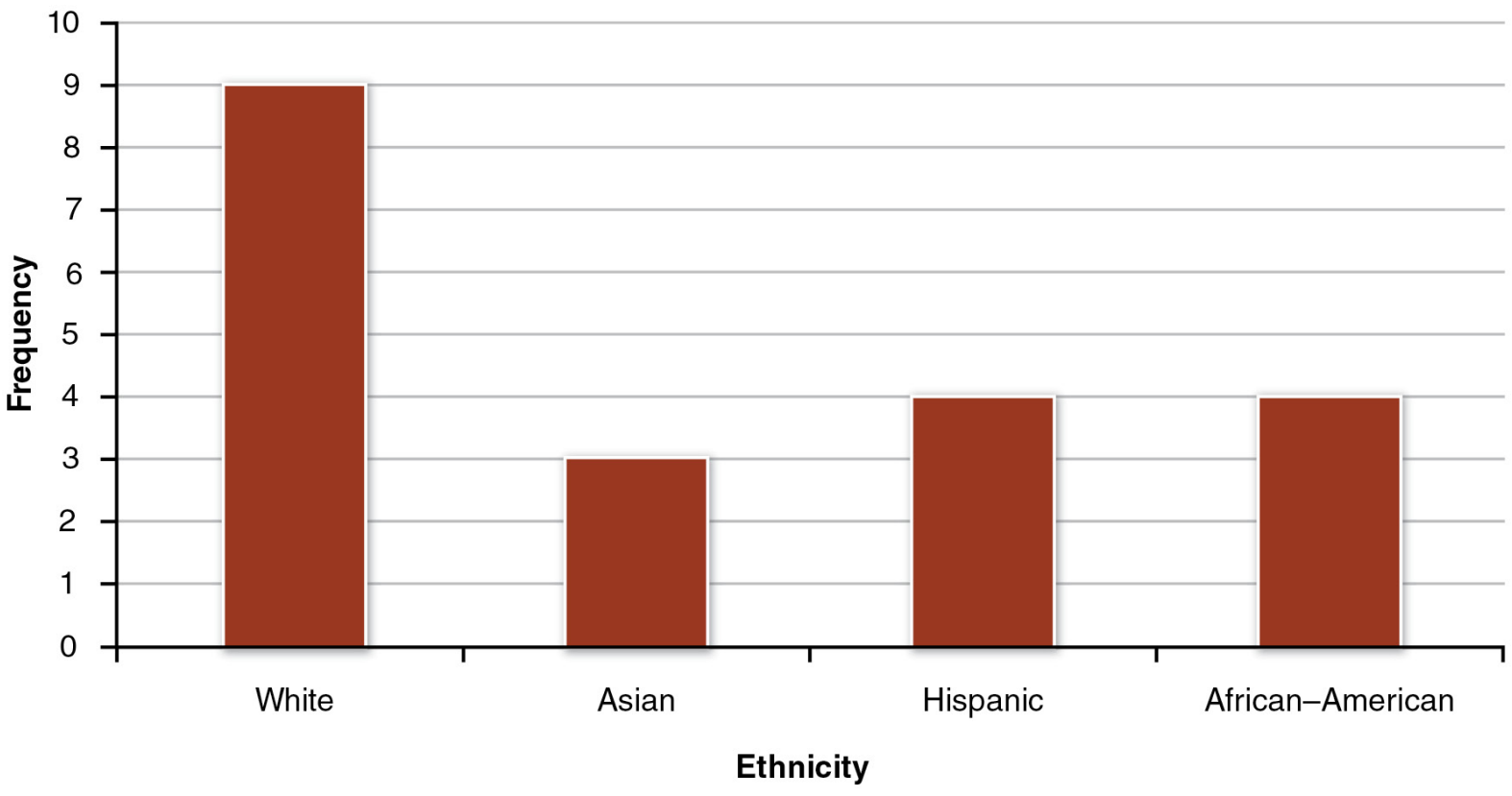
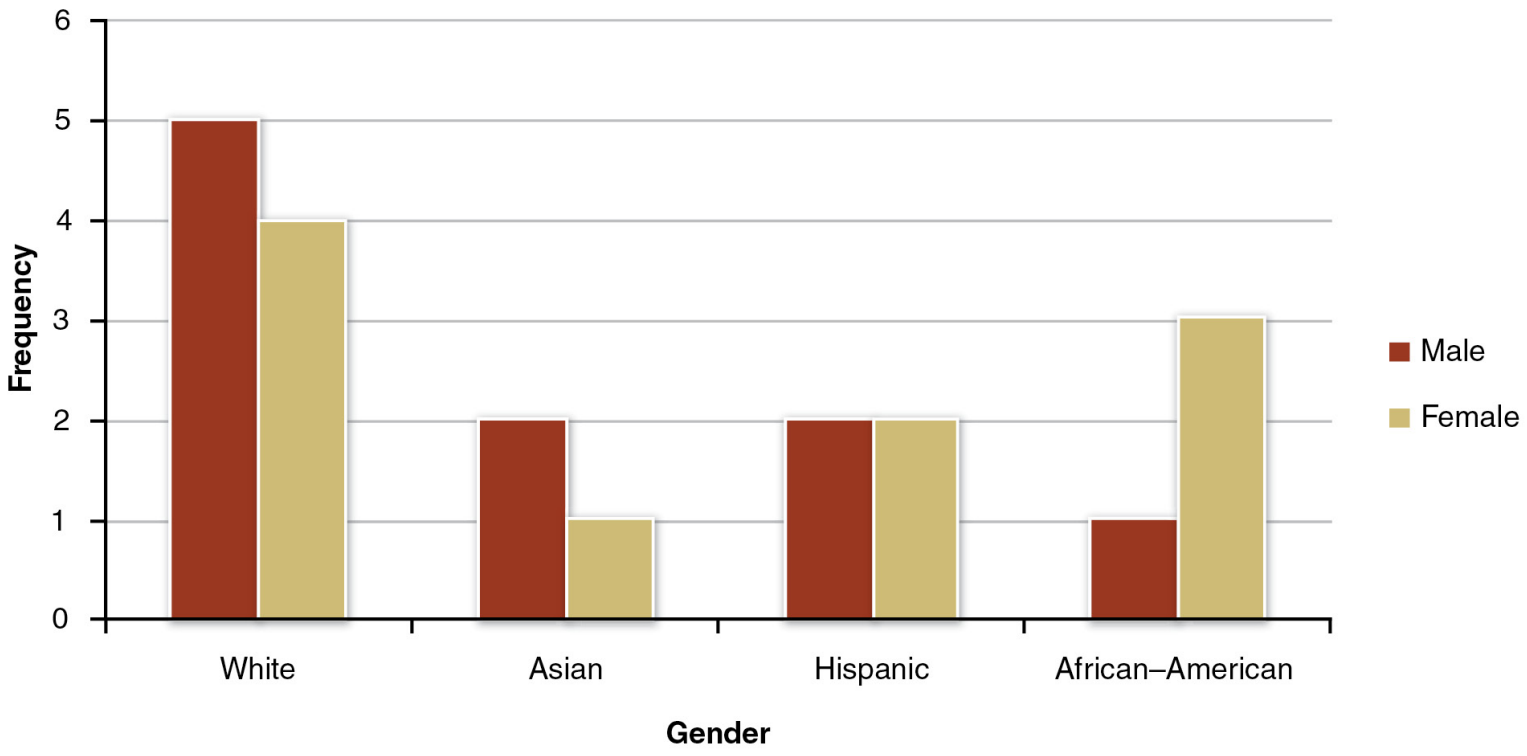


Figure 3.4b shows another variation on the bar graph, the **clustered bar graph**, which summarizes the frequency by two categories at one time. In this case, the bar graph displays information about both gender and ethnicity. As in the previous

graph, categories of ethnicity are displayed along the horizontal axis. But this time, we have divided the total number of each ethnicity by the gender of respondents—indicated using different colored bars. For example, notice that the nine White participants are divided into five males and four females. Similarly, the four African-American participants are divided into one male and three females.

Figure 3.4b: Clustered bar graph displaying frequency by ethnicity and gender



Keep in mind that bar graphs are used for qualitative, or nominal, categories. We could just as easily have listed Caucasian participants second, third, or fourth along the axis because ethnicity is measured on a nominal scale.

When we want to present quantitative data—that is, those values measured on an ordinal, interval, or ratio scale—we use a different kind of graph called a **histogram**. As Figure 3.5a shows, histograms are drawn with the bars touching one another to indicate that the categories are quantitative and on a continuous scale. This figure has broken down the "life-satisfaction" values into three categories (less than 30, 31–40, and 41–50) and displayed the frequencies for each category in numerical order. For example, six people had life satisfaction scores falling between 31 and 40.

Finally, all of our bar graphs and histograms so far have displayed data that have been split into categories. However, as Figure 3.5b illustrates, histograms can also present data on a continuous scale. Figure 3.5b also has an additional new feature—a curved line overlaid on the graph. This curve represents a normal distribution and allows us to gauge visually how close our sample data are to being normally distributed.

Figure 3.5a: Histogram showing frequencies by life satisfaction (quantitative) categories

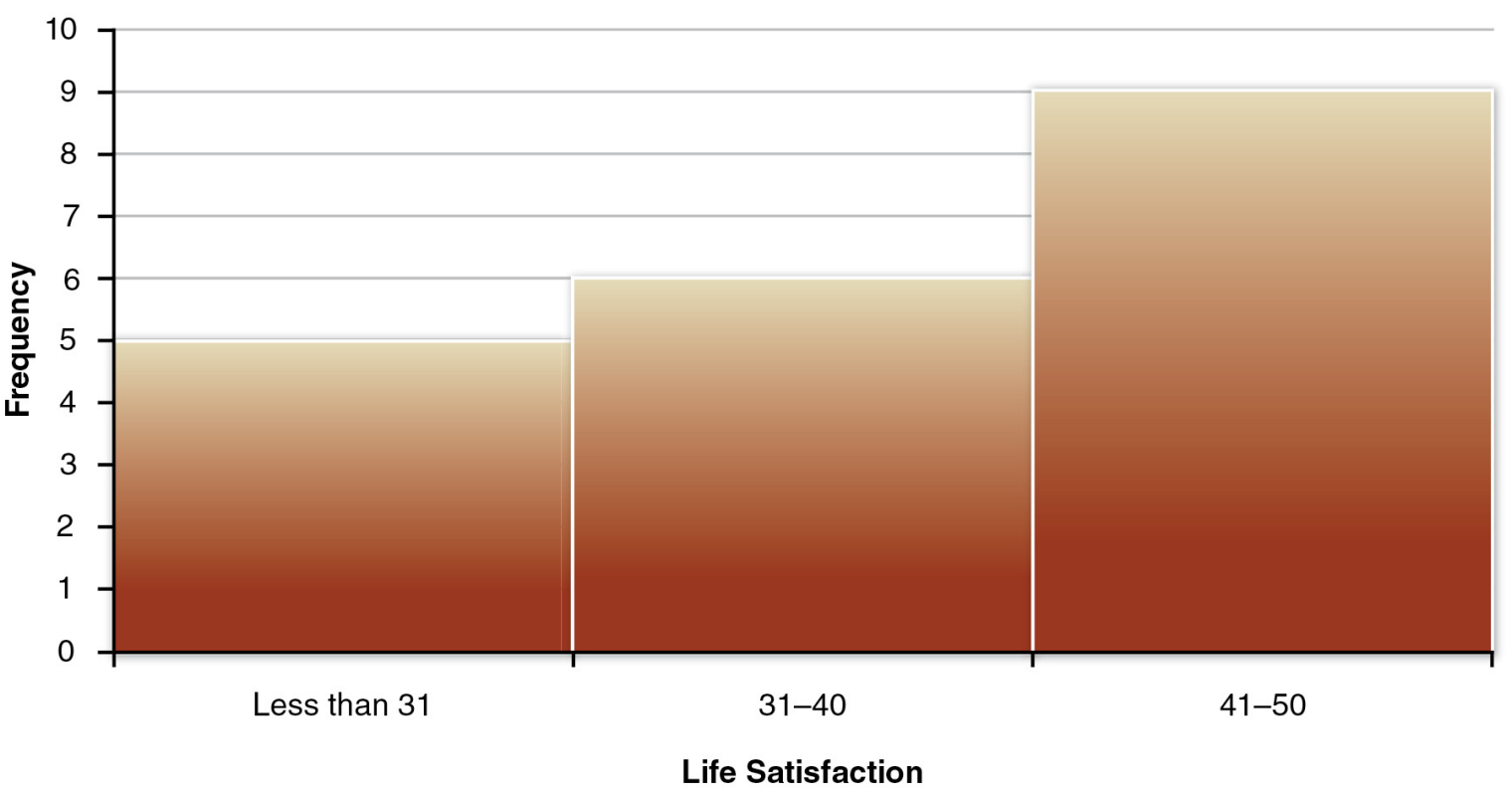
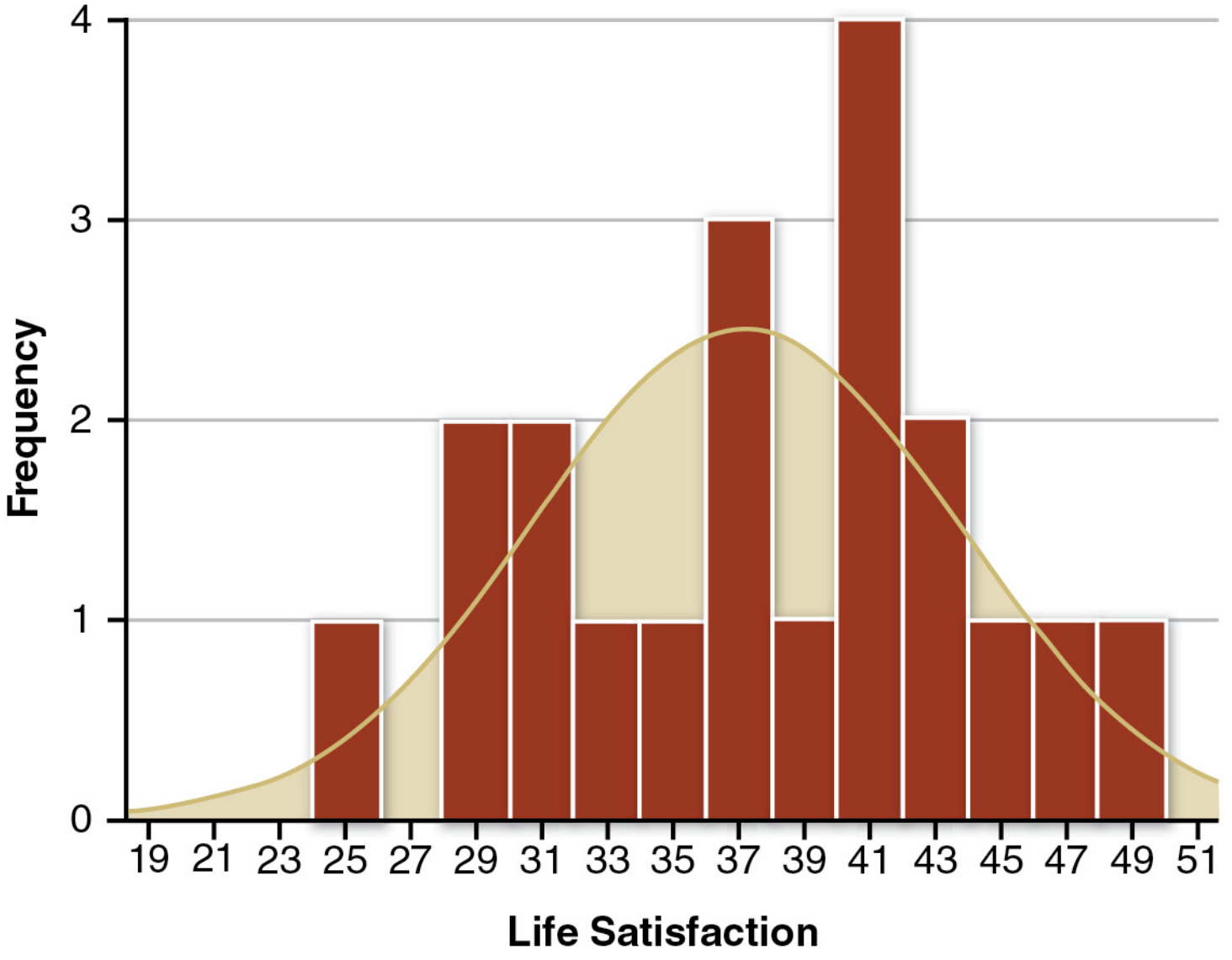


Figure 3.5b: Histogram showing life satisfaction scores on a continuous scale

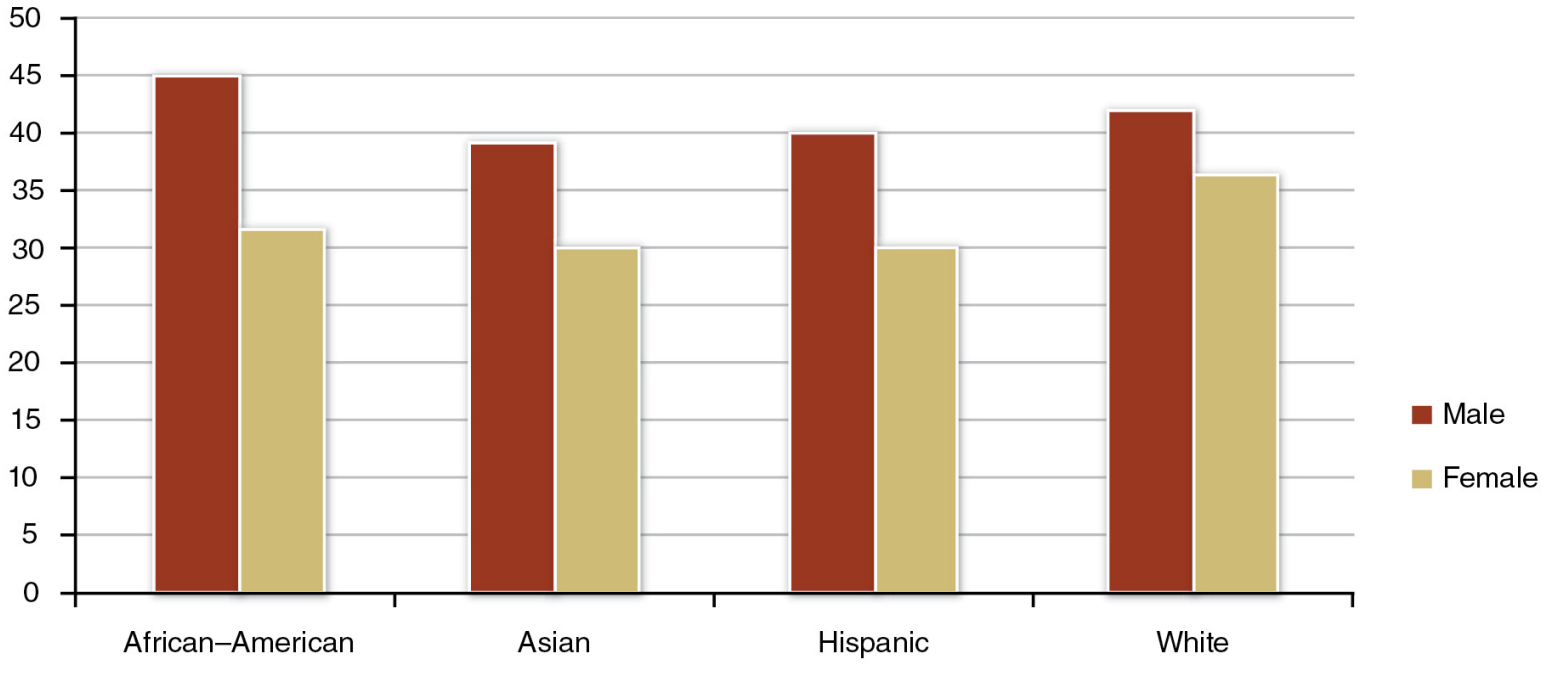


Displaying Central Tendency

Graphs are also commonly used to display numeric descriptors in an easy-to-understand visual format. Referring back to the sample data in Table 3.4 provides information about ethnicity and gender but also about reports of daily stress and life satisfaction. Thus, a natural question is whether there are gender or ethnic differences in these two variables. Figure 3.6 displays a clustered bar graph showing the mean level of life satisfaction in each group of participants. Of note is that males appear to report more life satisfaction than females, as revealed by the fact that the red bars are always higher than the gold bars. We can also see some variation in satisfaction levels by ethnicity: African-American males (45) appear to

report slightly more satisfaction than White males (42).

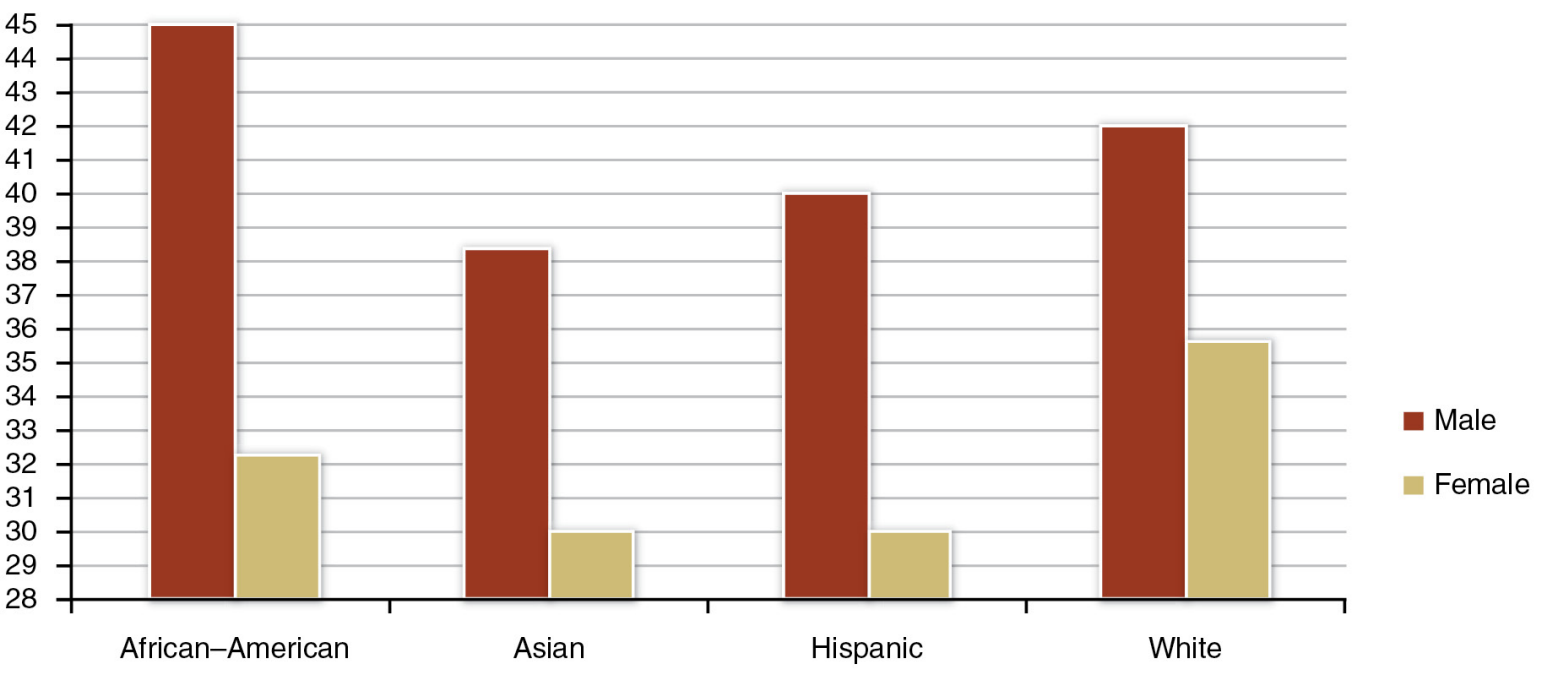
Figure 3.6: Clustered bar graph displaying life satisfaction scores by gender and ethnicity



These particular data are fictional, of course, but even if our graph depicted real data, we would want to be cautious in interpreting them. One reason for caution is that the data represent a descriptive study. We might be able to state which demographic groups report more life satisfaction, but we would be unable to determine the reasons for the difference. Another, more important, reason for caution is that visual presentations can be misleading, and we would need to conduct statistical analyses to discover the real patterns of differences.

The best way to appreciate this latter point is to notice what happens when we tweak the graph a little bit. The original graph in Figure 3.6 is a fair representation of the data: The scale starts at zero, and the *y*-axis on the left side increases by reasonable intervals. However, if we were trying to win an argument about gender differences in happiness, we could always alter the scale, as Figure 3.7 shows. These bars represent the same set of means, but we have compressed the *y*-axis to show only a small part of the range of the scale. That is, rather than ranging from 0 to 50, this misleading graph ranges from 28 to 45, in increments of 1. To the uncritical eye, the graph appears to show an enormous gender difference in life satisfaction; to the trained eye, it shows an obvious attempt to make the findings seem more interesting. Anytime we encounter a bar graph used to support a particular argument, we must always pay close attention to the scale of the results: Does it represent the actual range of the data, or is it compressed to exaggerate the difference? Likewise, any time researchers create a graph to display results, they have a responsibility to ensure that the graph is an accurate representation of the data.

Figure 3.7: Clustered bar graph altered to exaggerate the differences


45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
28
African–American
Asian
Hispanic
White
Male
Female

Summary and Resources

Chapter Summary

This chapter has focused on descriptive designs, the first of three specific research designs the text will discuss. As the name implies, the primary goal of descriptive designs is to describe attitudes and behavior, without any pretense of making causal claims. One common feature of all descriptive designs is that they are able to assess behaviors in their natural environment, or at least in something very close to it. The chapter covered three types of descriptive research: case studies, archival research, and observational research. Because each of these methods has the goal of describing attitudes, feelings, and behaviors, each one can be used from either a quantitative or a qualitative perspective.

In a case study, the researcher studies one person in great detail over a period of time. This approach is often used to study special populations and to gather detailed information about rare phenomena. On the one hand, case studies represent the lowest point on the continuum of control because of the lack of a comparison group and the difficulty of generalizing from a single case. On the other hand, case studies are a valuable tool for beginning to study a phenomenon in depth. We discussed the example of Phineas Gage, who suffered severe brain damage and showed drastic changes in his personality and cognitive skills. Although it is difficult to generalize from the specifics of Gage's experience, this case has helped to inspire more than a century's worth of research into the connections among mind, brain, and behavior.

Archival research involves drawing new conclusions by analyzing existing sources of data. This approach is often used to track changes over time or to study things that would be impossible to measure in a laboratory setting. For example, we discussed Phillips's study of copycat suicides, which he conducted by matching newspaper coverage of suicides to subsequent spikes in fatality rates. There would be no practical or ethical way to study these connections other than examining the patterns as they occurred naturally. Archival studies are still relatively low on the continuum of control, primarily because the researcher does not have much control over how the data are collected. In many cases, analyzing archives involves a process known as content analysis, or developing a coding strategy to extract relevant information from a broader collection of content. Content analysis involves a three-step process: identifying the most relevant archives, sampling from these archives, and finally coding and recording behaviors. For example, Weigel and colleagues studied race relations on television by sampling a week's worth of prime-time programming and recording the screen time dedicated to portraying interactions between characters of different races.

Lastly, observational research involves directly observing behavior and recording observations in a systematic way. This approach is well suited to a wide variety of research questions, provided the variables can be directly observed. That is, researchers can observe *what* people do but not *why* they do it. In exchange for giving up access to internal processes, researchers gain access to unfiltered behavioral responses—especially when they find ways to observe people unobtrusively. We discussed three main types of observational research. Structured observation involves creating a standardized situation, often in a laboratory setting, and tracking people's responses. Naturalistic observation involves observing behavior as it occurs naturally, often in its natural context. Participant observation involves having the researcher take part in the same activities as the participants in order to gain greater insight into their private behaviors. All three of these variations go through a similar three-step process as archival research: choose a hypothesis, choose a sampling strategy, and then code and record behaviors.

Finally, this chapter discussed principles for describing data in both visual and numeric form. To move toward conducting statistical analyses, summarizing data in numeric form is also useful. We discussed two categories of numeric summaries, central tendency and dispersion. Measures of central tendency (i.e., mean, median, and mode) provide information about the "typical" score in a dataset, while measures of dispersion (i.e., range, variance, and standard deviation) provide information about the distribution of scores around the central tendency—that is, they tell us *how typical* the typical score is. Finally, the chapter described the process of translating scores into standard scores (aka, z scores), which express individual scores in terms of standard deviations. This technique is useful for comparing results from different studies and using different measures. The chapter also discussed guidelines for visual presentation. Remember that the sole purpose of visual information is to communicate research findings to an audience. Thus, a researcher's descriptions should always be

accurate, concise, and easy to understand. The most common visual displays for summarizing data are bar graphs (for nominal data) and histograms (for quantitative data). Regardless of the choice of visual display, it should represent the data accurately; it is especially important to make sure that the *y*-axis accurately represents the range of data.

Key Terms

archival research

bar graph

case study

central tendency

clustered bar graph

content analysis

deviation score

dispersion

ecological validity

ethnography

event sampling

frequency tables

histogram

individual sampling

mean

median

mode

naturalistic observation

normal distribution (or "bell curve")

observational research

participant observation

participant reactivity

range

sample

standard scores (or *z* scores)

structured observation

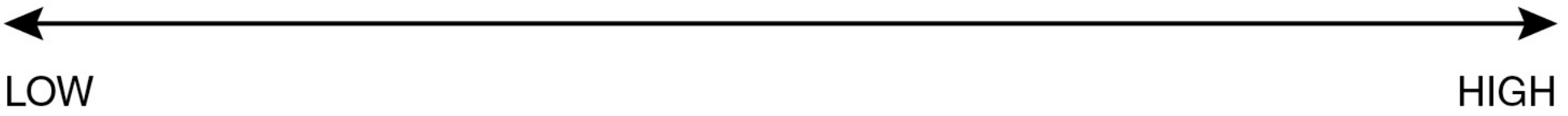
time sampling

variance

Chapter 3 Flashcards

Apply Your Knowledge

1. Compare and contrast the sets of the following terms. Your answers should demonstrate that you understand each term.
    a. individual sampling versus event sampling
    b. participant observation versus naturalistic observation
    c. mean versus median versus mode
    d. variance versus standard deviation
    e. bar graph versus histogram
2. Place each of the three research methods we have discussed in this chapter (listed below) on the continuum of control.



archival research

case study

naturalistic observation

3. For each of the following research methods, list one advantage and one disadvantage.

a. archival research

advantage:

disadvantage:

b. case studies

advantage:

disadvantage:

c. observation studies

advantage:

disadvantage:

4. For each of the following data sets, compute the mean, median, mode, and standard deviation. Once you have determined all three measures of central tendency, decide which one best represents the data.
   a. 2, 2, 4, 5
   b. 10, 13, 15, 100
5. Mike scores an 80 on a math test that has a mean of 100 and a standard deviation of 20. Convert Mike's test score into a *z* score.
6. For each of the following relationships, state the best way to present it graphically (bar graph, clustered bar graph, or histogram).
   a. average income by years of school completed (ratio scale)
   b. average income based on category of school completed (high school, some college, college degree, master's degree, and doctoral degree)
   c. average income based on gender and category of school completed
7. For each of the following questions, state how you would test them using an observational design.
   a. Are people who own red cars more likely to drive recklessly?
      (1) What would your hypothesis be?
      (2) Where would you acquire your sample and how (i.e., which type)?
      (3) What categories of behavior would you record? How would you define them?
   b. Are men more likely than women to "lose control" at a party?
      (1) What would your hypothesis be?
      (2) Where would you acquire your sample and how (i.e., which type)?
      (3) What categories of behavior would you record? How would you define them?
   c. How many fights break out in an average NHL (hockey) game?
      (1) What would your hypothesis be?
      (2) Where would you acquire your sample and how (i.e., which type)?
      (3) What categories of behavior would you record? How would you define them?

Critical Thinking Questions

1. Explain the tradeoffs involved in taking a qualitative versus a quantitative approach to a research question. What are the pros and cons of each one?
2. What are the advantages and disadvantages of conducting participant observation?

Research Scenarios: Try It