

7



Asking People About Themselves: Survey Research

- **Why Conduct Surveys?**
- **Constructing Questions to Ask**
 - Defining the Research Objectives
 - Attitudes and Beliefs*
 - Facts and Demographics*
 - Behaviors*
 - Question Wording
 - Simplicity*
 - Double-Barreled Questions*
 - Loaded Questions*
 - Negative Wording*
 - “Yea-Saying” and “Nay-Saying”*
- **Responses to Questions**
 - Closed- Versus Open-Ended Questions
 - Number of Response Alternatives
 - Rating Scales
 - Graphic Rating Scale*
 - Semantic Differential Scale*
 - Nonverbal Scale for Children*
- **Labeling Response Alternatives**
- **Finalizing the Questionnaire**
 - Formatting the Questionnaire
 - Refining Questions
- **Administering Surveys**
 - Questionnaires
 - Personal Administration to Groups or Individuals*
 - Mail Surveys*
 - Internet Surveys*
 - Other Technologies*
 - Interviews
 - Face-to-Face Interviews*
 - Telephone Interviews*
 - Focus Group Interviews*
- **Survey Designs to Study Changes Over Time**
- **Sampling From a Population**
 - Confidence Intervals
 - Sample Size

■ Sampling Techniques

Probability Sampling

Simple Random Sampling

Stratified Random Sampling

Cluster Sampling

Nonprobability Sampling

Haphazard Sampling

Quota Sampling

■ Evaluating Samples

Sampling Frame

Response Rate

Reasons for Using Convenience Samples

Study Terms

Review Questions

Activity Questions

Survey research employs questionnaires and interviews to ask people to provide information about themselves—their attitudes and beliefs, demographics (age, gender, income, marital status, and so on) and other facts, and past or intended future behaviors. This chapter will explore methods of designing and conducting surveys, including sampling techniques.

WHY CONDUCT SURVEYS?

During the past few weeks, I have read numerous stories about survey findings in my local newspaper. One was a survey of California voters to measure their preference for the two candidates for governor. Another was a nationwide survey asking adults about their current thoughts and emotions regarding the events of September 11, 2001. A nationwide survey of teenagers reported their experiences with violence. I also encountered surveys at work. My department surveyed recent graduates of the master's degree program to ask about their employment and perceptions of their educational experience in the program. A student administered a questionnaire to victims of domestic violence living at two local shelters to measure their experiences with police. Closer to home, my wife purchased a new car a few weeks ago and received a survey from the manufacturer asking about her satisfaction with the dealer and car quality. Surveys are clearly a common and important method of studying behavior.

Surveys provide us with a methodology for asking people to tell us about themselves. They have become extremely important as society demands data about issues rather than only intuition and anecdotes. My department needs data from graduates to help determine changes that should be made to the curriculum. Auto companies want data from buyers to assess and improve product quality and customer satisfaction. Without collecting such data, we are totally dependent upon stories we might hear or letters that a graduate or customer might write. Other surveys can be important for making public policy decisions by lawmakers and public agencies. In basic research, many important variables,

FIGURE 7.1
Relationship between
hours of work and
grade point average

Source: “Negative Correlates of Part-time Employment During Adolescence,” by L. Steinberg and S. M. Dornbusch, 1991, *Developmental Psychology*, 27, pp. 303–313. Copyright © 1991 by the American Psychological Association. Reprinted by permission.



including attitudes, current emotional states, and self-reports of behaviors, are most easily studied using questionnaires or interviews.

We often think of survey data providing a “snapshot” of how people think and behave at a given point in time. However, the survey method is also an important way for researchers to study relationships among variables and ways that attitudes and behaviors change over time. For example, Steinberg and Dornbusch (1991) examined the relationship between the number of hours that high school students work and variables such as grade point average, drug and alcohol use, and psychosomatic distress. The sample consisted of 3,989 students in grades 10–12 at nine high schools in California and Wisconsin. The researchers found that “long work hours during the school year are associated with lower investment and performance in school, greater psychological and somatic distress, drug and alcohol use, delinquency, and autonomy from parents” (Steinberg & Dornbusch, 1991, p. 304). Figure 7.1 shows a typical finding: There are some positive aspects of working fewer than 10 hours per week (as opposed to not being employed); however, increasingly negative effects are associated with longer work hours.

Survey research is also important as a complement to experimental research findings. Recall from Chapter 2 that Winograd and Soloway (1986) conducted experiments on the conditions that lead to forgetting where we place something. To study this topic using survey methods, Brown and Rahhal (1994) asked both younger and older adults about their actual experiences when they hid something and later forgot its location. They reported that older adults take longer than younger adults to find the object, older adults hide objects from potential thieves, whereas younger people hide things from friends and relatives. Interestingly, most lost objects are eventually found, usually by accident in a location that had been searched previously. This research illustrates a point made in previous chapters that multiple methods are needed to understand any behavior.

An assumption that underlies the use of questionnaires and interviews is that people are willing and able to provide truthful and accurate answers. Re-

searchers have addressed this issue by studying possible biases in the way people respond. A **response set** is a tendency to respond to all questions from a particular perspective rather than to provide answers that are directly related to the questions. Thus, response sets can affect the usefulness of data obtained from self-reports. The most common response set is called social desirability, or “faking good.” The social desirability response set leads the individual to answer in the most socially acceptable way—the way that “most people” are perceived to respond or the way that would reflect most favorably on the person. Social desirability can be a problem in many research areas, but it is probably most acute when the question concerns a sensitive topic such as violent or aggressive behavior, substance abuse, or sexual practices. However, it should not be assumed that people consistently misrepresent themselves. Jourard (1969) suggested that people are most likely to lie when they don’t trust the researcher. If the researcher openly and honestly communicates the purposes and uses of the research, promises to provide feedback about the results, and assures confidentiality, then the participants can reasonably be expected to give honest responses.

We turn now to the major considerations in survey research: constructing the questions that are asked, the methods for presenting the questions, and sampling of the individuals taking part in the research.

CONSTRUCTING QUESTIONS TO ASK

A great deal of thought must be given to writing questions for questionnaires and interviews. This section describes some of the most important factors to consider when constructing questions.

Defining the Research Objectives

When constructing questions for a survey, the first thing the researcher must do is explicitly determine the research objectives: What is it that he or she wishes to know? The survey questions must be tied to the research questions that are being addressed. Too often, surveys get out of hand when researchers begin to ask any question that comes to mind about a topic without considering exactly what useful information will be gained by doing so. This process will usually require the researcher to decide on the type of questions to ask. As noted previously, there are three general types of survey questions (Judd, Smith, & Kidder, 1991).

Attitudes and beliefs Questions about attitudes and beliefs focus on the ways that people evaluate and think about issues. Should more money be spent on mental health services? Are you satisfied with the way that police responded to your call? How do you evaluate this instructor?

Facts and demographics Factual questions ask people to indicate things they know about themselves and their situation. In most studies, asking some demographic information is necessary to adequately describe your sample. Age

and gender are typically asked. Depending on the topic of the study, questions on such information as ethnicity, income, marital status, employment status, and number of children might be included. Obviously, if you are interested in making comparisons among groups, such as males and females, you must ask the relevant information about group membership. It is unwise to ask such questions if you have no real reason to use the information, however.

Other factual information you might ask will depend on the topic of your survey. Each year, *Consumer Reports* magazine asks me to tell them about the repairs that have been necessary on many of the products that I own, such as my car and dishwasher. Factual questions about illnesses and other medical information would be asked in a survey of health and quality of life.

Behaviors Other survey questions can focus on past behaviors or intended future behaviors. How many times last week did you exercise for 20 minutes or longer? How many children do you plan to have? Have you ever been so depressed that you called in sick to work?

Question Wording

A great deal of care is necessary to write the very best questions for a survey. Cognitive psychologists have identified a number of potential problems with question wording (see Graesser, Kennedy, Wiemer-Hastings, & Ottati, 1999). Many of the problems stem from a difficulty with understanding the question, including (a) unfamiliar technical terms, (b) vague or imprecise terms, (c) ungrammatical sentence structure, (d) phrasing that overloads working memory, and (e) embedding the question with misleading information. Here is a question that illustrates some of the problems identified by Graesser et al.:

Did your mother, father, full-blooded sisters, full-blooded brothers, daughters, or sons ever have a heart attack or myocardial infarction?

There is memory overload because of the length of the question and the need to keep track of all those relatives while reading the question, and the respondent must worry about two different diagnoses with regard to each relative. Further, the term *myocardial infarction* may be unfamiliar to most people. How do you write questions to avoid such problems? The following items are important to consider when you are writing questions.

Simplicity The questions asked in a survey should be relatively simple. People should be able to easily understand and respond to the questions. Avoid jargon and technical terms that people won't understand. Sometimes, however, you have to make the question a bit more complex to make it easier to understand. Usually this occurs when you need to define a term or describe an issue prior to asking the question. Thus, before asking whether someone approves of Proposition J, you will probably want to provide a brief description of the content of this ballot measure.

Double-barreled questions Avoid “double-barreled” questions that ask two things at once. A question such as “Should senior citizens be given more money for recreation centers and food assistance programs?” is difficult to answer because it taps two potentially very different attitudes. If you are interested in both issues, ask two questions.

Loaded questions A loaded question is written to lead people to respond in one way. For example, the questions “Do you favor eliminating the wasteful excesses in the public school budget?” and “Do you favor reducing the public school budget?” will likely elicit different answers. Or consider that men are less likely to say they have “raped” someone than that they have “forced sex”; similarly, women are less likely to say they have been raped than forced to have unwanted sex (Koss, 1992). Questions that include emotionally charged words such as *rape*, *waste*, *immoral*, *ungodly*, or *dangerous* may influence the way that people respond and thus lead to biased conclusions.

Negative wording Avoid phrasing questions with negatives. This question is phrased negatively: “The city should not approve the proposed women’s shelter.” Agreement with this question means disagreement with the proposal. This phrasing can confuse people and result in inaccurate answers. A better format would be: “The city should approve the proposed women’s shelter.”

“Yea-saying” and “nay-saying” When you ask several questions about a topic, there is a possibility that a respondent will employ a response set to agree or disagree with all the questions. Such a tendency is referred to as “**yea-saying**” or “**nay-saying**.” The problem here is that the respondent may in fact be expressing true agreement, but alternatively may simply be agreeing with anything you say. One way to detect this response set is to word the questions so that consistent agreement is unlikely. For example, a study of family communication patterns might ask people how much they agree with the following statements: “The members of my family spend a lot of time together” and “I spend most of my weekends with friends.” Similarly, a measure of loneliness (e.g., Russell, Peplau, & Cutrona, 1980) will phrase some questions so that agreement means the respondent is lonely (“I feel isolated from others”) and others with the meaning reversed so that disagreement indicates loneliness (e.g., “I feel part of a group of friends”). Although it is possible that someone could legitimately agree with both items, consistently agreeing or disagreeing with a set of related questions phrased in both standard and reversed formats is an indicator that the individual is “yea-saying” or “nay-saying.”

Graesser and his colleagues have developed a computer program called *QUAID* (Question Understanding Aid) that analyzes question wording. Researchers can try out their questions online at the *QUAID* Web site (<http://mnemosyne.csl.psc.memphis.edu/quaid>).

RESPONSES TO QUESTIONS

Closed- Versus Open-Ended Questions

Questions may be either closed- or open-ended. With closed-ended questions, a limited number of response alternatives are given; with open-ended questions, respondents are free to answer in any way they like. Thus, you could ask a person: “What is the most important thing for children to prepare them for life?” followed by a list of answers from which to choose (a closed-ended question) or you could leave this question open-ended for the person to provide the answer.

Using closed-ended questions is a more structured approach; they are easier to code and the response alternatives are the same for everyone. Open-ended questions require time to categorize and code the responses and are therefore more costly. Sometimes a respondent’s response cannot be categorized at all because the response doesn’t make sense or the person couldn’t think of an answer. Still, an open-ended question can yield valuable insights into what people are thinking. Open-ended questions are most useful when the researcher needs to know what people are thinking and how they naturally view their world; closed-ended questions are more likely to be used when the dimensions of the variables are well defined.

Schwarz (1999) points out that the two approaches can sometimes lead to different conclusions. He cites the results of a survey question about preparing children for life. When “To think for themselves” was one alternative in a closed-ended list, 62% chose this option; however, only 5% gave this answer when the open-ended format was used. This finding points to the need to have a good understanding of the topic when asking closed-ended questions.

Number of Response Alternatives

With closed-ended questions, there are a fixed number of response alternatives. In public opinion surveys, a simple “yes or no” or “agree or disagree” dichotomy is often sufficient. In more basic research, it is often preferable to have a sufficient number of alternatives to allow people to express themselves—for example, a 5- or 7-point scale ranging from “strongly agree to strongly disagree” or “very positive to very negative.” Such a scale might appear as follows:

Strongly agree _____ Strongly disagree

Rating Scales

Rating scales such as the one just shown are very common in many areas of research. Rating scales ask people to provide “how much” judgments on any number of dimensions—amount of agreement, liking, or confidence, for example. Rating scales can have many different formats. The format that is used depends on factors such as the topic being investigated. Perhaps the best way to gain an

understanding of the variety of formats is simply to look at a few examples. The simplest and most direct scale presents people with five or seven response alternatives with the endpoints on the scale labeled to define the extremes. For example,

Students at the university should be required to pass a comprehensive examination to graduate.

Strongly agree _____ Strongly disagree

How confident are you that the defendant is guilty of attempted murder?

Not at all confident _____ Very confident

Graphic rating scale A graphic rating scale requires a mark along a continuous 100-millimeter line that is anchored with descriptions at each end.

How would you rate the movie you just saw?

Not very enjoyable _____ Very enjoyable

A ruler is then placed on the line to obtain the score on a scale that ranges from 0 to 100.

Semantic differential scale The semantic differential scale is a measure of the meaning of concepts that was developed by Osgood and his associates (Osgood, Suci, & Tannenbaum, 1957). Respondents rate any concept—persons, objects, behaviors, ideas—on a series of bipolar adjectives using 7-point scales.

Smoking Cigarettes

Good	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	Bad
Strong	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	Weak
Active	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	:	_____	Passive

Research on the semantic differential shows that virtually anything can be measured using this technique. Ratings of specific things (marijuana), places (the student center), people (the governor, accountants), ideas (abortion, tax reduction), and behaviors (attending church, using public transit) can be obtained. A large body of research shows that the concepts are rated along three basic dimensions: the first and most important is *evaluation* (e.g., adjectives such as good–bad, wise–foolish, kind–cruel); the second is *activity* (active–passive, slow–fast, excitable–calm); and the third is *potency* (weak–strong, hard–soft, large–small).

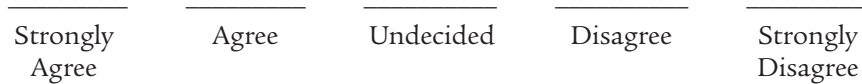
Nonverbal scale for children Young children may not understand the types of scales we’ve just described, but they are able to give ratings. For example, you could ask children to point to the face that shows how they feel about a toy:

Point to the face that shows how you feel about the toy.



Labeling Response Alternatives

The examples thus far have labeled only the endpoints on the rating scale. Respondents decide the meaning of the other response alternatives. This is a reasonable approach, and people are usually able to use such scales without difficulty. Sometimes researchers need to provide labels to more clearly define the meaning of each alternative. Here is a fairly standard alternative to the agree-disagree scale shown previously:



This type of scale assumes that the middle alternative is a “neutral” point halfway between the endpoints. Sometimes, however, a perfectly balanced scale may not be possible or desirable. Consider a scale asking a college professor to rate a student for a job or graduate program. This particular scale asks for comparative ratings of students:

In comparison with other graduates, how would you rate this student’s potential for success?



Notice that most of the alternatives are asking people to make a rating in terms of the top 25% of students. This is done because students who apply for such programs tend to be very bright and motivated, and so professors rate them favorably. The wording of the alternatives attempts to force the raters to make finer distinctions among generally very good students.

Labeling alternatives is particularly interesting when asking about the frequency of a behavior. For example, you might ask, “How often do you exercise for at least 20 minutes?” What kind of scale should you use to let people answer

this question? You could list (1) never, (2) rarely, (3) sometimes, (4) frequently. These terms convey your meaning but they are vague. Here is another set of alternatives, similar to ones described by Schwarz (1999):

- _____ less than twice a week
- _____ about twice a week
- _____ about four times a week
- _____ about six times a week
- _____ at least once each day

A different scale might be:

- _____ less than once per month
- _____ about once a month
- _____ about once every two weeks
- _____ about once a week
- _____ more than once per week

Schwarz (1999) calls the first scale a high-frequency scale because most alternatives indicate a high frequency of exercise. The other scale is referred to as low frequency. Schwarz points out that the labels should be chosen carefully because people may interpret the meaning of the scale differently, depending on the labels used. If you were actually asking the exercise question, you might decide on alternatives different from the ones described here. Moreover, your choice should be influenced by factors such as the population you are studying. If you are studying people who generally exercise a lot, you will be more likely to use a higher-frequency scale than you would if you were studying people who generally don't exercise a great deal.

FINALIZING THE QUESTIONNAIRE

Formatting the Questionnaire

A written questionnaire itself should appear attractive and professional. It should be neatly typed and free of spelling errors. People should find it easy to identify the questions and the response alternatives to the questions. Leave enough space between questions so people don't become confused when reading the questionnaire. If you have a particular scale format, such as a 5-point rating scale, use it consistently. Don't change from 5- to 4- to 7-point scales, for example.

It is also a good idea to carefully consider the sequence in which you will ask your questions. In general, it is best to ask the most interesting and important questions first to capture the attention of your respondents and motivate them to complete the survey. Roberson and Sundstrom (1990) obtained the highest

return rates in an employee attitude survey when important questions were presented first and demographic questions were asked last. In addition, it is a good idea to group questions together when they address a similar theme or topic. Doing so will make your survey appear more professional, and your respondents are more likely to take it seriously.

Refining Questions

Before actually administering the survey, it is a good idea to give the questions to a small group of people and have them “think aloud” while answering them. The participants might be chosen from the population being studied, or they could be friends or colleagues who can give reasonable responses to the questions. For the “think aloud” procedure, you will need to ask the individuals to tell you how they interpret each question and how they respond to the response alternatives. This procedure can provide valuable information that you can use to improve the questions. (The importance of pilot studies such as this is discussed further in Chapter 9.)

ADMINISTERING SURVEYS

There are two ways to administer surveys. One is to use a written questionnaire. Respondents read the questions and indicate their responses on a form. The other way is to use an interview format. An interviewer asks the questions and records the responses in a personal verbal interaction. Both questionnaires and interviews can be presented to respondents in several ways. Let’s examine the various methods of administering surveys.

Questionnaires

With questionnaires, the questions are presented in written format and the respondents write their answers. There are several positive features of using questionnaires. First, they are generally less costly than interviews. They also allow the respondent to be completely anonymous as long as no identifying information (e.g., name, Social Security or driver’s license number) is asked. However, questionnaires require that the respondents be able to read and understand the questions. In addition, many people find it boring to sit by themselves reading questions and then providing answers; thus, there may be a problem of motivation. Questionnaires can be administered in person to groups or individuals, through the mail, on the Internet, and with other technologies.

Personal administration to groups or individuals Often researchers are able to distribute questionnaires to groups or individuals. This might be a college class, parents attending a school meeting, people attending a new employee orientation, or individual students waiting for an appointment with an advisor. An advantage of this approach is that you have a captive audience

that is likely to complete the questionnaire once they start it. Also, the researcher is present so people can ask questions if necessary.

Mail surveys Surveys can be mailed to individuals at a home or business address. This is a very inexpensive way of contacting the people who were selected for the sample. However, the mail format is a drawback because of potentially low response rates: The questionnaire can easily be placed aside and forgotten among all the other tasks that people must attend to at home and work. Even if people start to fill out the questionnaire, something may happen to distract them, or they may become bored and simply throw the form in the trash. Some of the methods for increasing response rates are described later in this chapter. Another drawback is that no one is present to help if the person becomes confused or has a question about something.

Internet surveys It is very easy to design a questionnaire for administration on the Internet. Both open- and closed-ended questions can be written and presented to respondents. After the questionnaire is completed, the responses are immediately sent to the researcher. One of the first problems to consider is how to sample people. Most commonly, surveys are listed on search engines so people who are interested in a topic can discover that someone is interested in collecting data. Some of the major polling organizations are building a database of people interested in participating in surveys. Every time they conduct a survey, they select a sample from the database and send an e-mail invitation to participate. The Internet is also making it easier to obtain samples of people with particular characteristics. There are all sorts of Internet special interest groups for people with a particular illness or of a particular age, marital status, or occupational group. They use newsgroups, e-mail discussions, bulletin boards, and chat rooms to exchange ideas and information. Researchers can ask people who use these resources to volunteer for surveys. One concern about Internet data collection is whether the results will be at all similar to what might be found using traditional methods. Although research on this topic is not extensive, data indicate that Internet results are in fact comparable (Krantz, Ballard, & Scher, 1997; Stanton, 1998).

One problem with Internet data is that ultimately there is an ambiguity about the characteristics of the individuals providing information for the study. To meet ethical guidelines, the researcher will usually state that only persons 18 years of age or older are eligible; yet how is that controlled? People may also misrepresent their age, gender, or ethnicity. We simply do not know if this is a major problem. However, for most research topics it is unlikely that people will go to the trouble of misrepresenting themselves on the Internet to a greater extent than they would with any other method of collecting data.

Other technologies Researchers are taking advantage of new technologies to assist with the collection of data. An interesting application is seen in studies aimed at sampling people's behaviors and emotions over an extended

period of time. The usual approach would be to ask people to provide retrospective accounts of their behaviors or emotions (e.g., how often have you felt angry during the last week?). With pagers, cell phones, and other wireless communication devices, it is possible to contact people at various times and ask them to provide an immediate report of their current activities and emotional reactions. Feldman Barrett & Barrett (2001) refer to this as “computerized experience-sampling.” Responses might be provided on a paper questionnaire to be turned in later, or some other technology might be used such as a series of questions administered via Touch-Tone phone or a program running on a personal digital assistant.

Interviews

The fact that an interview involves an interaction between people has important implications. First, people are often more likely to agree to answer questions for a real person than to answer a mailed questionnaire. Good interviewers become quite skilled in convincing people to participate. Thus, response rates tend to be higher when interviews are used. The interviewer and respondent often establish a rapport that helps motivate the person to answer all the questions and complete the survey. People are more likely to leave questions unanswered on a written questionnaire than in an interview. An important advantage of an interview is that the interviewer can clarify any problems the person might have in understanding questions. Further, an interviewer can ask follow-up questions if needed to help clarify answers.

One potential problem in interviews is called **interviewer bias**. This term describes all of the biases that can arise from the fact that the interviewer is a unique human being interacting with another human. Thus, one potential problem is that the interviewer could subtly bias the respondent’s answers by inadvertently showing approval or disapproval of certain answers. Or, if there are several interviewers, each could possess different characteristics (e.g., level of physical attractiveness, age, or race) that might influence the way respondents answer. Another problem is that interviewers may have expectations that could lead them to “see what they are looking for” in the respondents’ answers. Such expectations could bias their interpretations of responses or lead them to probe further for an answer from certain respondents but not from others—for example, when questioning whites but not people from other groups or when testing boys but not girls. Careful screening and training of interviewers help to limit such biases.

We can now examine three methods of conducting interviews: face-to-face, telephone, and focus groups.

Face-to-face interviews Face-to-face interviews require that the interviewer and respondent meet to conduct the interview. Usually the interviewer travels to the person’s home or office, although sometimes the respondent goes to the interviewer’s office. Such interviews tend to be quite expensive and time-consuming. Therefore, they are most likely to be used when the sample size is fairly small and there are clear benefits to a face-to-face interaction.

Telephone interviews Almost all interviews for large-scale surveys are done via telephone. Telephone interviews are less expensive than face-to-face interviews, and they allow data to be collected relatively quickly because many interviewers can work on the same survey at once. Also, computerized telephone survey techniques lower the cost of telephone surveys by reducing labor and data analysis costs. With a computer-assisted telephone interview (CATI) system, the interviewer's questions are prompted on the computer screen, and the data are entered directly into the computer for analysis.

Focus group interviews A focus group is an interview with a group of about 6 to 10 individuals brought together for a period of usually 2–3 hours. Virtually any topic can be explored in a focus group. Often the group members are selected because they have a particular knowledge or interest in the topic. Because the focus group requires people to both spend time and incur some costs traveling to the focus group location, there is usually some sort of monetary or gift incentive to participate.

The questions tend to be open-ended and they are asked of the whole group. An advantage here is that group interaction is possible: People can respond to one another, and one comment can trigger a variety of responses. The interviewer must be skilled in working with groups both to facilitate communication and to deal with problems that may arise, such as one or two persons trying to dominate the discussion or hostility between group members. The group discussion is usually recorded and may be transcribed. The tapes and transcripts are then analyzed to find themes and areas of group consensus and disagreement. Sometimes the transcripts are analyzed with a computer program to search for certain words and phrases. Researchers usually prefer to conduct at least two or three discussion groups on a given topic to make sure that the information gathered is not unique to one group of people. However, because each focus group is time-consuming and costly and provides a great deal of information, researchers don't do very many such groups on any one topic.

SURVEY DESIGNS TO STUDY CHANGES OVER TIME

Surveys most frequently study people at one point in time. On many occasions, however, researchers wish to make comparisons over time. For example, my local newspaper hires a firm to conduct an annual random survey of county residents. Because the questions are the same each year, it is possible to track changes over time in such variables as satisfaction with the area, attitudes toward the school system, and perceived major problems facing the county. Similarly, a large number of new freshman students are surveyed each year at colleges throughout the United States to study changes in the composition, attitudes, and aspirations of this group (Astin, 1987). Often researchers will test hypotheses concerning how behavior may change over time. For example, Sebald (1986) compared surveys of teenagers in 1963, 1976, and 1982. The survey questions asked about whom the

teenagers seek advice from on a variety of issues. The primary finding was that seeking advice from peers rather than from parents increased from 1963 to 1976 but that this peer orientation decreased from 1976 to 1982.

Another way to study changes over time is to conduct a **panel study** in which the same people are surveyed at two or more points in time. In a “two-wave” panel study, people are surveyed at two points in time; in a “three-wave” panel study, there are three surveys; and so on. Panel studies are particularly important when the research question addresses the relationship between one variable at “time one” and another variable at some later “time two.” For example, Hill, Rubin, and Peplau (1976) surveyed dating couples to study variables such as attitude similarity. The same people were surveyed later to determine whether they were still in the dating relationship and, if so, how satisfied they were. The results showed that attitude similarity, measured at time one, is a predictor of how long the dating relationship will last.

SAMPLING FROM A POPULATION

Most research projects involve **sampling** participants from a population of interest. The **population** is composed of all individuals of interest to the researcher. One population of interest in a large public opinion poll, for instance, might be all eligible voters in the United States. This implies that the population of interest does not include people under the age of 18, convicted felons, visitors from other countries, and anyone else not eligible to vote. You might conduct a survey in which your population consists of all students at your college or university. With enough time and money, a survey researcher could conceivably contact everyone in the population. The United States attempts to do this every 10 years with an official census of the entire population. With a relatively small population, you might find it easy to study the entire population.

In most cases, however, studying the entire population would be a massive undertaking. Fortunately, it can be avoided by selecting a sample from the population of interest. With proper sampling, we can use information obtained from the participants (or “respondents”) who were sampled to precisely estimate characteristics of the population as a whole. Statistical theory allows us to infer what the population is like, based on data obtained from a sample (the logic underlying what is called *statistical significance* will be addressed in Chapter 13).

Confidence Intervals

When researchers make inferences about populations, they do so with a certain degree of confidence. Here is a statement that you might see when you read the results of a survey: “The results from the survey are accurate within 3 percentage points, using a 95% level of confidence.” What does this tell you? Suppose you asked students to tell you whether they prefer to study at home or at school, and the survey results indicate that 61% prefer to study at home. You now know that

the actual population value is probably between 58% and 64%. This is called a **confidence interval**—you can have 95% confidence that the true population value lies within this interval around the obtained sample result. Your best estimate of the population value is the sample value. However, because you have only a sample and not the entire population, your result may be in error. The confidence interval gives you information about the likely amount of the error. The formal term for this error is *sampling error* although you are probably more familiar with the term *margin of error*. Recall that the concept of measurement error was discussed in Chapter 5: When you measure a single individual on a variable, the obtained score may deviate from the true score because of measurement error. Similarly, when you study one sample, the obtained result may deviate from the true population value because of sampling error.

The surveys you often read about in newspapers and the previous example deal with percentages. What about questions that ask for more quantitative information? The logic in this instance is very much the same. For example, if you also ask students to report how many hours and minutes they studied during the previous day, you might find that the average amount of time was 76 minutes. A confidence interval could then be calculated based on the size of the sample; for example, the 95% confidence interval is 76 minutes plus or minus 10 minutes. It is highly likely that the true population value lies within the interval of 66 to 86 minutes. The topic of confidence intervals is discussed again in Chapter 13.

Sample Size

It is important to note that a larger sample size will reduce the size of the confidence interval. Although the size of the interval is determined by several factors, the most important is sample size. Larger samples are more likely to yield data that accurately reflect the true population value. This statement should make intuitive sense to you; a sample of 200 people from your school should yield more accurate data about your school than a sample of 25 people.

How large should the sample be? The sample size can be determined using a mathematical formula that takes into account the size of the confidence interval and the size of the population you are studying. Table 7.1 shows the sample size needed for a sample percentage to be accurate within plus or minus 3%, 5%, and 10%, given a 95% level of confidence. Note first that you need a larger sample size for increased accuracy. With a population size of 10,000, you need a sample of 370 for 5% accuracy; the needed sample size increases to 964 for 3% accuracy. It is also important to note that sample size is *not* a constant percentage of the population size. Many people believe that proper sampling requires a certain percentage of the population; these people often complain about survey results when they discover that a survey of an entire state was done with “only” 700 or 1,000 people. However, you can see in the table that the needed sample size does not change much even as the population size increases from 5,000 to 100,000 or more. As Fowler (1984) notes, “a sample of 150 people will describe a population of 1,500 or 15 million with virtually the same degree of accuracy . . .” (p. 41).

TABLE 7.1 Sample size and precision of population estimates (95% confidence level)

Size of population	Precision of Estimate		
	±3%	±5%	±10%
2,000	696	322	92
5,000	879	357	94
10,000	964	370	95
50,000	1,045	381	96
100,000	1,056	383	96
Over 100,000	1,067	384	96

Note: The sample sizes were calculated using conservative assumptions about the nature of the true population values.

SAMPLING TECHNIQUES

There are two basic techniques for sampling individuals from a population: probability sampling and nonprobability sampling. In **probability sampling**, each member of the population has a specifiable probability of being chosen. Probability sampling is very important when you want to make precise statements about a specific population on the basis of the results of your survey. In **nonprobability sampling**, we don't know the probability of any particular member of the population being chosen. Although this approach is not as sophisticated as probability sampling, we shall see that nonprobability sampling is quite common and useful in many circumstances.

Probability Sampling

Simple random sampling With **simple random sampling**, every member of the population has an equal probability of being selected for the sample. If the population has 1,000 members, each has one chance out of a thousand of being selected. Suppose you want to sample students who attend your school. A list of all students would be needed; from that list, students would be chosen at random to form the sample.

When conducting telephone interviews, researchers commonly have a computer randomly generate a list of telephone numbers with the dialing prefixes used for residences in the city or area being studied. This will produce a random sample of the population because most residences have telephones (if many people do not have phones, the sample would be biased). Some companies will even provide researchers with a list of telephone numbers for a survey in which the phone numbers of businesses and numbers that phone companies do not use

have been removed. You might note that this procedure results in a random sample of households rather than individuals. Survey researchers use other procedures when it is important to select one person at random from the household.

Stratified random sampling A somewhat more complicated procedure is **stratified random sampling**. The population is divided into subgroups (or strata), and random sampling techniques are then used to select sample members from each stratum. Any number of dimensions could be used to divide the population, but the dimension (or dimensions) chosen should be relevant to the problem under study. For instance, a survey of sexual attitudes might stratify on the basis of age, gender, and amount of education because these factors are related to sexual attitudes. Stratification on the basis of height or hair color would be ridiculous.

Stratified random sampling has the advantage of a built-in assurance that the sample will accurately reflect the numerical composition of the various subgroups. This kind of accuracy is particularly important when some subgroups represent very small percentages of the population. For instance, if African Americans make up 5% of a city of 100,000, a simple random sample of 100 people might not include any African Americans; a stratified random sample would include five African Americans chosen randomly from the population. In practice, when it is important to represent a small group within a population, researchers will “oversample” that group to ensure that a representative sample of the group is surveyed; a large enough sample must be obtained to be able to make inferences about the population. Thus, if your campus has a distribution of students similar to the city described here and you need to compare attitudes of African Americans and Whites, you will need to sample a large percentage of the African American students and only a small percentage of the White students to obtain a reasonable number of respondents from each group.

Cluster sampling It might have occurred to you that obtaining a list of all members of a population might be difficult. What if officials at your school decide that you cannot have access to a list of all students? What if you want to study a population that has no list of members, such as people who work in county health care agencies? In such situations, a technique called **cluster sampling** can be used. Rather than randomly sampling from a list of individuals, the researcher can identify “clusters” of individuals and then sample from these clusters. After the clusters are chosen, all individuals in each cluster are included in the sample. For example, you might conduct the survey of students using cluster sampling by identifying all classes being taught—the classes are the clusters of students. You could then randomly sample from this list of classes and have all members of the chosen classes complete your survey (making sure, of course, that no one completes the survey twice).

Most often, use of cluster sampling requires a series of samples from larger to smaller clusters—a “multistage” approach. For example, a researcher interested in studying county health care agencies might first randomly determine a

number of states to sample and then randomly sample counties from each state chosen. The researcher would then go to the health care agencies in each of these counties and study the people who work in them. Note that the main advantage of cluster sampling is that the researcher does not have to sample from lists of individuals to obtain a truly random sample of individuals.

Nonprobability Sampling

In contrast, nonprobability sampling techniques are quite arbitrary. A population may be defined, but little effort is expended to ensure that the sample accurately represents the population. However, among other things, nonprobability samples are cheap and convenient. Two types of nonprobability sampling are haphazard sampling and quota sampling.

Haphazard sampling One form of nonprobability sampling is **haphazard sampling** or “convenience” sampling. Haphazard sampling could be called a “take-them-where-you-find-them” method of obtaining participants. Thus, you would select a sample of students from your school in any way that is convenient. You might stand in front of the student union at 9 A.M., ask people who sit around you in your classes to participate, or visit a couple of fraternity and sorority houses. Unfortunately, such procedures are likely to introduce biases into the sample so that the sample may not be an accurate representation of the population of all students. Thus, if you selected your sample from students walking by the student union at 11 A.M., your sample excludes students who don’t frequent this location, and it may also eliminate afternoon and evening students. On my own campus, this sample would differ from the population of all students by being younger, working fewer hours, and being more likely to belong to a fraternity or sorority. Sample biases such as these limit your ability to use your sample data to estimate the actual population values. Your results may not generalize to your intended population but instead may describe only the biased sample that you obtained.

Quota sampling Another form of nonprobability sampling is **quota sampling**. A researcher who uses this technique chooses a sample that reflects the numerical composition of various subgroups in the population. Thus, quota sampling is similar to the stratified sampling procedure previously described; however, random sampling does not occur when you use quota sampling. To illustrate, suppose you want to ensure that your sample of students includes 19% freshmen, 23% sophomores, 26% juniors, 22% seniors, and 10% graduate students because these are the percentages of the classes in the total population. A quota sampling technique would make sure you have these percentages, but you would still collect your data using haphazard techniques. If you didn’t get enough graduate students in front of the student union, perhaps you could go to a graduate class to complete the sample. Although quota sampling is a bit more sophisticated than haphazard sampling, the problem re-

mains that no restrictions are placed on how individuals in the various subgroups are chosen. The sample does reflect the numerical composition of the whole population of interest, but respondents within each subgroup are selected in a haphazard manner.

EVALUATING SAMPLES

Samples should be representative of the population from which they are drawn. A completely unbiased sample is one that is highly representative of the population. How do you create a completely unbiased sample? First, you would randomly sample from a population that contains *all* individuals in the population. Second, you would contact and obtain completed responses from *all* individuals selected to be in the sample. Such standards are rarely achieved. Even if random sampling is used, there can be bias from two sources: the sampling frame used and poor response rates. Moreover, even though nonprobability samples have more potential sources of bias than probability samples, there are many reasons why they are used and should be evaluated positively.

Sampling Frame

The **sampling frame** is the *actual* population of individuals (or clusters) from which a random sample will be drawn. Rarely will this perfectly coincide with the population of interest—some biases will be introduced. If you define your population as “residents in my city,” the sampling frame may be a list of telephone numbers that you will use to contact residents between 5 P.M. and 9 P.M. This sampling frame excludes persons who do not have telephones or whose schedule prevents them from being at home when you are making calls. Also, if you are using the telephone directory to obtain numbers, you will exclude persons who have unlisted numbers. As another example, suppose you want to know what doctors think about the portrayal of the medical profession on television. A reasonable sampling frame would be all doctors listed in your telephone directory. Immediately you can see that you have limited your sample to a particular geographical area. More important, you have also limited the sample to doctors who have private practices—doctors who work in clinics and hospitals have been excluded. When evaluating the results of the survey, you need to consider how well the sampling frame matches the population of interest. Often the biases introduced are quite minor; however, they could be consequential.

Response Rate

The **response rate** in a survey is simply the percentage of people in the sample who actually completed the survey. Thus, if you mail 1,000 questionnaires to a random sample of adults in your community and 500 are completed and returned to you, the response rate is 50%. Response rate is important because it indicates how much bias there might be in the final sample of respondents.

Nonrespondents may differ from respondents in any number of ways including age, income, marital status, and education. The lower the response rate, the greater the likelihood that such biases may distort the findings and in turn limit the ability to generalize the findings to the population of interest.

In general, mail surveys have lower response rates than telephone surveys. With both methods, however, steps can be taken to maximize response rates. With mail surveys, an explanatory postcard or letter can be sent a week or so prior to mailing the survey. Follow-up reminders and even second mailings of the questionnaire are often effective in increasing response rates. It often helps to have a personally stamped return envelope rather than a business reply envelope. Even the look of the cover page of the questionnaire can be important (Dillman, 2000). With telephone surveys, respondents who aren't home can be called again and people who can't be interviewed today can be scheduled for a call at a more convenient time. Sometimes an incentive may be necessary to increase response rates. Such incentives can include cash, a gift, or a gift certificate for agreeing to participate. A crisp dollar bill "thank you" can be included with a mailed questionnaire. Another incentive is a chance to win a drawing for a prize. Finally, researchers should attempt to convince people that the survey's purposes are important and their participation will be a valuable contribution.

Reasons for Using Convenience Samples

Much of the research in psychology uses nonprobability sampling techniques to obtain participants for either surveys or experiments. The advantage of these techniques is that the investigator can obtain research participants without spending a great deal of money or time on selecting the sample. For example, it is common practice to select participants from students in introductory psychology classes. Often, these students are required to participate in studies being conducted by faculty and their students; the introductory psychology students can then choose which studies they wish to participate in.

Even in studies that do not use college students, the sample is often based on convenience rather than concern for obtaining a random sample. One of my colleagues studies children, but they are almost always from one particular elementary school. You can guess that this is because my colleague has established a good relationship with the teachers and administrators; thus, obtaining permission to conduct the research is fairly easy. Even though the sample is somewhat biased because it includes only children from one neighborhood that has certain social and economic characteristics, my colleague is not terribly concerned.

Why aren't researchers more worried about obtaining random samples for their research? The most important reason is that the research is being conducted to study relationships between variables rather than to accurately estimate population values. In the study of high school students cited previously (Steinberg & Dornbusch, 1991), the sample included only high school students

in California and Wisconsin. The sample data provide information on the average hours that these students work and their average use of alcohol. However, the researchers were not interested in giving a precise estimate of the number of hours that teenagers in the entire United States work; they were more interested in whether the number of hours that teenagers work is related to variables such as grade point average and alcohol use.

These issues will be explored further in Chapter 14. For now, it is also important to recognize that some nonprobability samples are more representative than others. The sample studied by Steinberg and Dornbusch appears to be highly representative of U.S. teenagers in general; even though the students came from only two states, they are from different geographical areas and several different high schools. Introductory psychology students are fairly representative of college students in general, and most college student samples are fairly representative of young adults. There aren't many obvious biases, particularly if you are studying basic psychological processes. Other samples might be much less representative of an intended population. Not long ago, a public affairs program on my local public television station asked viewers to dial a telephone number or send e-mail to vote for or against a gun control measure being considered by the legislature; the following evening, the program announced that almost 90% of the respondents opposed the measure. The sampling problems here are obvious: Groups opposed to gun control could immediately contact members to urge them to vote, and there were no limits on how many times someone could respond. In fact, the show received about 100 times more votes than it usually receives when it does such surveys. It is likely, then, that this sample was not at all representative of the population of the city or even viewers of the program.

You now have a great deal of information about methods for asking people about themselves. If you engage in this type of research, you will often need to design your own questions by following the guidelines described in this chapter and consulting sources such as Judd et al. (1991) and Converse and Presser (1986). However, you can also adapt questions and entire questionnaires that have been used in previous research. For example, Greenfield (1999) studied the new phenomenon of Internet addiction by adapting questions from a large body of existing research on addiction to gambling. Consider using previously developed questions, particularly if they have proven useful in other studies (make sure you don't violate any copyrights, however). A variety of measures of social, political, and occupational attitudes developed by others have been compiled by Robinson and his colleagues (Robinson, Athanasiou, & Head, 1969; Robinson, Rusk, & Head, 1968; Robinson, Shaver, & Wrightsman, 1991).

We noted in Chapter 4 that both nonexperimental and experimental research methods are necessary to fully understand behavior. The previous chapters have focused on nonexperimental approaches. In the next chapter, we begin a detailed description of experimental research design.

Study Terms

CATI	Population
Closed-ended questions	Probability sampling
Cluster sampling	Quota sampling
Confidence interval	Random sample
Face-to-face interview	Rating scale
Focus group	Response rate
Graphic rating scale	Response set
Group survey administration	Sampling
Haphazard (convenience) sampling	Sampling error
High-frequency scale	Sampling frame
Internet survey	Semantic differential scale
Interviewer bias	Simple random sampling
Mail survey	Stratified random sampling
Nonprobability sampling	Survey research
Open-ended questions	Telephone interview
Panel study	Yea-saying and nay-saying response set

Review Questions

1. What is a survey? Describe some research questions you might address with a survey.
2. Distinguish between probability and nonprobability sampling techniques. What are the implications of each?
3. Distinguish between haphazard and quota sampling.
4. Distinguish between simple random, stratified random, and cluster sampling.
5. Why don't researchers who want to test hypotheses about the relationships between variables worry a great deal about random sampling?
6. What are the advantages and disadvantages of using questionnaires versus interviews in a survey?
7. Distinguish between the different questionnaire, interview, and Internet survey administration methods.
8. What are some factors to take into consideration when constructing questions for surveys (including both questions and response alternatives)?
9. Define interviewer bias.

10. What is a social desirability response set?
11. How does sample size affect the interpretation of survey results?

Activity Questions

1. In the Steinberg and Dornbusch (1991) study on teenage employment (see Figure 7.1), longer work hours were associated with lower grade point averages. Can you conclude that working longer hours *causes* lower grades? Why or why not? How might you expand the scope of this investigation through a panel study?
2. Select a topic for a survey. Write at least five closed-ended questions that you might include in the survey. For each question, write one “good” version and one “poor” version. For each poor question, state what elements make it poor and why the good version is an improvement.
3. Suppose you want to know how many books in a bookstore have only male authors, only female authors, or both male and female authors (the “bookstore” in this case might be a large retail store, the textbook section of your college bookstore, or all the books in the stacks of your library). Because there are thousands of books in the store, you decide to study a sample of the books rather than examine every book there. Describe a possible sampling procedure using a nonprobability sampling technique. Then describe how you might sample books using a probability sampling technique. Now speculate on the ways that the outcomes of your research might differ using the two techniques.