

# 4 Survey Designs— Predicting Behavior



*Duncan Smith/Photodisc/Thinkstock*

## Learning Objectives

---

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

- Describe the distinguishing features of survey research.
- Outline best practices for designing questionnaires to ensure quality responses.
- Explain the reasons for sampling for the population.
- Distinguish the different types of sampling strategies.
- Explain the logic behind common approaches to analyzing survey data.

In a highly influential book published in the 1960s, the sociologist Erving Goffman (1963) defined *stigma* as an unusual characteristic that triggers a negative evaluation. In his words, “the stigmatized person is one who is reduced in our minds from a whole and usual person to a tainted, discounted one” (p. 3). People’s beliefs about stigmatized characteristics exist largely in the eye of the beholder, but have substantial influence on social interactions with the stigmatized (see Snyder, Tanke, & Berscheid, 1977). A large research tradition in psychology has been devoted to understanding both the origins of stigma and the consequences of being stigmatized. According to Goffman and others, the characteristics associated with the greatest degree of stigma have three features in common: they are highly visible, they are perceived as controllable, and they are misunderstood by the public.

Recently, researchers have taken considerable interest in people’s attitudes toward members of the gay and lesbian community. Although these attitudes have become more positive over time, this group still encounters harassment and other forms of discrimination on a regular basis (see Almeida, Johnson, Corliss, Molnar, & Azrael, 2009; Berrill, 1990). One of the top recognized experts on this subject is Gregory Herek, professor of psychology at the University of California at Davis (<http://psychology.ucdavis.edu/herek/>). In a 1988 article, Herek conducted a survey of heterosexuals’ attitudes toward both lesbians and gay men, with the goal of understanding the predictors of negative attitudes. Herek approached this research question by constructing a questionnaire to measure people’s attitudes toward these groups. In three studies, participants were asked to complete this attitude measure, along with other existing scales assessing attitudes about gender roles, religion, and traditional ideologies.

Herek’s (1988) research revealed that, as hypothesized, heterosexual males tended to hold more negative attitudes about gay men and lesbians than did heterosexual females. However, the same psychological mechanisms seemed to explain the prejudice in both genders. That is, negative attitudes toward gays and lesbians were associated with increased religiosity, more traditional beliefs about family and gender, and fewer experiences actually interacting with gay men and lesbians. These associations meant that Herek could predict people’s attitudes toward gay men and lesbians based on knowing their views about family, gender, and religion, as well as their past interactions with the stigmatized group. In this paper, Herek’s primary contribution to the literature was the insight that reducing stigma toward gay men and lesbians “may require confronting deeply held, socially reinforced values” (1988, p. 473). This insight was only possible because people were asked to report these values directly.

This chapter continues along the continuum of control, moving on to survey research, in which the primary goal is either describing or predicting attitudes and behavior. For our purposes, **survey research** refers to any method that relies on people’s direct reports of their own attitudes, feelings, and behaviors. So, for example, in Herek’s (1988) study, the participants reported their attitudes toward lesbians and gay men, rather than these attitudes being somehow directly observed by the researchers. Compared to the descriptive designs we discussed in Chapter 3, survey designs tend to have more control over both data collection and question content. Thus, survey research falls somewhere between purely descriptive research (Chapter 3) and the explanatory power of experimental designs (Chapter 5). This chapter provides an overview of survey research from conceptualization through analysis. It will discuss the types of research questions that are best suited to survey research and provide an overview of the decisions to consider in designing and conducting a survey study. We will then cover the process of data collection, with a focus on selecting the people who will complete surveys. Finally, the chapter will describe the three most common approaches for analyzing survey data.

## Research: Making an Impact

### Kinsey Reports

Alfred Kinsey's research on human sexuality is an example of social research that changed the way society thought about a complex issue—in this case, ideas about “normal” sexual behavior. Kinsey's research, particularly two books on male and female sexuality known together as the Kinsey Reports, illuminated the discrepancies between the assumptions made by a “moral public” and the actual behavior of individuals. His shift in the approach to studying sex—applying scientific methods and reasoning rather than basing conclusions on medical speculation and dogmatic opinions—changed the nature of sex research and the general public's view of sex for decades to come.

Kinsey's major contribution was in challenging the prevailing assumptions about sexual activity in the United States and obtaining descriptive data from both men and women that described their own sexual practices (Bullough, 1998). By collecting actual data instead of relying on speculation, Kinsey made the study of sexuality more scientifically based. The results of his surveys revealed a variety of sexual behaviors that shocked many members of society and redefined the sexual morality of modern America.

Until Kinsey's research, the general, Victorian viewpoint was that women should not show any interest in sex and should submit to their husband without any sign of pleasure (Davis, 1929). Kinsey's data challenged the prevailing assumption that women were asexual. His studies revealed that 25% of the women studied had experienced an orgasm by the age of 15 and more than half by the age of 20 (Kinsey, Pomeroy, Martin, & Gebhard, 1953). Eventually, these results were bundled into the various elements that fueled the women's movement of the 1960s and encouraged further examination of female sexuality (Bullough, 1998).

Kinsey's data also contributed to the budding gay and lesbian liberation movement. Until the Kinsey Reports, studies of human sexuality were based on the assumption that homosexuals were mentally ill (Bullough, 1998). When Kinsey's data revealed that many males and females practiced homosexuality to some degree, he suggested that sexuality was more of a continuum than a series of categories into which people fit. In addition, the Kinsey Reports revealed that the number of extramarital relationships people were having was higher than most expected. Forty percent of married American males reported having an extramarital relationship (Kinsey, et al., 1953).

These ideas, though controversial, prompted society to take a realistic look at the actual sexual practices of its members. The topic of sexuality became less dogmatic as society became more open about sexual activities and preferences.

Kinsey's data not only encouraged social change but also revolutionized the way in which scientists study sexuality. By examining data and studying sex from an unbiased standpoint, Kinsey successfully transformed the study of human sexuality into a science. His research not only changed our way of studying sexual behavior but also allowed society to become less restrictive in its expectations of “normal” sexual behavior.

### Think About It

1. What type of data formed the basis of Kinsey's reports? What are the pros and cons of this type?
2. How did applying the scientific method change the national conversation about sexuality?

## 4.1 Introduction to Survey Research

Whether you aware of it or not, most people encounter survey research throughout most of their lives. Every time we decide to answer that call from an unknown number, and the person on the other end of the line insists on knowing the call recipient's household income and favorite brand of laundry detergent, we are helping to conduct survey research. When news programs try to predict the winner of an election two weeks early, these reports are based on survey research of eligible voters. In both cases, the researcher is trying to make predictions about the products people buy or the candidates they will elect based on what people say about their own attitudes, feelings, and behaviors.



shironosov/iStock/Thinkstock

**Surveys are used to describe or predict attitudes and behavior.**

Surveys can be used in a variety of contexts and are most appropriate for questions that involve people describing their attitudes, their behaviors, or a combination of the two. For example, if we want to examine the predictors of attitudes toward the death penalty, we could ask people their opinions on this topic and also ask them about their political party affiliation. Based on these responses, we could test whether political affiliation predicted attitudes toward the death penalty. Or, imagine we want to know whether students who spend more time studying are more likely to do well on their exams.

This question could be answered using

a survey that asked students about their study habits and then tracked their exam grades. We will return to this example near the end of the chapter, as we discuss the process of analyzing survey data to test our hypotheses about predictions.

The common thread of these two examples is that they require people to report either their thoughts (e.g., opinions about the death penalty) or their behaviors (e.g., the hours they spend studying). Contrast these with an example that might be a poor fit for survey research: If a researcher wanted to test whether a new drug led to increased risk of developing blood clots, it would be much safer to test for these clots using medical technology, rather than asking people for their beliefs ("on a scale from 1 to 5, how many clots have you developed this week?"). Thus, when deciding whether a survey is the best fit for a research question, a researcher must consider whether people will be both able and willing to report the opinions or behaviors accurately. The next section expands on both of these issues.

### Distinguishing Features of Surveys

Survey research designs have three distinguishing features that set them apart from other designs. First, all survey research relies on either written or verbal **self-reports** of people's attitudes, feelings, and behaviors. This self-reporting means that researchers will ask

participants a series of questions and record their responses. The approach has several advantages, including being relatively straightforward and allowing a degree of access to psychological processes (e.g., “*Why do you support candidate X?*”). However, researchers should also be also cautious in their interpretation of self-report data because participants’ responses often reflect a combination of their true attitude and concern over how this attitude will be perceived. Scientists refer to this concern as **social desirability**, which means that people may be reluctant to report unpopular attitudes. For example, if we were to ask people their attitudes about different racial groups, their answers might reflect both their true attitude and their desire not to appear racist. We return to the issue of social desirability later in this chapter and discuss some tactics for designing questions that can help to sidestep these concerns and capture respondents’ true attitudes.

The second distinguishing feature of survey research is its ability to access internal states that cannot be measured through direct observation. The discussion of observational designs in Chapter 3 explained that one limitation of these designs was a lack of insight into *why* people behave the way they do. Survey research can address this limitation directly: By asking people what they think, how they feel, and why they behave in certain ways, researchers come closer to capturing the underlying psychological processes. However, people’s reports of their internal states should be taken with a grain of salt, for two reasons. First, as mentioned, these reports may be biased by social-desirability concerns, particularly when unpopular attitudes are involved. Second, a large body of literature in social psychology suggests that people may not understand the true reasons for their behavior. In an influential review paper, psychologists Richard Nisbett and Tim Wilson (1977) argued that we make poor guesses after the fact about why we do things, based more on our assumptions than on any real introspection. Thus, survey questions can provide access to internal states, but researchers should always interpret responses with caution.

Third, on a more practical note, survey research allows us to collect large amounts of data with relatively little effort and few resources. Many of the descriptive designs Chapter 3 discussed require observing one person at a time, and the same will hold true when Chapter 5 explores experimental designs. Survey-research designs stand out as the most efficient, because surveys can be distributed to large groups of people simultaneously. Still, their actual efficiency depends on the decisions researchers make during the design process. In reality, efficiency is often in a delicate balance with the accuracy and completeness of the data.

Broadly speaking, survey research can be conducted using either verbal or written self-reports (or a combination of the two). Before diving into the details of writing and formatting a survey, we need to understand the pros and cons of administering a survey as an interview (i.e., a verbal survey) or a questionnaire (i.e., a written survey).

## Interviews

An **interview** involves a verbal question-and-answer exchange between the researcher and the participant. This verbal exchange can take place either face-to-face or over the phone. So, our earlier telemarketer example represents an interview because the questions are asked verbally via phone. Likewise, if we are approached in a shopping mall and asked to answer questions about our favorite products, we experience a survey in interview form because

the questions are administered verbally face-to-face. And, if a person has ever participated in a focus group, during which a group of people gives their reactions to a new product, the researchers are essentially conducting an interview with the group.

### ***Interview Schedules***

Regardless of how the interview is administered, the interviewer (i.e., the researcher) has a predetermined plan, or script, for how the interview should go. This plan, or script, for the progress of the interview is known as an **interview schedule**. When conducting an interview—including those telemarketing calls—the researcher/interviewer has a detailed plan for the order of questions to be asked, along with follow-up questions that depend on the participant's responses.

Broadly speaking, researchers employ two types of interview schedules. A **linear** (also called “structured”) **schedule** will ask the same questions, in the same order, for all participants. In contrast, a **branching schedule** unfolds more like a flowchart, with the next question dependent on participants’ answers. Interviewers typically use a branching schedule in cases with follow-up questions that only make sense for some of the participants. For example, a researcher might first ask people whether they have children; if they answer “yes,” the interviewer might then follow up by asking how many.

One danger in using a branching schedule is that it is based partly on the researcher’s assumptions about the relationships between variables. Granted, to ask only people with children to indicate how many they have is fairly uncontroversial. Imagine the following scenario, however. Say we first ask participants for their household income, and then ask about their political donations:

- “How much money do you make? \$18,000? OK, how likely are you to donate to the Democratic Party?”
- “How much money do you make? \$250,000? OK, how likely are you to donate money to the Republican Party?”

The way these questions branch implicitly assumes that wealthier people are more likely to be Republicans, and less wealthy people are more likely to be Democrats. The data might support this assumption or they might not. By planning the follow-up questions in this way, though, we are unable to capture cases that do not fit our stereotypes (i.e., the wealthy Democrats and the poor Republicans). Researchers must therefore be careful about letting their biases shape the data-collection process.

### ***Advantages and Disadvantages of Interviews***

Interviews offer a number of advantages over written surveys. For one, people are often more motivated to talk than they are to write. Consider the example of an actual undergraduate research assistant who was dispatched to a local shopping mall to interview people about their experiences in romantic relationships. He had no trouble at all recruiting participants, many of whom would go on and on (and on, and on) about recent relationships—one woman

even confided to him that she had just left an abusive spouse earlier that week. For better or for worse, these experiences would have been more difficult to capture in writing.

Related to this benefit, people's oral responses are typically richer and more detailed than their written responses. Think of the difference between asking someone to "describe your views on gun control" and asking someone to "indicate on a scale of 1 to 7 the degree to which you support gun control." The former is more likely to capture the richness and subtlety involved in people's attitudes about guns. On a practical note, an interview format also allows the researcher to ensure that respondents understand the questions. Poorly worded written-questionnaire items force survey participants to guess at the researcher's meaning, and these guesses introduce a large source of error variance. On the other hand, if an interview question is poorly asked, people can easily ask the interviewer to clarify. Finally, using an interview format allows researchers to reach a broader cross-section of people and to include those who are unable to read and write—or, perhaps, unable to read and write the language of the survey.

Interviews also have two clear disadvantages compared to written surveys. First, interviews cost more in terms of both time and money. It took more time for the graduate assistant to go to a shopping mall than it would have taken to mail out packets of surveys (but no more money—research-assistant positions tend to be unpaid). Second, the interview format allows many opportunities for interviewers to pass on their personal biases. These biases are unlikely to be deliberate, but participants can often pick up on body language and subtle facial expressions when the interviewer disagrees with their answers. Such cues may influence them to shape their responses to make the interviewer happier. The best way to understand the pros and cons of interviewing is to recognize that both are a consequence of personal interaction. The interaction between interviewer and interviewee allows for richer responses but also the potential for these responses to be biased. Researchers must weigh these pros and cons and decide which method is the best fit for their survey. The next section turns to the process of administering surveys in writing.

One additional problem with interviews is the increasing difficulty of obtaining representative samples for interviews over the telephone due to low or declining use of landline phones, coupled with the use of unlisted numbers and call-screening devices. In the United States, the Pew Research Center (2012) reports that overall response rate—a ratio of completed interviews to the number of phone numbers dialed—was just 9% in 2012, one-fourth of the 36% level from 1997. Thus, significant differences may exist between people who elect to respond to phone surveys and those who do not.

## Questionnaires

A **questionnaire** is a survey that involves a written question-and-answer exchange between the researcher and the participant. The exchange is a bit different from interview formats—in this case, the questions are designed ahead of time, then distributed to participants, who write their responses and return the questionnaire to the researcher. The next section discusses details for designing these questions. First, however, we will take a quick look at the process of administering written surveys.

## Distribution Methods

Questionnaires can be distributed in three primary ways, each with its own pattern of advantages and disadvantages:

**Distributing by mail:** Until recently, researchers commonly distributed surveys by sending paper copies through the mail to a group of participants (see the section on “Sampling” for more discussion on how this group is selected). Mailing surveys is relatively cheap and relatively easy to do, but it is unfortunately one of the worst methods in terms of response rates. People tend to ignore questionnaires that they receive in the mail, dismissing them as one more piece of junk. Researchers have a few methods available for increasing response rates, including providing incentives, making the survey interesting, and making it as easy as possible to return the results (e.g., with a postage-paid envelope). However, even using all of these tactics, researchers consider themselves extremely lucky to obtain a 30% response rate from a mail survey. That means a researcher who mails 1,000 surveys will be doing well to receive 300 back. More typical response rates for mail surveys can be in the single digits. Because of this low return on investment, researchers have begun relying on other methods for their written surveys.

**Distributing in person:** Another option for researchers is to distribute a written survey in person, simply handing out copies and asking participants to fill them out on the spot. This method is certainly more time-consuming; a researcher has to be stationed for long periods of time to collect data. In addition, people are less likely to answer the questions honestly because the presence of a researcher makes them worry about social desirability. Last, the sample for this method is limited to people who are in the physical area at the time that questionnaires are being distributed. As the chapter discusses later, this limitation might lead to problems in the composition of the sample. On the plus side, however, this method tends to result in higher compliance rates because people find it harder to say no to someone face-to-face than to ignore a piece of mail.

**Distributing online:** During the last two decades, online surveys have become the dominant method of data collection, for both market research and academic research. Online distribution involves posting a questionnaire on a web page, and then directing participants to this web page to complete the questionnaire. Online surveys offer many benefits over other forms of data collection, including the ability to: present audio and visual stimuli, randomize the order of questions, and implement complex branching logic (e.g., asking people to evaluate local grocery stores depending on where they live).

Most recently, researchers have begun exploring the best ways to design surveys for mobile devices. According to a report from the International Telecommunications Union, in 2013, 6.8 billion mobile phones were in use, compared to a world population of 7.2 billion. In 2012, 44% of Americans slept next to



*AndreyPopov/iStock/Thinkstock*

**Approximately 20–30% of online surveys are completed on a mobile device.**

their phones (Pew Research Center, 2012). Not surprisingly, consensus in the market research industry is that approximately 20–30% of *online* surveys are actually completed on a *mobile* device (Poynter, Williams, & York, 2014). Why does this matter? People take surveys on their smartphones because it is convenient (or, in some cases, because it is their only Internet device). However, despite recent exponential advancement, mobile phones still have smaller screens, less functional keyboards, and less predictability in displaying images and videos. (Imagine someone being asked to view a set of two-minute-long advertisements on an iPhone while trying to complete a survey before a doctor’s appointment.) Researchers do have ways to make this experience more pleasant for respondents and consequently to increase the quality of data obtained. For example, mobile surveys work best when they are shorter overall, when the question text is short and straightforward, and when response scales (discussed below) are kept at five points (see Poynter et al., 2014, for a review). The latter point is a direct result of small screen size: Longer response scales require respondents to scroll back and forth on their screens to see the entire scale. Unfortunately, but understandably, some applied research suggests that people tend to ignore the scale points that they cannot see—perhaps using only four points out of a ten-point scale.

Because these methods are relatively new, the jury is still out on whether online and mobile distribution results in biased samples or biased responses. However, worth keeping in mind is that approximately 13% of the U.S. population does not have Internet access (Internet Users by Country, 2014). This group is disproportionately older (65+) and represents the lowest income and least educated segments of the population. Thus, if research questions involve reaching these groups, it is necessary to supplement online surveys with other distribution methods. For readers interested in more information on designing and conducting Internet research, Sam Gosling and John Johnson’s (2010) recent book provides an excellent resource. In addition, several groups of psychological researchers have been attempting to understand the psychology of Internet users (read about recent studies on this website: <http://www.spring.org.uk/2010/10/internet-psychology.php>).

### ***Advantages and Disadvantages of Questionnaires***

Just as interview methods do, written questionnaires claim their own set of advantages and disadvantages. Written surveys allow researchers to collect large amounts of data with little cost or effort, and they can offer a greater degree of anonymity than interviews. Anonymity can be a particular advantage in dealing with sensitive or potentially embarrassing topics. That is, people may be more willing to answer a questionnaire about their alcohol use or their sexual history than they would be to discuss these things face-to-face with an interviewer. On the downside, written surveys miss out on one advantage of interviews because no one is available to clarify confusing questions. Fortunately, researchers have one relatively easy way of minimizing this problem: make survey questions as clear as possible. The next section explains the process of questionnaire design.

## **4.2 Questionnaire Design**

One of the most important steps in conducting survey research is deciding how to construct and assemble the questionnaire items. In some cases, a researcher will be able to answer research questions using questionnaires that other researchers have already developed. For

example, quite a bit of psychology research uses standard scales that measure self-esteem, prejudice, depression, or stress levels. The advantage of these ready-made measures is that other people have already gone to the trouble of making sure they are valid and reliable. So, someone interested in the relationship between stress and depression could distribute the Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983) and the Beck Depression Inventory (Beck, Steer, Ball, & Ranieri, 1996) to a group of participants and more quickly move along on to the fun part of data analyses.

However, in many cases, no perfect measure exists for a research question—either because no one has studied the topic before or because the current measures are all flawed in some way. When this happens, researchers need to go through the process of designing their own questions. This section discusses strategies for writing questions and choosing the most appropriate response format.

## Five Rules for Better Questions

Each of the rules listed below is designed to make research questions as clear and easy to understand as possible so as to minimize the potential for error variance. We discuss each rule below and illustrate it with contrasting pairs of items: “bad” items that do not follow the rule and “better” items that do.

1. **Use simple language.** One of the simplest and most important rules to keep in mind is that people have to be able to understand the survey questions. This means avoiding jargon and specialized language whenever possible.

BAD: “Have you ever had an STD?”

BETTER: “Have you ever had a sexually transmitted disease?”

BAD: “What is your opinion of the S-CHIP program?”

BETTER: “What is your opinion of the State Children’s Health Insurance Program?”

It is also a good idea to simplify the language as much as possible, so that people spend time answering the question rather than trying to decode its meaning. For example, words like *assist* and *consider* can be replaced with simpler words like *help* and *think*. This may seem odd—or perhaps even condescending to participants—but it is always better to err on the side of simplicity. Remember, when people are forced to guess at the meaning of questions, these guesses add error variance to their answers.

2. **Be precise.** Another way to ensure that people understand the question is to be as precise as possible with wording. Ambiguously (or vaguely) worded questions will introduce an extra source of error variance into the data because people may interpret these questions in varying ways.

BAD: "What drugs do you take?" (Legal drugs? Illegal drugs? Now? In college?)

BETTER: "What prescription drugs are you currently taking?"

BAD: "Do you like sports?" (Playing? Watching? Which sports??)

BETTER: "How much do you enjoy watching basketball on television?"

3. **Use neutral language.** Questions should be designed to *measure* participants' attitudes, feelings, or behaviors rather than to *manipulate* these things. That is, avoid **leading questions** that are written in such a way that they suggest an answer.

BAD: "Do you beat your children?" (Who would say yes?)

BETTER: "Is it acceptable to use physical forms of discipline?"

BAD: "Do you agree that the president is an idiot?"

BETTER: "How would you rate the president's job performance?"

This guideline can be used to sidestep social desirability concerns. If the researcher suspects that people may be reluctant to report holding an attitude—for example, using corporal punishment with their children—it helps to phrase the question in a nonthreatening way: "using physical forms of discipline" versus "beating your children." Many current measures of prejudice adopt this technique. For example, McConahay's (1986) "modern racism" scale contains items such as "Discrimination against Blacks is no longer a problem in the United States." People who hold prejudicial attitudes are more likely to confess agreement with statements like this one than with blunter ones, like "I hate people from Group X."

4. **Ask one question at a time.** One remarkably common error that people make in designing questions is to include a **double-barreled question** (one which asks more than one question at a time). A new-patient questionnaire at a doctor's office often asks whether the patient suffer from "headaches and nausea." What if an individual only suffers from one of these or has a lot of nausea and an occasional headache? The better approach is to ask about each of these symptoms separately.

BAD: "Do you suffer from pain and numbness?"

BETTER: "How often do you suffer from pain?" "How often do you suffer from numbness?"

BAD: "Do you like watching football and boxing?"

BETTER: "How much do you enjoy watching football?" "How much do you enjoy watching boxing?"

5. **Avoid negations.** One final and simple way to clarify questions is to avoid questions with negative statements because these can often be difficult to understand. The first example below may be a little silly, but the second comes from a real survey of voter opinion.

BAD: "Do you never not cheat on your exams?" (Wait, what? Do I cheat? Do I not cheat? What is this asking?)

BETTER: "Have you ever cheated on an exam?"

BAD: "Are you against rejecting the ban on pesticides?" (Wait, so, am I for the ban? Against the ban? What is this asking?)

BETTER: "Do you support the current ban on pesticides?"

## Participant-Response Options

This section discusses the issue of deciding how participants should respond to survey questions. The decisions researchers make at this stage will affect the type of data they ultimately collect, so it is important to choose carefully. This section reviews the primary decisions a researcher will need to make about response options, as well as the pros and cons of each one.

One of the first choices to make is whether to collect open-ended or fixed-format responses. As the names imply, **fixed-format responses** require participants to choose from a list of options (e.g., "Choose your favorite color"), while **open-ended responses** ask participants to provide unstructured responses to a question or statement (e.g., "How do you feel about legalizing marijuana?"). Open-ended responses tend to be richer and more flexible but harder to translate into quantifiable data—analogous to the tradeoff we discussed in comparing written versus oral survey methods. To put it another way, some concepts are difficult to reduce to a seven-point fixed-format scale, but number ratings on these scales are easier to analyze than a paragraph of free-flowing text.

Another reason to think carefully about this decision is that fixed-format responses will, by definition, restrict people's options in answering the question. In some cases, these restrictions can even act as leading questions. In a study of people's perceptions of history, Dario Páez Rovira and his colleagues (Rovira, Deschamps, & Pennebaker, 2006) asked respondents to indicate the "most significant event over the last 50 years." When this was asked in an open-ended way (i.e., "list the most significant event"), 2% of participants listed the invention of computers. Another version of the survey asked the question using a fixed-format way (i.e., "choose



*Eduard Lysenko/iStock/Thinkstock*

**Thirty percent of participants selected the invention of computers as the most significant event of the past 50 years when presented with fixed-format responses, but when a different group was asked the same question in an open-ended format, only 20% listed the invention of computers.**

the most significant event"). When asked to select from a list of four options (World War II, invention of computers, Tiananmen Square, or man on the moon), 30% chose the invention of computers. In exchange for having easily coded data, the researchers accidentally forced participants into a smaller number of options. The result, in this case, was a distorted sense of the importance of computers in people's perceptions of history.

### **Fixed-Format Options**

Although fixed-format responses can sometimes constrain or skew participants' answers, researchers tend to use them more often than not. This decision is largely practical; fixed-format responses allow for more efficient data collection from a much larger sample. (Imagine the chore of having to hand-code 2,000 essays.) But once researchers have decided on this option for the questionnaire, the decision process is far from over. In this section, we discuss three possibilities for constructing a fixed-format response scale.

**True/false.** One fixed-format option asks questions using a true/false format, which asks participants to indicate whether they endorse a statement. For example:

"I attended church last Sunday."	True	False
"I am a U.S. citizen."	True	False
"I am in favor of abortion."	True	False

This last example may strike you as odd, and in fact it illustrates an important limitation in the use of true/false formats: They are best used for statements of facts rather than attitudes. It is relatively straightforward to answer whether we attended church or are a U.S. citizen. However, people's attitudes toward abortion are often complicated—one might be "pro-choice" but still support some restrictions, or "pro-life" but support exceptions (e.g., in cases of rape). For most people, a true/false question cannot even come close to capturing the complexity of these beliefs. However, for survey items that involve simple statements of fact, the true/false format can be a good option.

**Multiple choice.** A second option uses a multiple-choice format, which asks participants to select from a set of predetermined responses.

"Which of the following is your favorite fast-food restaurant?"

- a) McDonald's
- b) Burger King
- c) Wendy's
- d) Taco Bell

"Whom did you vote for in the 2012 presidential election?"

- a) Mitt Romney
- b) Barack Obama

"How do you travel to work most days? (Select all that apply.)"

- a) drive alone
- b) carpool
- c) public transportation

As these examples show, multiple-choice questions offer quite a bit of freedom in both the content and the response-scaling of questions. A researcher can ask participants either to select one answer or, as in the last example, to select all applicable answers. A survey can cover everything from preferences (e.g., favorite fast-food restaurant) to behaviors (e.g., how people travel to work).

Multiple-choice formats do have a downside. Whenever the survey provides a set of responses, it restricts participants' responses to that set. This is the problem that Rovira and colleagues (2006) encountered in asking people about the most significant events of the last century. In each of the examples above, the categories fail to capture all possible responses. What if someone's favorite restaurant is In-and-Out Burger? What if a respondent voted for Ralph Nader? What if a person telecommutes or bicycles to work? Researchers have two relatively easy ways to avoid (or at least minimize) this problem. First, when choosing the response options, plan carefully. During the design process, it helps to brainstorm with other people to ensure the survey is capturing the most likely range of responses. However, it is often impossible to provide every option that people might conceive. The second solution is to provide an "other" response to a multiple-choice question, which allows people to write in an option that the survey neglected to include. For example, our last question about traveling to work could be rewritten as:

"How do you travel to work on most days? (Select all that apply.)"

- a) drive alone
- b) carpool
- c) public transportation
- d) other (please specify): \_\_\_\_\_

This way, people who telecommute, or bicycle, or even ride their trained pony to work will have a way to respond rather than skipping the question. And, if researchers start to notice a pattern in these write-in responses (e.g., 20% of people added "bicycle"), then they have valuable knowledge to improve the next incarnation of the survey.

**Rating scales.** Last, but certainly not least, another option uses a rating-scale format, which asks participants to respond on a scale representing a continuum.

"Sometimes it is necessary to sacrifice liberty in the name of security."

1	2	3	4	5
not at all necessary				very necessary

"I would vote for a candidate who supported the death penalty."

1	2	3	4	5
not at all likely				very likely

"The political party in power right now has really messed things up."



This format is well suited to capturing attitudes and opinions, and, indeed, is one of the most common approaches to attitude research. Rating scales are easy to score, and they give participants some flexibility in indicating their agreement with or endorsement of the questions. Researchers have two critical decisions to make about the construction of rating-scale items; both have implications for how they analyze and interpret results.

First, a researcher needs to decide the **anchors**, or labels, for the response scale. Rating scales offer a good deal of flexibility in these anchors, as the examples above demonstrate. A survey can frame questions in terms of "agreement" with a statement or "likelihood" of a behavior, or researchers can customize the anchors to match their questions (e.g., "not at all necessary"). Scales that use anchors of "strongly agree" and "strongly disagree" are also referred to as **Likert scales**. At a fairly simple level, the choice of labels affects the interpretation of the results. For example, if we asked the "political party in power" question above, we have to be aware that the anchors are phrased in terms of agreement with the statement. In discussing these results, we would be able to discuss how much people agreed with the statement, on average, and whether agreement correlated with other things. If this seems like an obvious point, readers would be amazed how often researchers (or the media) will take an item like this and spin the results to talk about the "likelihood of voting" for the party in power—confusing an attitude with a behavior. So, in short, researchers must make sure they are being honest when presenting and interpreting research data.

At a more conceptual level, a researcher needs to decide whether the anchors for the rating scale make use of a **bipolar scale**, which has polar opposites at its endpoints, or a **unipolar scale**, which assesses a single construct. The difference between these options is best illustrated by an example:

**Bipolar:** How would you rate your current mood?

Sad—————Happy

**Unipolar:** How would you rate your current mood?

1	2	3	4	5	6	7
not at all sad						very sad

1	2	3	4	5	6	7
not at all happy						very happy

The bipolar option requires participants to place themselves on a continuous scale somewhere between "sad" and "happy," which are polar opposites. The bipolar scale assumes that the endpoints represent the only two options; participants can be sad, happy, or somewhere in between. In contrast, the unipolar option asks participants to rate themselves on two scales, indicating their level of both "sadness" and "happiness." A pair of unipolar scales

assumes that it is possible to experience varying degrees of each item—participants can be moderately happy, but also a little bit sad, for example. The decision to use a bipolar or a unipolar scale comes down to the context. What is the most logical way to think about these constructs? What have previous researchers done?

In the 1970s, Sandra Lipsitz Bem revolutionized the way researchers thought about gender roles by arguing against a bipolar approach. Previously, gender role identification had been measured on a bipolar scale from “masculine” to “feminine”; the scale assumed that a person could be one or the other. Bem (1974) argued instead that people could easily have varying degrees of masculine and feminine traits. Her scale, the Bem Sex Role Inventory, asks respondents to rate themselves on a set of 60 unipolar traits. Someone with mostly feminine and hardly any masculine traits would be described as “feminine.” Someone with high ratings on both masculine and feminine traits would be described as “androgynous.” And, someone with low ratings on both masculine and feminine traits would be described as “undifferentiated.” View and complete Bem’s scale online at: <http://garote.bdmonkeys.net/bsri.html>.



*iulianvalentin/iStock/Thinkstock*

**Sandra Lipsitz Bem insisted that people have varying degrees of masculine and feminine traits.**

After settling on the best way to anchor the scale, the researcher’s second critical decision is to decide on the number of points in the response scale. Notice that all of the examples in this section have an odd number of points (i.e., five or seven). Odd numbers are usually preferable for rating-scale items because the middle of the scale (i.e., “3” or “4”) allows respondents to give a neutral, middle-of-the-road answer. That is, on a scale from “strongly disagree” to “strongly agree,” the midpoint can be used to indicate “neither” or “I’m not sure.” However, in some cases, a researcher may not want to allow a neutral option in a scale. Using an even number of points (e.g., four or six) essentially compels people either to agree or disagree with the statement; this type of scaling is referred to as **forced choice**.

So, how many points should the scale have? As a general rule, more points will translate into more variability in responses—the more choice people have (up to a point), the more likely they are to distribute their responses among those choices. From a researcher’s perspective, the big question is whether this variability is meaningful.

For example, if we assess college students’ attitudes about a student-fee increase, student opinions will likely vary depending on the size of the fee and the ways in which it will be used. Thus, we might prefer a five- or seven-point scale to a two-point (yes or no) scale. However, past a certain point, increases in the scale range cease to connect to meaningful variation in attitudes. In other words, the difference between a 5 and a 6 on a seven-point scale is fairly intuitive for participants to grasp. What is the real difference, though, between an 80 and an 81 on a 100-point scale? When scales become too large, researchers risk introducing another source of error variance as participants impose their own interpretations on the scaling. In sum, more points do not always translate to a better scale.

Back to the question: How many points should the scale have? The ideal compromise supported by most statisticians is to use a seven-point scale whenever possible because of the differences between scales of measurement. As the discussion in Chapter 2 explained, the way variables are measured has implications for data analyses. For the most popular statistical tests to be legitimate, variables need to be on an interval scale (i.e., with equal intervals between points) or a ratio scale (i.e., with a true zero point). Based on mathematical modeling research, statisticians have concluded that the variability generated by a seven-point scale is most likely to mimic an interval scale (e.g., Nunnally, 1978). So, from a statistical perspective, a seven-point scale is ideal because it allows us the most flexibility in data analyses.

## Finalizing the Questionnaire

After constructing the questionnaire items, researchers face one last important step before beginning data collection. This section discusses a few guidelines for assembling the items into a coherent questionnaire. One main goal at this stage is to think carefully about the order of the individual items.

First, keep in mind that the first few questions will set the tone for the rest of the questionnaire. It is best to start with questions that are both interesting and nonthreatening to help ensure that respondents complete the questionnaire with open minds. For example:

BAD OPENING: "Do you agree that your child's teacher is an idiot?" (threatening, and also a leading question)

BETTER OPENING: "How would you rate the performance of your child's teacher?"

BAD OPENING: "Would you support a 1% sales tax increase?" (boring)

BETTER OPENING: "How do you feel about raising taxes to help fund education?"

Second, strive whenever possible to have continuity in the different sections of the questionnaire. Imagine constructing a survey to give to college freshmen. It might include questions on family background, stress levels, future plans, campus engagement, and so on. The survey will be most effective if it groups questions by topic. So, for instance, students respond to a set of questions about future plans on one page and then answer a set of questions about campus engagement on another page. This approach makes it easier for participants to progress through the questions without having to switch mentally between topics.

Third, remember that individual questions are always read in context. This means that if the college-student survey begins with a question about plans for the future and then asks about stress, respondents will likely have their future plans in mind when they think about their stress level. Consider again the example of the graduate assistant. His department used to administer a gigantic survey packet (on paper) to the 2,000 students enrolled in Introductory Psychology each semester. One year, a faculty member included a measure of identity, asking participants to complete the statements "I am\_\_\_\_" and "I am not\_\_\_\_." As researchers started to analyze data from this survey, they discovered an astonishing 60% of students had filled in the blank with "I am not a homosexual!" This response seemed downright strange, until the surveyors realized that the questionnaire immediately preceding the identity one measured prejudice toward gay and lesbian individuals. So, as these students completed the

identity measure, they had homosexuality on their minds and felt compelled to point out that they were not homosexual. In other words, responses are all about context.

Finally, after assembling a draft version of the questionnaire, perform a test run. This test run, called **pilot testing**, involves giving the questionnaire to a small sample of people, getting their feedback, and making any necessary changes. One of the best ways to pilot test is to find a patient group of friends to complete the questionnaire who will provide extensive feedback. In soliciting their feedback, ask questions like the following:

Was anything confusing or unclear?

Was anything offensive or threatening?

How long did the questionnaire take you to complete?

Did it seem repetitive or boring? Did it seem too long?

Were there particular questions that you liked or disliked? Why?

The answers to these questions will supply valuable information to revise and clarify the questionnaire before devoting resources to a full round of data collection. The next section turns to the question of how to find and select participants for this stage of the research.

## Research: Thinking Critically

### Beauty and First Impressions

Follow the link below to a press release from the University of British Columbia, describing a recent publication by researchers in the psychology department. This study suggests that physical beauty may play a role in how easily we form first impressions of other people. As you read the article, consider what you have learned so far about the research process, and then respond to the questions below.

<http://news.ubc.ca/2010/12/21/beautiful-people-convey-personality-traits-better-during-first-impressions/>

#### Think About It:

1. Suppose the following questions were part of the questionnaire given after the three-minute one-on-one conversations in this study. Based on the goals of the study and the rules discussed in this chapter, identify the problem with each of the following questions and suggest a better item.

a) Jane is very neat.

1            2            3            4            5

strongly disagree            strongly agree

main problem:

better item:

b) Jane is generous and organized.

1            2            3            4            5

strongly disagree            strongly agree

main problem:

better item:

c) Jane is extremely attractive.            TRUE            FALSE

main problem:

better item:

2. What are the strengths and weaknesses of using a fixed-format questionnaire in this study versus open-ended responses?
3. The researchers state that they took steps to control for the “positive bias that can occur in self-reporting.” How might social desirability influence the outcome of this particular study? What might the researchers have done to reduce the effect of social desirability?

## 4.3 Sampling From the Population

At this point, the chapter should have conveyed an understanding of how to construct survey items. The next step is to find a group of people to fill out the survey. But where does a researcher find this group? And how many people are needed? On the one hand, researchers want as many people as possible to capture the full range of attitudes and experiences. On the other hand, they have to conserve time and other resources, which often means choosing a smaller sample of people. This section examines the strategies researchers can use to select samples for their studies.

Researchers refer to the entire collection of people who could possibly be relevant to a study as the **population**. For example, if we were interested in the effects of prison overcrowding, our population would consist of prisoners in the United States. If we wanted to study voting behavior in the next presidential election, the population would be U.S. residents eligible to vote. And if we wanted to know how well college students cope with the transition from high school, our population would include every college student enrolled in every college in the country.

These populations suggest an obvious practical complication. How can we get every college student—much less every prisoner—in the country to fill out our questionnaire? We cannot; instead, researchers will collect data from a sample, a subset of the population. Instead of trying to reach all prisoners, we might sample inmates from a handful of state prisons. Rather than attempt to survey all college students in the country, researchers often restrict their studies to a collection of students at one university.

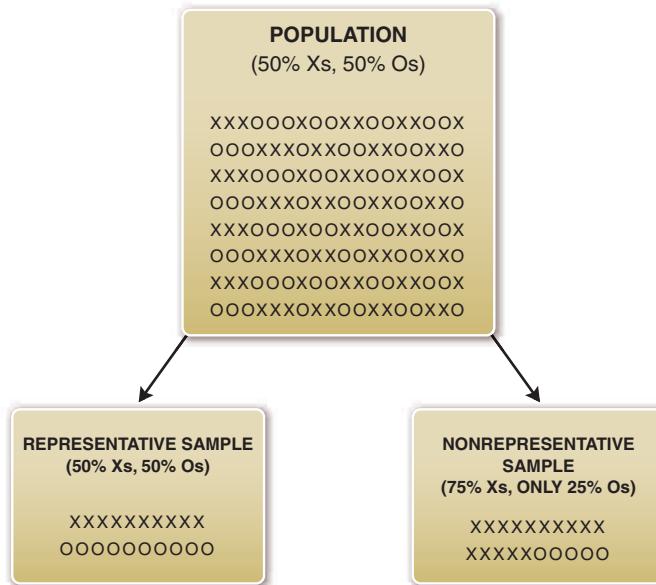
The goal in choosing a sample is to make it as representative as possible of the larger population. That is, if researchers choose students at one university, they need to be reasonably similar to college students elsewhere in the country. If the phrase “reasonably similar” sounds vague, this is because the basis for evaluating a sample varies depending on the hypothesis

and the key variables. For example, if we wanted to study the relationship between family income and stress levels, we would need to make sure that our sample mirrored the population in the distribution of income levels. Thus, a sample of students from a state university might be a better choice than students from, say, Harvard (which costs about \$60,000 per year including room and board). On the other hand, if the research question deals with the pressures faced by students in selective private schools, then Harvard students could be a representative sample for the study.

Figure 4.1 shows a conceptual illustration of both a representative and nonrepresentative sample, drawn from a larger population. The population in this case consists of 144 individuals, split evenly between Xs and Os. Thus, we would want our sample to come as close as possible to capturing this 50/50 split. The sample of 20 individuals on the left is representative of the sample because it is split evenly between Xs and Os. But the sample of 20 individuals on the right is nonrepresentative because it contains 75% Xs. Because the population has far fewer Os than we might expect, this sample does not accurately represent the population. This failure of the sample to represent the population is also referred to as **sampling bias**.

From where do these samples come? Broadly speaking, researchers have two broad categories of sampling strategies at their disposal: probability sampling and nonprobability sampling.

**Figure 4.1: Representative and nonrepresentative samples of a population**



## Probability Sampling

Researchers use **probability sampling** when each person in the population has a known chance of being in the sample. This is possible *only* in cases where researchers know the exact size of the population. For instance, the current population of the United States is 322.1 million people ([www.census.gov/popclock/](http://www.census.gov/popclock/)). If we were to select a U.S. resident at random, each

resident would have a one in 322.1 million chance of being selected. Whenever researchers have this information, probability-sampling strategies are the most powerful approach because they greatly increase the odds of getting a representative sample. Within this broad category of probability sampling are three specific strategies: simple random sampling, stratified random sampling, and cluster sampling.

**Simple random sampling**, the most straightforward approach, involves randomly picking study participants from a list of everyone in the population. The term for this list is a **sampling frame** (e.g., imagine a list of every resident of the United States). To have a truly representative random sample, researchers must have a sampling frame; they must choose from it randomly; and they must have a 100% response rate from those selected. (As Chapter 2 discussed, if people drop out of a study, it can threaten the validity of the hypothesis test.)

Researchers use **stratified random sampling**, a variation of simple random sampling, when subgroups of the population might be left out of a purely random sampling process. Imagine a city with a population that is 80% Caucasian, 10% Hispanic, 5% African American, and 5% Asian. If we were to pick 100 residents at random, the chances are very good that our entire sample would consist of Caucasian residents and ignore the perspective of all ethnic minority residents. To prevent this problem, researchers use stratified random sampling—breaking the sampling frame into subgroups and then sampling a random number from each subgroup. In this example, we could divide the list of residents into four ethnic groups and then pick a random 25 from each of these groups. The end result would be a sample of 100 people that captured opinions from each ethnic group in the population. Notice that this approach results in a sample that does not exactly represent the underlying population—that is, Hispanics constitute 25% of the sample, rather than 10%. One way to correct for this issue is to use a statistical technique known as “weighting” the data. Although the full details are beyond the scope of this book, weighting involves trying to correct for problems in representation by assigning each participant a weighting coefficient for analyses. In essence, people from groups that are underrepresented would have a weight greater than 1, while those from groups that are overrepresented would have a weight less than 1. For more information on weighting and its uses, see <http://www.applied-survey-methods.com/weight.html>.



*bowdenimages/iStock/Thinkstock*

**In a neighborhood with a majority of Caucasian residents, stratified random sampling is needed to capture the perspective of all ethnic groups in the community.**

Finally, researchers employ **cluster sampling**, another variation of random sampling, when they do not have access to a full sampling frame (i.e., a full list of everyone in the population). Imagine that we want to study how cancer patients in the United States cope with their illness. Because no list exists of every cancer patient in the country, we have to get a little creative with our sampling. The best way to think about cluster sampling is as

“samples within samples.” Just as with stratified sampling, we divide the overall population into groups, but cluster sampling differs in that we are dividing into groups based on more than one level of analysis. In our cancer example, we could start by dividing the country into regions, then randomly selecting cities from within each region, and then randomly selecting hospitals from within each city, and finally randomly selecting cancer patients from each hospital. The end result would be a random sample of cancer patients from, say, Phoenix, Miami, Dallas, Cleveland, Albany, and Seattle; taken together, these patients would provide a fairly representative sample of cancer patients around the country.

## Nonprobability Sampling

The other broad category of sampling strategies is known as **nonprobability sampling**. These strategies are used in the (remarkably common) case in which researchers do not know the odds of any given individual’s being in the sample. This uncertainty represents an obvious shortcoming—if we do not know the exact size of the population and do not have a list of everyone in it, we have no way to know that our sample is representative. Despite this limitation, researchers use nonprobability sampling on a regular basis. We will discuss two of the most common nonprobability strategies here.

In many cases, it is not possible to obtain a sampling frame. When researchers study rare or hard-to-reach populations or study potentially stigmatizing conditions, they often recruit by word-of-mouth. The term for this is **snowball sampling**—imagine a snowball rolling down a hill, picking up more snow (or participants) as it goes. If we wanted to study how often homeless people took advantage of social services, we would be hard pressed to find a sampling frame that listed the homeless population. Instead, we could recruit a small group of homeless people and ask each of them to pass the word along to others, and so on. If we wanted to study changes in people’s identities following sex-reassignment surgery, we would find it difficult to track down this population via public records. Instead, we could recruit one or two patients and ask for referrals to others. The resulting sample in both cases is unlikely to be representative, but researchers often have to compromise for the sake of obtaining access to a population. Snowball sampling is most often used in qualitative research, where the advantages of gaining a rich narrative from these individuals outweigh the loss of representativeness.

One of the most popular nonprobability strategies is known as **convenience sampling**, or simply including people who show up for the study. Any time a 24-hour news station announces the results of a viewer poll, they are likely based on a convenience sample. CNN and Fox News do not randomly select from a list of their viewers; they post a question onscreen or online, and people who are motivated (or bored) enough to respond will do so. As a matter of fact, the vast majority of psychology research studies are based on convenience samples of undergraduate college students. Research in psychology departments often works like this: Experimenters advertise their studies on a website, and students enroll in these studies, either to earn extra cash or to fulfill a research requirement for a course. Students often pick a particular study based on whether it fits their busy schedules or whether the advertisement sounds interesting. These decisions are hardly random and, consequently, neither is the sample. The goal here is not to disparage all psychology research—that would be self-defeating—but to emphasize that all of the decisions researchers make have both pros and cons.

## Choosing a Sampling Strategy

Although researchers always strive for a representative sample, no such thing as a perfectly representative one exists. Some degree of **sampling error**, defined as the degree to which the characteristics of the sample differ from the characteristics of the population, is always present. Instead of aiming for perfection, then, researchers aim for an estimate of how far from perfection their samples are. These estimates are known as the **margin of error**, or the degree to which the results from a particular sample are expected to deviate from the population as a whole.

One of the main advantages of a probability sample is that we are able to calculate these errors, as long as we know our sample size and desired level of confidence. In fact, most of us encounter margins of error every time we see the results of an opinion poll. For example, CNN may report that “Candidate A is leading the race with 60% of the vote,  $\pm 3\%$ .” This means Candidate A’s approval percentage in the sample is 60%, but based on statistical calculations, her real percentage is between 57% and 63%. The smaller the error (3% in this example), the more closely the results from the sample match the population. Naturally, researchers conducting these opinion polls want the error of estimation to be as small as possible. How persuaded would anyone be to learn that “Candidate A has a 10-point lead, plus or minus 20 points?” This margin of error ought to trigger our skepticism, because the real difference is between 30 points and -10 points—i.e., a 10-point lead for *the other candidate*.

Researchers’ most direct means of controlling the margin of error is by changing the sample size. Most survey research aims for a margin of error of less than five percentage points. Based on standard calculations, this requires a sample size of 400 people per group. That is, if we want to draw conclusions about the entire sample (e.g., “30% of registered voters said X”), then we would need at least 400 respondents to say this with some confidence. If we want to draw conclusions about subgroups (e.g., “30% of women compared to 50% of men”), then we would actually need at least 400 respondents of each gender to draw conclusions with confidence.

The magic number of 400 represents a compromise—a researcher is willing to accept 5% error for the sake of keeping time and costs down. It is worth noting, however, that some types of research have more stringent standards: For political polls to be reported by the media, they must have at least 1,000 respondents, which brings the margin of error down to three percentage points. In contrast, some areas of applied research may have more relaxed standards. In marketing research, for example, budget considerations sometimes lead to smaller samples, which means drawing conclusions at lower levels of confidence. For example, with a sample size of 100 people per group, researchers have to contend with 8–10% margin of error—almost double the error, but at a fraction of the costs.

If probability sampling is so powerful, why are nonprobability strategies so popular? One reason is that convenience samples are more practical; they are cheaper, easier, and almost always possible to conduct with relatively few resources because researchers can avoid the costs of large-scale sampling. A second reason is that convenience is often a good-enough starting point for a new line of research. For example, if we wanted to study the predictors of relationship satisfaction, we could start by testing our hypotheses in a controlled setting using college student participants and then extend the research to the study of adult married couples. Finally, and relatedly, in many cases it is acceptable to have a nonrepresentative sample because researchers do not need to generalize results. If we want to study the prevalence

of alcohol use in college students, it may be perfectly acceptable to use a convenience sample of college students. Although, even in this case, researchers would have to keep in mind that they are studying drinking behaviors among students who volunteered to complete a study on drinking behaviors.

In some cases, however, it is critical to use probability sampling, despite the extra effort required. Specifically, researchers use probability samples any time it is important to generalize and any time it is important to predict behavior of a population. The best way of understanding these criteria is to think of political polls. In the lead-up to an election, each campaign is invested in knowing exactly what the voting public thinks of its candidate. In contrast to a CNN poll, which is based on a convenience sample of viewers, polls conducted by a campaign will be based on randomly selected households from a list of registered voters. The resulting sample is much more likely to be representative, much more likely to tell the campaign how the entire population views its candidate, and therefore much more likely to be useful.

## 4.4 Analyzing Survey Data

Now comes the fun part. Once researchers have designed a survey, chosen an appropriate sample, and collected some data, it is time for analyses. As with the descriptive designs Chapter 3 explained, the goal of these analyses is to subject hypotheses to a statistical test. Surveys can be used both to describe and predict thoughts, feelings, and behaviors. Since Chapter 3 already covered the basics of descriptive analysis, this section will focus on predictive analyses, which are designed to assess the associations between and among variables. Researchers typically use three approaches to test predictive hypotheses: correlational analyses, chi-square analyses, and regression analyses. Each has its advantages and disadvantages, and each is most appropriate for a different kind of data. This section will walk through the basics of each analysis. Because the statistics course discusses these approaches in more detail, the goal here is to acquire a more conceptual overview of each technique and its usefulness in answering research questions.

### Correlational Analysis

The beginning of this chapter described an example of a survey research question: What is the relationship between the number of hours that students spend studying and their grades in the class? In this case, the hypothesis claims that we can predict something about students' grades by knowing how many hours they spend studying.

Imagine we collected a small amount of data (shown in Table 4.1) to test this hypothesis. (Of course, a true test of this hypothesis would require more than 10 people in the sample, but these data will do as an illustration.)

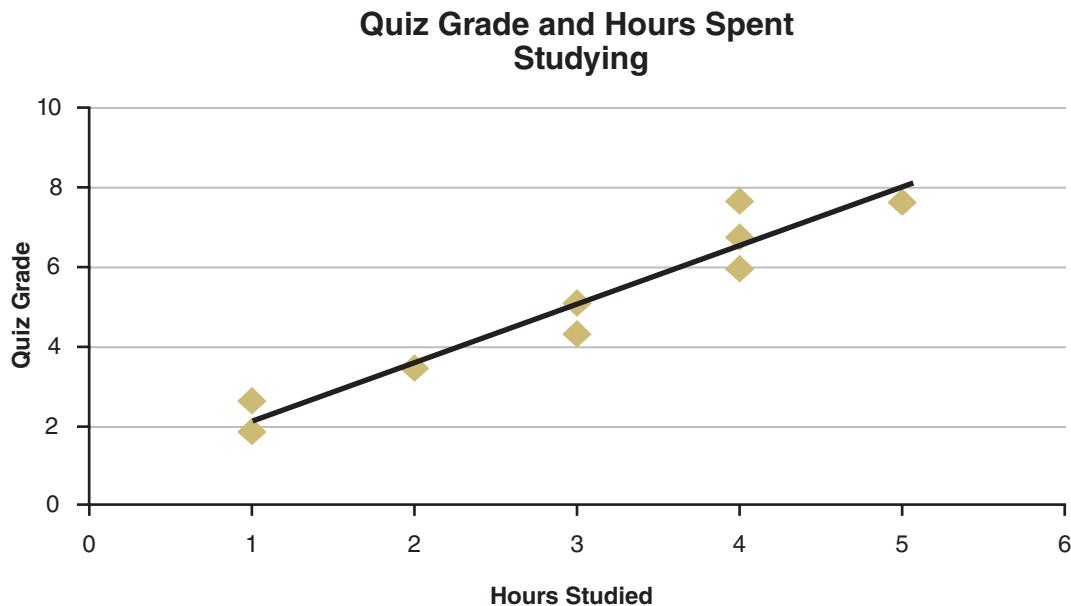
**Table 4.1: Data for quiz grade/hours studied example**

Participant	Hours Studied	Quiz Grade
1	1	2
2	1	3
3	2	4
4	3	5
5	3	6
6	3	6
7	4	7
8	4	8
9	4	9
10	5	9

### ***The Logic of Correlation***

The important question here is whether and to what extent we can predict grades based on study time. One common statistic for testing these kinds of hypotheses is a **correlation**, which gives an assessment of the linear relationship between two variables. A stronger correlation between two variables indicates a stronger association between them. In the case of the current example, the stronger the correlation between study time and quiz grade, the more accurately we can predict grades based on knowing how long the student spends studying.

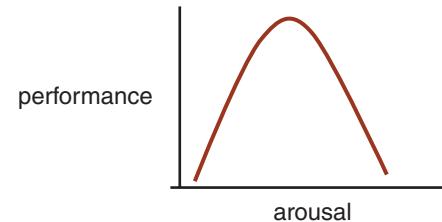
Before we calculate the correlation between these variables, it is always a good idea to visualize the data on a graph. Chapter 3 discussed a type of graph, called a scatterplot, that displays points of data on two variables at a time. The scatterplot in Figure 4.2 shows our sample data from the studying/quiz grade study.

**Figure 4.2: Scatterplot for quiz grade/hours studied example**

Each point on the graph represents one participant. For example, the point in the top right corner represents a student who studied for five hours and earned a 9 on the quiz. The two points in the bottom right represent students who studied for only one hour and earned a 2 and a 3 on the quiz.

Researchers have three reasons to graph data before conducting statistical tests. First, a graph allows us to get a general sense of the pattern—in this case, students who study less appear to do worse on the quiz. As a result, we will be better informed going into our statistical calculations. Second, the graph lets us examine the raw data for any outliers, or points that stand out as clear exceptions to the overall pattern. These outlier points may indicate that a respondent misunderstood the question and should be dropped from analyses. On the other hand, a cluster of outlier points could indicate the presence of subgroups within our data. Perhaps most students do worse if they study less, but a group of students is able to ace the quizzes without any preparation. Examining this cluster of people in more detail might suggest either a refinement of our hypothesis or an interesting direction for future research.

Third, the graph assures researchers that there is a linear relationship between the variables. This is a very important point about correlations: The math of the standard correlation formula is based on how well the data points fit a straight line, which means nonlinear relationships might be overlooked. Figure 4.3 demonstrates a robust nonlinear finding in psychology regarding the relationship between task performance and physiological arousal. As this graph shows, people tend to perform their best on just about any task when they have a moderate level of arousal.

**Figure 4.3: Curvilinear relationship between arousal and performance**

When arousal is too high, people find it difficult to calm down and concentrate; when arousal is too low, people find it difficult to care about the task at all. If we simply ran a standard correlation with data on performance and arousal, the correlation would be zero because the points do not fit a straight line. Thus, it is critical to visualize the data before jumping ahead to the statistics. Otherwise, researchers risk overlooking an important finding in the data. (It is important to note that non-linear relationships like this one *can* still be analyzed, but the calculations quickly become complex. In fact, these analyses even require specialized knowledge to use statistical software.)

## Interpreting Coefficients

Once we are satisfied that our data look linear, it is time to calculate our statistics. Researchers typically calculate using a computer software program, such as SPSS, SAS, or Microsoft Excel. The number used to quantify the correlation is called the **correlation coefficient**. This number ranges from  $-1$  to  $+1$  and contains two important pieces of information:

- The *direction* of the relationship is based on the sign of the correlation coefficient. A  $+0.8$  would indicate a positive correlation, meaning that as one variable increases, so does the other variable. A  $-0.8$  would indicate a negative correlation, meaning that as one variable increases, the other variable decreases. (Refer back to Section 2.1 for a review of these two terms.)
- The *size* of the relationship is based on the absolute value of the correlation coefficient. The farther the coefficient is from zero in either direction, the stronger the relationship between variables. For example, both a  $+0.8$  and a  $-0.8$  indicate strong relationships.

So, for example, a  $+0.2$  represents a weak positive relationship and a  $-0.7$  represents a strong negative relationship.

Calculating the correlation for our quiz-grade study produces a coefficient of  $0.962$ , indicating a strong positive relationship between studying and quiz grade. What does this mean in plain English? Students who spend more hours studying tend to score higher on the quiz.

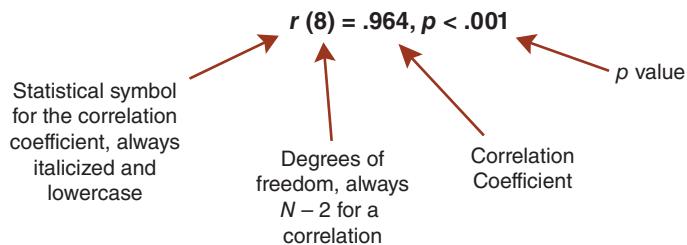
How do we know whether to get excited about a correlation of  $0.962$ ? As with all of our statistical analyses, we look this value up in a critical value table, or, more commonly, let the computer software do this for us. The critical value table provides a *p* value representing the odds that our correlation is due to random chance. In this case, the *p* value is less than  $0.001$ . This means that the chance of our correlation being a random fluke is less than 1 in 1,000; we can feel pretty confident in our results.

When interpreting correlation results, realize that statistical significance is closely tied to the sample size. In a small sample, it is possible to see moderate to strong relationships that do not meet the threshold for statistical significance. One good option in these cases is to collect additional data. If the correlation maintains its size and also attains statistical significance, researchers can have some confidence in the results. It is also possible to have the opposite problem: Large sample sizes can make even the smallest relationships show high levels of statistical significance. In a 2008 journal article, Newman, Groom, Handelman and Pennebaker analyzed differences in language use between men and women. Because the authors had a sample of over 14,000 text samples, even the tiniest differences in language were statistically

significant. For example, men used words related to anger about 4% more than women; with such a large sample, this trivial difference was significant at  $p < 0.05$ . To deal with this issue, the authors chose to use a more conservative threshold of  $p < 0.001$ , considering all other results to be too trivial.

Returning to our quiz-grade study, we now have all the information we need to report this correlation in a research paper. The standard way of reporting a correlation coefficient includes information about the sample size ( $N$ ) and  $p$  value, as well as the coefficient itself. Our quiz-grade study would be reported as Figure 4.4 depicts.

**Figure 4.4: Correlation coefficient diagram**



Where, then, does this leave our hypothesis? We started by predicting that students who spent more time studying would perform better on their quizzes than those who spent less time studying. We then designed a study to test this hypothesis by collecting data on study habits and quiz grades. Finally, we analyzed these data and found a significant, strong, positive correlation between hours studied and quiz grade. Based on this study, our hypothesis has been confirmed—students who study more have higher quiz grades. Of course, because this is a correlational study, we are unable to make causal statements. It could be that studying more for an exam helps students to learn more. Or, it could be the case that previous low quiz grades make students give up and study less. A third variable of motivation could cause students both to study more and perform better on the quizzes. To tease these explanations apart and determine causality calls for a different type of research design, which Chapter 5 will discuss.

## Multiple Regression Analysis

Correlations are the best tool to test the linear relationship between pairs of quantitative variables. However, in many cases, researchers are interested in comparing the influence of several variables at once. Imagine we want to expand the study about hours studying and quiz grade by looking at other variables that might predict students' quiz grades. We have already learned that the hours students spend studying correlate positively with their grades. But what about SAT scores? Will students with higher standardized-test scores do better in all of their college classes? What about the number of classes that students have previously taken in the subject area? Will increased familiarity with the subject be associated with higher scores? To compare the influence of all three variables, we can use a slightly different analytic approach. **Multiple regression** is a variation on correlational analysis in which more than one predictor variable is used to predict a single outcome variable. In this example, we would

attempt to predict the outcome variable of quiz scores based on three predictor variables: SAT scores, number of previous classes, and hours studied.

Multiple regression requires an extensive set of calculations; consequently, it is always performed by computer software. A detailed look at these calculations is beyond the scope of this book, but a conceptual overview will help convey the unique advantages of this analysis. Essentially, the calculations for multiple regression are based on the correlation coefficients between each of our predictor variables, as well as between each of these variables and the outcome variable. Table 4.2 shows these correlations for our revised quiz-grade study. If we scan the top row, we see the correlations between quiz grade and the three predictor variables: SAT ( $r = 0.14$ ), previous classes ( $r = 0.24$ ), and hours studied ( $r = 0.25$ ). The remainder of the table shows correlations between the various predictor variables; for example, hours studied and previous classes correlate at  $r = 0.24$ . When researchers conduct multiple regression analysis using computer software, the software will use all of these correlations in performing its calculations.

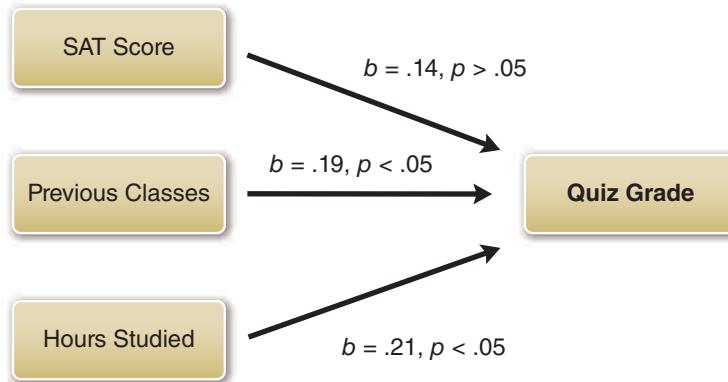
**Table 4.2: Correlations for a multiple regression analysis**

	Quiz Grade	SAT Score	Previous Classes	Hours Studied
Quiz Grade	—	0.14	0.24*	0.25*
SAT Score		—	.02	-.02
Previous Classes			—	0.24*
Hours Studied				—

*Note.* Correlations marked with an asterisk (\*) are statistically significant at the 95% confidence level. This notation in results tables is common and allows researchers to quickly spot the most interesting findings.

The advantage of multiple regression is that it considers both the individual and the combined influence of the predictor variables. Figure 4.5 shows a visual diagram of the individual predictors of quiz grades. The numbers along each line are known as **regression coefficients**, or **beta weights**. These values are very similar to correlation coefficients but differ in an important way: They represent the effects of each predictor variable while controlling for the effects of all the other predictors. That is, the value of  $b = 0.21$  linking hours studied with quiz grades is the independent contribution of hours studied, controlling for SAT scores and previous classes. If we compare the size of these regression coefficients, we see that, in fact, hours spent studying is still the largest predictor of quiz grades ( $b = 0.21$ ), compared to both SAT scores ( $b = 0.14$ ) and previous classes ( $b = 0.19$ ).

Even if individual variables only have a small influence, they can add up to a larger combined influence. So, if we were to analyze the predictors of quiz grades in this study, we would find a combined **multiple correlation coefficient** of  $r = 0.34$ . The multiple correlation coefficient represents the combined association between the outcome variable and the full set of predictor variables. Note that in this case, the combined  $r$  of 0.34 is larger than any of the individual correlations in Table 4.2, which ranged from 0.14 to 0.25. These numbers mean that we are better able to predict quiz grades from examining all three variables than we are from examining any single variable. Or, as the saying goes, the whole is greater than the sum of its parts.

**Figure 4.5: Predictors of quiz grades**

Multiple regression is an incredibly useful and powerful analytic approach, but it can also be a difficult concept to grasp. Before moving on, we will revisit the concept in the form of an analogy. Imagine someone has just eaten the most delicious hamburger of his life and is determined to understand what makes it so good. Many things contribute to the taste of the hamburger: the quality of the meat, the type and amount of cheese, the freshness of the bun, perhaps the smoked chili peppers layered on top. If the diner were to approach this investigation using multiple regression, he would be able to distinguish the influence of each variable (how important is the cheese compared to the smoked peppers?) as well as take into account the full set of ingredients (does the freshness of the bun really matter when the other elements taste so good?). Ultimately, the individual would be armed with the knowledge of which elements are most important in crafting the perfect hamburger and would understand more about the perfect hamburger than if he had examined each ingredient in isolation.

## Chi-Square Analyses

Both correlations and regressions are well suited to testing hypotheses about prediction, as long as we can demonstrate a linear relationship between two variables. Linear relationships, however, require that variables be measured on one of the quantitative scales, that is, ordinal, interval, or ratio scales (see Section 2.3 for a review). What if we want to test an association between nominal, or categorical, variables? In these cases, we need an alternative statistic called the **chi-square statistic**, which determines whether two nominal variables are independent from or related to one another. Chi-square is often abbreviated with the symbol  $\chi^2$ , which shows the Greek letter chi with the superscript 2 for squared. (This statistic is also referred to as the chi-square test for independence—a slightly longer but more descriptive synonym.)

The idea behind this test is similar to that of the correlation coefficient. If two variables are independent, then knowing the value of one variable does not tell us anything about the value of the other variable. As we will see in the example below, a larger chi-square reflects a larger deviation from what we would expect by chance and is thus an index of statistical significance.

Imagine that we want to know whether people in rural or urban areas are more likely to support a sales-tax increase. We can easily speculate why either group might be more likely to do so—perhaps people living in cities are more politically liberal or perhaps people living in small towns are better able to see benefits of higher local taxes. So, we might survey a sample of 100 people, asking them to indicate both their location (rural or urban) and their support for a sales-tax proposal. The survey produces the following results (in Table 4.3), presented in a **contingency table**, which displays the number of individuals in each combination of our nominal variables. Notice that we have more urban than rural residents, reflecting the higher population density in cities.

But, as it turns out, the raw numbers are less important than the ratios within each group. The chi-square calculation works by first considering what each cell in the table would look like *if there were no relationship at all* (i.e., under the null hypothesis), and then determining how much the data differ from that reference point.

**Table 4.3: Chi-square example: support for a sales tax increase**

	Rural	Urban	Total
Support	10	45	55
Don't Support	30	15	45
Total	40	60	100

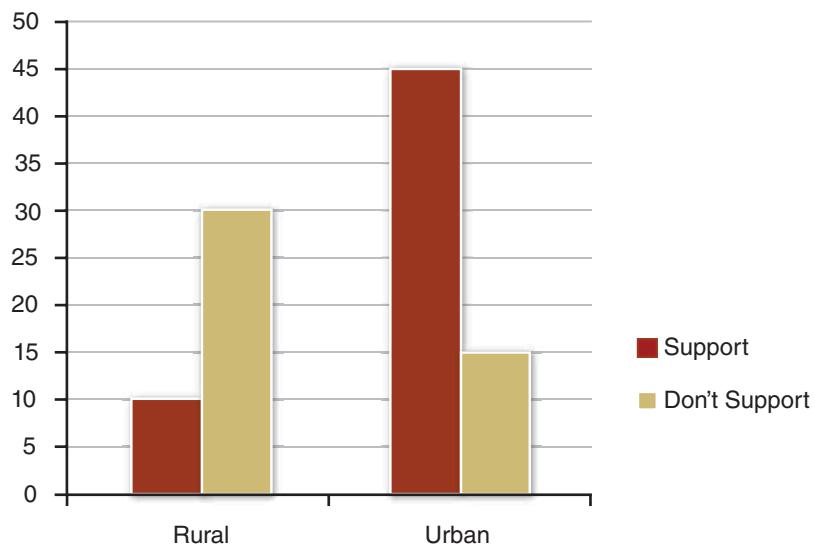
In this example, our final chi-square value is 34.55; this represents the total difference across the table between actual and expected data. The larger this number is, the more our observed data differ from the expected frequencies, and the more our variables relate to one another. In the current example, this means we can predict a person's support for a sales-tax increase based on where he or she lives, which is consistent with our initial hypothesis.

Still, how do we know if our value of 34.55 is meaningful? As with the other statistical tests we have discussed, determining the significance requires looking up the result in a critical-value table to assess whether the calculated value is above threshold. In this case, the critical value for a chi-square with a  $2 \times 2$  table = 3.84, so we can feel confident in our value of 34.55—almost 10 times higher than the threshold value.

However, unlike correlation and regression coefficients, our chi-square results cannot tell us anything about the direction or magnitude of the relationship. A larger chi-square reflects a larger deviation from what we would expect by chance and is thus an index of statistical significance. To interpret the patterns of our data, we need to visually inspect the numbers in our data table. Better yet, we can create a bar graph like we did in Chapter 3 to display these frequencies visually.

As Figure 4.6 shows, the cell frequencies suggest a fairly clear interpretation: People who live in urban settings are much more likely than people who live in rural settings to support a sales-tax increase. In fact, urban residents *support* the increase by a 3-to-1 margin, while rural residents *oppose* the increase by a 3-to-1 margin.

**Figure 4.6: Graph of chi-square results**



## Research: Thinking Critically

### Self-Esteem in Youth and Early Adulthood

Follow the link below to read a press release from the American Psychological Association, describing recent research on self-esteem during adolescence. This study, by a group of Swiss researchers, challenges some of our popular assumptions about gender differences in self-esteem. 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.apa.org/news/press/releases/2011/07/youth-self-esteem.aspx>

#### Think About It:

1. Why is self-esteem a good topic to study using survey research methods? Does using a survey to study self-esteem present any weaknesses?
2. What type of sampling was used in this study? Was this an appropriate strategy?
3. What type of data analysis discussed in this chapter is appropriate to understanding the influence of multiple variables (mastery, health, income) on self-esteem?

## Summary and Resources

### Chapter Summary

This chapter has covered the process of survey research from conceptualization through analysis. We first discussed the types of research questions that are best suited to survey research—essentially, those that can be answered based on people's observations of their own behavior. Survey research can involve either verbal reports (i.e., interviews) or written reports (i.e., questionnaires). In both cases, surveys are distinguished by their reliance on people's self-reports of their attitudes, feelings, and behaviors.

This chapter covered several key points for writing survey items. The key takeaway of the five rules for better questions is that questions should be written as clearly and unambiguously as possible. This helps to minimize the error variance that might result from participants imposing their own guesses and interpretations on the material. When designing survey items, researchers also have a broad choice between open-ended and fixed-format responses. The former provide richer and more extensive data but are harder to score and code; the latter are easier to code but can constrain people's responses to a researcher's choice of categories. If and when researchers settle on a fixed-format response, they have another set of decisions to make regarding the response scaling, labels, and general format.

Once researchers have constructed the scale, it is time to begin data collection. This chapter discussed the concept of sampling, or choosing a portion of the population to use for a study. Broadly speaking, sampling can be either "probability" or "nonprobability," depending on whether researchers have a known population size from which they sample randomly. Probability sampling is more likely to result in a representative sample, but this approach is not possible in all studies. In fact, a significant proportion of psychology research studies use a form of nonprobability sampling called convenience sampling, meaning that the sample consists of those who show up for the study.

Finally, this chapter covered three approaches to analyzing survey data and testing hypotheses about prediction. The first, correlational analysis, is a very popular way to analyze survey data. The correlation is a statistical test that assesses the linear relationship between two variables. The stronger the correlation between variables, the more we can accurately predict one based on knowing the other. Second, regression analyses allow us to expand our investigations into multiple predictors. Multiple regression offers the advantage of considering both the individual and the combined influence of the predictor variables. However, both correlation and regression require the variables to be quantitative—that is, measured on an ordinal, interval, or ratio scale. In cases where our survey produces nominal or categorical data, we use an alternative called the chi-square statistic, which determines whether two nominal variables are independent or related. The chi-square works by examining the extent to which our observed data deviate from the pattern we would expect if the variables were unrelated.

The common thread in all these analyses is that while they measure the association between variables, they do not tell us anything about the causal relationship between them. To make causal statements, we have to conduct experiments, which the next chapter will discuss.

## Key Terms

**anchors** Labels, or endpoints, for a rating scale.

**bipolar scale** Rating scale that has polar opposites as its anchors.

**branching schedule** An interview format in which questions branch in different directions depending on participants' answers.

**chi-square statistic** A statistical test similar to the correlation coefficient; determines whether two nominal variables are independent or related.

**cluster sampling** A variation of simple random sampling that involves dividing the sample into groups based on more than one level of analysis.

**contingency table** A data summary table that shows the number of individuals in each combination of the nominal variables; used as the first step in calculating chi-square.

**convenience sampling** A nonprobability sampling strategy that involves simply enrolling people who show up for the study.

**correlation** Statistical test that gives an assessment of the linear relationship between two variables; the stronger the correlation between variables, the more accurately the prediction about one is based on knowing the other.

**correlation coefficient** The number used to quantify a correlation; this coefficient ( $r$ ), ranges from  $-1$  to  $+1$  and contains information about both the size and direction of a correlation.

**double-barreled question** A flawed survey item that asks more than one question at a time.

**fixed-format response** Answer to a limiting question or statement, involving choosing from a list of options; on a survey, fixed-format responses are easier to code but can constrain the data into narrow categories.

**forced choice** A rating scale that requires respondents to agree or disagree with a statement, usually through the use of an even number of scale points.

**interview** A survey that is administered verbally.

**interview schedule** A plan or script for the progress of an interview that describes the list of questions and the order in which they should be asked.

**leading question** A flawed survey item worded in a way that suggests an answer.

**Likert scale** Format that uses anchors of "strongly agree" and "strongly disagree" to rate responses to a survey question.

**linear schedule** An interview format that asks the same questions in the same order for all participants.

**margin of error** The degree to which the data from the sample are expected to deviate from the population as a whole.

**multiple-choice format** A type of fixed-format response that asks participants to select from a set of predetermined responses.

**multiple correlation coefficient** A number that represents the combined association between the outcome variable and the full set of predictor variables.

**multiple regression** A variation on correlational analysis in which more than one predictor variable is used to predict a single outcome variable.

**nonprobability sampling** A group of sampling strategies used when the odds of any given individual's being in the sample are unknown.

**open-ended response** Unstructured answer to a question or statement; on a survey, open-ended responses provide rich data but are hard to code.

**pilot testing** "Test run" of a survey that involves giving the questionnaire to a small sample of people, getting their feedback, and making any necessary changes.

**population** The entire collection of people who could possibly be relevant for a study.

**probability sampling** A group of data-collection strategies used when each person in the population has a known chance of being in the sample.

**questionnaire** A survey that is administered in writing.

**rating scale** A fixed-format response that asks participants to place responses on a continuum.

**regression coefficients (beta weights)** Values that represent the effects of each predictor variable while controlling for the effects of all the other predictors.

**sampling bias** The failure of the sample to represent the underlying distribution in the population.

**sampling error** The degree to which the characteristics of the sample differ from the characteristics of the population.

**sampling frame** A list of all members of a particular population (e.g., a list of every resident of the United States) and a necessary requirement for probability sampling strategies.

**self-reports** Participants' reports of their own attitudes, feelings, and behaviors.

**simple random sampling** A probability sampling strategy that involves randomly picking participants from a list of everyone in the population.

**snowball sampling** A nonprobability sampling strategy that involves recruiting by word-of-mouth referrals.

**social desirability** Participants' reluctance to give unpopular answers to survey questions.

**stratified random sampling** A variation of simple random sampling, used when subgroups of the population might be left out of a purely random sampling process; involves breaking the sampling frame into subgroups and then sampling a random number from each subgroup.

**survey research** Any method that relies on people's observations of their own behavior.

**true/false format** A fixed-format response that asks participants to indicate whether they endorse a statement.

**unipolar scale** Rating scale that assess a single construct.

## Apply Your Knowledge

1. For each of the following poorly written questionnaire items, identify the major problem and then rewrite it so that the problem is resolved.
  - a. How much do you like cats and ponies?

main problem:

better item:

- b. Do you think that John McCain's complete lack of personality proved that he would have been a terrible president?

main problem:

better item:

- c. Do you dislike not playing basketball?

main problem:

better item:

- d. Do you support SB 1070?

main problem:

better item:

- e. How often do you take drugs?

main problem:

better item:

2. Dr. Truxillo is interested in Arizona residents' thoughts and feelings about global warming. For each of the following examples, identify the sampling method used by her research assistants.
  - a. Alejandra sets up a table in the mall and hands a survey to people who approach her.
  - b. Catherine randomly chooses five cities, then chooses three neighborhoods in each, then randomly samples 5,000 households for a phone survey.
  - c. Isaiah starts with a list of the entire population of Arizona and selects participants by dialing random phone numbers.
  - d. Anna obtains the master list from Isaiah and divides the population according to education level. She then randomly chooses 500 high school dropouts, 500 college graduates, and 500 people with some postgraduate education.

3. Based on each of the following study descriptions, choose whether the best analysis would be a correlation, a multiple regression, or a chi-square.
  - a. Ahmad is interested in the relationship between annual income and self-reported happiness.
  - b. Sheila is interested in whether some ethnic groups are more likely to use counseling services (a yes-or-no question).
  - c. Angela is interested in knowing the best predictors of recovery from depression, comparing the influence of drugs, therapy, and family resources.
  - d. Kartik is interested in whether high school dropouts or college graduates are more likely to vaccinate their children.
  - e. Nicole is interested in understanding the best predictors of weight loss.
  - f. Isabella is interested in the relationship between self-esteem and prejudice.

### Critical Thinking Questions

1. In survey research, explain the trade-off between the “richness” of people’s responses, and the ease of analyzing their responses.
2. When conducting interviews, the researcher has a personal interaction with the subject. Why is this both good and bad?
3. What are some of the new challenges in conducting surveys over the Internet? On mobile devices?
4. Explain the compromises between confidence level and research costs. When might researchers be willing to accept more error in their findings?

### Research Scenarios: Try It

Mindfulness can be defined as “awareness of the present moment without judgment.” Recent research has found that a tendency to be mindful in everyday life is associated with reduced risk for emotional disorders such as depression and anxiety. Suppose you are interested in understanding the role that mindfulness may play in reducing negative reactions to stressful events. You decide to investigate this relationship in a sample of college students who undergo a significant stressor at the same time: final exams. You hypothesize that college students who are less mindful will be more stressed during final exams than those who are more mindful.

1. To test this hypothesis, what psychological construct(s) would you have to measure?
  - a. Mindfulness
  - b. Mindfulness and final exams
  - c. Mindfulness and well-being
  - d. Mindfulness and stress
2. Suppose you decide that a questionnaire survey would be the most appropriate method to measure mindfulness and stress. While you can use existing questionnaire measures of stress, you decide to create your own survey measure of mindfulness. What problem might be a particular concern in constructing a questionnaire asking about the tendency to be mindful?
  - a. Social desirability
  - b. Inefficient data collection
  - c. Double-barreled questions
  - d. May be too time-consuming

3. Suppose you want to create a questionnaire item with the following base: "When I have distressing thoughts or images, I just notice them and let them go." Which of the following response options would be both easy to score and appropriate for this question?
  - a. True/false
  - b. Multiple choice
  - c. Likert-style rating scale
  - d. Open-ended
4. The design of your study requires that participants be students who undergo a final exam. Suppose you run this study at a local college and advertise by posting flyers around school inviting all students to participate in the survey. The study has no eligibility requirements other than being a student and having a final exam at the end of the quarter. What type of sampling does this represent?
  - a. Snowball sampling
  - b. Convenience sampling
  - c. Cluster sampling
  - d. Stratified random sampling
5. Suppose you administer your mindfulness questionnaire at the beginning of the semester and administer your stress questionnaire to the same students during finals. Which of the following correlations between mindfulness and stress would provide support for your original hypothesis?
  - a.  $r = 0.65, p < 0.05$
  - b.  $r = -0.65, p < 0.05$
  - c.  $r = -0.20, p > 0.10$
  - d.  $r = 0.001, p > 0.1$

See Appendix A for answers to Research Scenarios: Try It questions.