CHAPTER 8

Producing Data: Sampling

Statistics, the science of data, provides ideas and tools that we can use in many settings. Sometimes we have data that describe a group of individuals and want to learn what the data say. That’s the job of exploratory data analysis. Sometimes we have specific questions but no data to answer them. To get sound answers, we must produce data in a way that is designed to answer our questions.

Suppose our question is “What percent of college students think that people should not obey laws that violate their personal values?” To answer the question, we interview undergraduate college students. We can’t afford to ask all students, so we put the question to a sample chosen to represent the entire student population. How shall we choose a sample that truly represents the opinions of the entire population? Statistical designs for choosing samples are the topic of this chapter.

Observation versus experiment

Our goal in choosing a sample is a picture of the population, disturbed as little as possible by the act of gathering information. Samples are one kind of observational study. In other settings, we gather data from an experiment. In doing an experiment, we don’t just observe individuals or ask them questions. We actively impose some treatment in order to observe the response. Experiments can answer questions such as “Does aspirin reduce the chance of a heart attack?” and “Do a
The rise and fall of hormone replacement

Should women take hormones such as estrogen after menopause, when natural production of these hormones ends? In 1992, several major medical organizations said “Yes.” In particular, women who took hormones seemed to reduce their risk of a heart attack by 35% to 50%. The risks of taking hormones appeared small compared with the benefits.

The evidence in favor of hormone replacement came from a number of observational studies that compared women who were taking hormones with others who were not. But women who choose to take hormones are very different from women who do not: they are richer and better educated and see doctors more often. These women do many things to maintain their health. It isn’t surprising that they have fewer heart attacks.

Experiments don’t let women decide what to do. They assign women to either hormone replacement or to dummy pills that look and taste the same as the hormone pills. The assignment is done by a coin toss, so that all kinds of women are equally likely to get either treatment. By 2002, several experiments with women of different ages agreed that hormone replacement does not reduce the risk of heart attacks. The National Institutes of Health, after reviewing the evidence, concluded that the observational studies were wrong. Taking hormones after menopause quickly fell out of favor.

When we simply observe women, the effects of actually taking hormones are confounded with (mixed up with) the characteristics of women who choose to take hormones.
Observation versus experiment

**CONFONDING**

Two variables (explanatory variables or lurking variables) are **confounded** when their effects on a response variable cannot be distinguished from each other.

Observational studies of the effect of one variable on another often fail because the explanatory variable is confounded with lurking variables. We will see that well-designed experiments take steps to defeat confounding.

**EXAMPLE 8.2 Wine, beer, or spirits?**

Moderate use of alcohol is associated with better health. Observational studies suggest that drinking wine rather than beer or spirits confers added health benefits. But people who prefer wine are different from those who drink mainly beer or stronger stuff. Wine drinkers as a group are richer and better educated. They eat more fruits and vegetables and less fried food. Their diets contain less fat, less cholesterol, and also less alcohol. They are less likely to smoke. The explanatory variable (What type of alcoholic beverage do you drink most often?) is confounded with many lurking variables (education, wealth, diet, and so on). A large study therefore concludes: “The apparent health benefits of wine compared with other alcoholic beverages, as described by others, may be a result of confounding by dietary habits and other lifestyle factors.” Figure 8.1 shows the confounding in picture form.

**FIGURE 8.1** Confounding: We can’t distinguish the effects of what people drink from the effects of their overall diet and lifestyle.
APPLY YOUR KNOWLEDGE

8.1 Cell phones and brain cancer. A study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same sex, age, and race who did not have brain cancer, then asked about use of cell phones. Result: “Our data suggest that use of handheld cellular telephones is not associated with risk of brain cancer.” Is this an observational study or an experiment? Why? What are the explanatory and response variables?

8.2 Teaching economics. An educational software company wants to compare the effectiveness of its computer animation for teaching about supply and demand curves with that of a textbook presentation. The company tests the economic knowledge of a number of first-year college students, then divides them into two groups. One group uses the animation, and the other studies the text. The company retests all the students and compares the increase in economic understanding in the two groups. Is this an experiment? Why or why not? What are the explanatory and response variables?

8.3 TV viewing and aggression. A typical hour of prime-time television shows three to five violent acts. Research shows that there is a clear association between time spent watching TV and aggressive behavior by adolescents. Nonetheless, it is hard to conclude that watching TV causes aggression. Suggest several lurking variables describing an adolescent’s home life that may be confounded with how much TV he or she watches.4

Sampling

A political scientist wants to know what percent of college-age adults consider themselves conservatives. An automaker hires a market research firm to learn what percent of adults aged 18 to 35 recall seeing television advertisements for a new gas-electric hybrid car. Government economists inquire about average household income. In all these cases, we want to gather information about a large group of individuals. Time, cost, and inconvenience forbid contacting every individual. So we gather information about only part of the group in order to draw conclusions about the whole.

POPULATION, SAMPLE, SAMPLING DESIGN

The population in a statistical study is the entire group of individuals about which we want information.

A sample is a part of the population from which we actually collect information. We use a sample to draw conclusions about the entire population.

A sampling design describes exactly how to choose a sample from the population.
Pay careful attention to the details of the definitions of “population” and “sample.” Look at Exercise 8.4 right now to check your understanding.

We often draw conclusions about a whole on the basis of a sample. Everyone has sipped a spoonful of soup and judged the entire bowl on the basis of that taste. But a bowl of soup is uniform, so that the taste of a single spoonful represents the whole. Choosing a representative sample from a large and varied population is not so easy. The first step in a proper sample survey is to say exactly what population we want to describe. The second step is to say exactly what we want to measure, that is, to give exact definitions of our variables. These preliminary steps can be complicated, as the following example illustrates.

**Example 8.3 The Current Population Survey**

The most important government sample survey in the United States is the monthly Current Population Survey (CPS). The CPS contacts about 60,000 households each month. It produces the monthly unemployment rate and much other economic and social information (see Figure 8.2). To measure unemployment, we must first specify the population we want to describe. Which age groups will we include? Will we include illegal aliens or people in prisons? The CPS defines its population as all U.S. residents (whether citizens or not) 16 years of age and over who are civilians and are not in an institution such as a prison. The unemployment rate announced in the news refers to this specific population.

The second question is harder: what does it mean to be “unemployed”? Someone who is not looking for work—for example, a full-time student—should not be called unemployed just because she is not working for pay. If you are chosen for the CPS sample, the interviewer first asks whether you are available to work and whether you actually looked for work in the past four weeks. If not, you are neither employed nor unemployed—you are not in the labor force.

If you are in the labor force, the interviewer goes on to ask about employment. If you did any work for pay or in your own business during the week of the survey, you...
are employed. If you worked at least 15 hours in a family business without pay, you are employed. You are also employed if you have a job but didn’t work because of vacation, being on strike, or other good reason. An unemployment rate of 4.7% means that 4.7% of the sample was unemployed, using the exact CPS definitions of both “labor force” and “unemployed.”

**APPLY YOUR KNOWLEDGE**

8.4 **Sampling students.** A political scientist wants to know how college students feel about the Social Security system. She obtains a list of the 3456 undergraduates at her college and mails a questionnaire to 250 students selected at random. Only 104 questionnaires are returned.

(a) What is the population in this study? Be careful: what group does she want information about?

(b) What is the sample? Be careful: from what group does she actually obtain information?

8.5 **The American Community Survey.** The American Community Survey (ACS) is replacing the “long form” sent to some households in the every-ten-years national census. Each month, the Census Bureau mails survey forms to 250,000 households. Telephone calls are made to households that don’t return the form. In the end, the Census Bureau gets responses from about 97% of the households it tries to contact. The survey asks questions about the people living in the household and about such things as plumbing, motor vehicles, and housing costs. What is the population for the ACS? What is the sample from which information is actually obtained?

8.6 **Customer satisfaction.** A department store mails a customer satisfaction survey to people who make credit card purchases at the store. This month, 45,000 people made credit card purchases. Surveys are mailed to 1000 of these people, chosen at random, and 137 people return the survey form. What is the population for this survey? What is the sample from which information was actually obtained?

**How to sample badly**

How can we choose a sample that we can trust to represent the population? A sampling design is a specific method for choosing a sample from the population. The easiest—but not the best—design just chooses individuals close at hand. If we are interested in finding out how many people have jobs, for example, we might go to a shopping mall and ask people passing by if they are employed. A sample selected by taking the members of the population that are easiest to reach is called a convenience sample. Convenience samples often produce unrepresentative data.

**EXAMPLE 8.4 Sampling at the mall**

A sample of mall shoppers is fast and cheap. But people at shopping malls tend to be more prosperous than typical Americans. They are also more likely to be teenagers or retired. Moreover, unless interviewers are carefully trained, they tend to question well-dressed, respectable people and avoid poorly dressed or tough-looking individuals. In short, mall interviews will not contact a sample that is representative of the entire population.
Interviews at shopping malls will almost surely overrepresent middle-class and retired people and underrepresent the poor. This will happen almost every time we take such a sample. That is, it is a systematic error caused by a bad sampling design, not just bad luck on one sample. This is bias: the outcomes of mall surveys will repeatedly miss the truth about the population in the same ways.

**BIAS**

The design of a statistical study is **biased** if it systematically favors certain outcomes.

**EXAMPLE 8.5  Online polls**

The American Family Association (AFA) is a conservative group that claims to stand for "traditional family values." It regularly posts online poll questions on its Web site—just click on a response to take part. Because the respondents are people who visit this site, the poll results always support AFA's positions. Well, almost always. In 2004, AFA's online poll asked about the heated issue of allowing same-sex marriage. Soon, email lists and social network sites favored mostly by young liberals pointed to the AFA poll. Almost 850,000 people responded, and 60% of them favored legalizing same-sex marriage. AFA claimed that homosexual rights groups had skewed its poll.

Online polls are now everywhere—some sites will even provide help in conducting your own online poll. As the AFA poll illustrates, you can't trust the results. **People who take the trouble to respond to an open invitation are usually not representative of any clearly defined population.** That's true of regular visitors to AFA's site, of the activists who made a special effort to vote in the marriage poll, and of the people who bother to respond to write-in, call-in, or online polls in general. Polls like these are examples of **voluntary response sampling.**

**VOLUNTARY RESPONSE SAMPLE**

A **voluntary response sample** consists of people who choose themselves by responding to a broad appeal. Voluntary response samples are biased because people with strong opinions are most likely to respond.

**APPLY YOUR KNOWLEDGE**

8.7  **Sampling on campus.**  You see a woman student standing in front of the student center, now and then stopping other students to ask them questions. She says that she is collecting student opinions for a class assignment. Explain why this sampling method is almost certainly biased.

8.8  **More sampling on campus.**  Your college wants to gather student opinion about parking for students on campus. It isn't practical to contact all students.

(a)  Give an example of a way to choose a sample of students that is poor practice because it depends on voluntary response.
(b) Give another example of a bad way to choose a sample that doesn’t use voluntary response.

**Simple random samples**

In a voluntary response sample, people choose whether to respond. In a convenience sample, the interviewer makes the choice. In both cases, personal choice produces bias. The statistician’s remedy is to allow impersonal chance to choose the sample. A sample chosen by chance allows neither favoritism by the sampler nor self-selection by respondents. Choosing a sample by chance attacks bias by giving all individuals an equal chance to be chosen. Rich and poor, young and old, black and white, all have the same chance to be in the sample.

The simplest way to use chance to select a sample is to place names in a hat (the population) and draw out a handful (the sample). This is the idea of simple random sampling.

**SIMPLE RANDOM SAMPLE**

A simple random sample (SRS) of size \( n \) consists of \( n \) individuals from the population chosen in such a way that every set of \( n \) individuals has an equal chance to be the sample actually selected.

An SRS not only gives each individual an equal chance to be chosen but also gives every possible sample an equal chance to be chosen. There are other random sampling designs that give each individual, but not each sample, an equal chance. Exercise 8.44 describes one such design.

When you think of an SRS, picture drawing names from a hat to remind yourself that an SRS doesn’t favor any part of the population. That’s why an SRS is a better method of choosing samples than convenience or voluntary response sampling. But writing names on slips of paper and drawing them from a hat is slow and inconvenient. That’s especially true if, like the Current Population Survey, we must draw a sample of size 60,000. In practice, samplers use software. The Simple Random Sample applet makes the choosing of an SRS very fast. If you don’t use the applet or other software, you can randomize by using a table of random digits.

**RANDOM DIGITS**

A table of random digits is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these two properties:

1. Each entry in the table is equally likely to be any of the 10 digits 0 through 9.
2. The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.
Table B at the back of the book is a table of random digits. Table B begins with the digits 19223950340575628713. To make the table easier to read, the digits appear in groups of five and in numbered rows. The groups and rows have no meaning—the table is just a long list of randomly chosen digits. There are two steps in using the table to choose a simple random sample.

**USING TABLE B TO CHOOSE AN SRS**

**Step 1. Label.** Give each member of the population a numerical label of the same length.

**Step 2. Table.** To choose an SRS, read from Table B successive groups of digits of the length you used as labels. Your sample contains the individuals whose labels you find in the table.

You can label up to 100 items with two digits: 01, 02, ..., 99, 00. Up to 1000 items can be labeled with three digits, and so on. Always use the shortest labels that will cover your population. As standard practice, we recommend that you begin with label 1 (or 01 or 001, as needed). Reading groups of digits from the table gives all individuals the same chance to be chosen because all labels of the same length have the same chance to be found in the table. For example, any pair of digits in the table is equally likely to be any of the 100 possible labels 01, 02, ..., 99, 00. Ignore any group of digits that was not used as a label or that duplicates a label already in the sample. You can read digits from Table B in any order—across a row, down a column, and so on—because the table has no order. As standard practice, we recommend reading across rows.

**EXAMPLE 8.6 Sampling spring break resorts**

A campus newspaper plans a major article on spring break destinations. The authors intend to call four randomly chosen resorts at each destination to ask about their attitudes toward groups of students as guests. Here are the resorts listed in one city:

<table>
<thead>
<tr>
<th></th>
<th>Aloha Kai</th>
<th>Captiva</th>
<th>Palm Tree</th>
<th>Sea Shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>02</td>
<td>Anchor Down</td>
<td>Casa del Mar</td>
<td>Radisson</td>
<td>Silver Beach</td>
</tr>
<tr>
<td>03</td>
<td>Banana Bay</td>
<td>Coconuts</td>
<td>Ramada</td>
<td>Sunset Beach</td>
</tr>
<tr>
<td>04</td>
<td>Banyan Tree</td>
<td>Diplomat</td>
<td>Sandpiper</td>
<td>Tradewinds</td>
</tr>
<tr>
<td>05</td>
<td>Beach Castle</td>
<td>Holiday Inn</td>
<td>Sea Castle</td>
<td>Tropical Breeze</td>
</tr>
<tr>
<td>06</td>
<td>Best Western</td>
<td>Lime Tree</td>
<td>Sea Club</td>
<td>Tropical Shores</td>
</tr>
<tr>
<td>07</td>
<td>Cabana</td>
<td>Outrigger</td>
<td>Sea Grape</td>
<td>Veranda</td>
</tr>
</tbody>
</table>

**Step 1. Label.** Because two digits are needed to label the 28 resorts, all labels will have two digits. We have added labels 01 to 28 in the list of resorts. Always say how you labeled the members of the population. To sample from the 1240 resorts in a major vacation area, you would label the resorts 0001, 0002, ..., 1239, 1240.
Step 2. Table. To use the Simple Random Sample applet, just enter 28 in the “Population =” box and 4 in the “Select a sample” box, click “Reset,” and click “Sample.” Figure 8.3 shows the result of one sample.

To use Table B, read two-digit groups until you have chosen four resorts. Starting at line 130 (any line will do), we find

69051 64817 87174 09517 84534 06489 87201 97245

Because the labels are two digits long, read successive two-digit groups from the table. Ignore groups not used as labels, like the initial 69. Also ignore any repeated labels, like the second and third 17s in this row, because you can’t choose the same resort twice. Your sample contains the resorts labeled 05, 16, 17, and 20. These are Beach Castle, Radisson, Ramada, and Sea Club.

We can trust results from an SRS, because it uses impersonal chance to avoid bias. Online polls and mall interviews also produce samples. We can’t trust results from these samples, because they are chosen in ways that invite bias. The first question to ask about any sample is whether it was chosen at random.

**EXAMPLE 8.7 Do you avoid soda?**

A Gallup Poll on the American diet asked subjects about their attitudes toward various foods. The press release mentioned “the increasing proportion of Americans who say they try to avoid ‘soda or pop’ (51%, up from 41% in 2002).” Can we trust that 51%?
Ask first how Gallup selected its sample. Later in the press release we read this: “These results are based on telephone interviews with a randomly selected national sample of 1,005 adults, aged 18 and older, conducted July 8–11, 2004.”

This is a good start toward gaining our confidence. Gallup tells us what population it has in mind (people at least 18 years old who live anywhere in the United States). We know that the sample from this population was of size 1005 and, most important, that it was chosen at random. There is more to say, but we have at least heard the comforting words “randomly selected.”

### APPLY YOUR KNOWLEDGE

#### 8.9 Apartment living.
You are planning a report on apartment living in a college town. You decide to select three apartment complexes at random for in-depth interviews with residents. Use the Simple Random Sample applet, other software, or Table B to select a simple random sample of three of the following apartment complexes. If you use Table B, start at line 117.

<table>
<thead>
<tr>
<th>Ashley Oaks</th>
<th>Country View</th>
<th>Mayfair Village</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bay Pointe</td>
<td>Country Villa</td>
<td>Nobb Hill</td>
</tr>
<tr>
<td>Beau Jardin</td>
<td>Crestview</td>
<td>Pemberly Courts</td>
</tr>
<tr>
<td>Bluffs</td>
<td>Del-Lynn</td>
<td>Peppermill</td>
</tr>
<tr>
<td>Brandon Place</td>
<td>Fairington</td>
<td>Pheasant Run</td>
</tr>
<tr>
<td>Briarwood</td>
<td>Fairway Knolls</td>
<td>River Walk</td>
</tr>
<tr>
<td>Brownstone</td>
<td>Fowler</td>
<td>Sagamore Ridge</td>
</tr>
<tr>
<td>Burberry Place</td>
<td>Franklin Park</td>
<td>Salem Courthouse</td>
</tr>
<tr>
<td>Cambridge</td>
<td>Georgetown</td>
<td>Village Square</td>
</tr>
<tr>
<td>Chauncey Village</td>
<td>Greenacres</td>
<td>Waterford Court</td>
</tr>
<tr>
<td>Country Squire</td>
<td>Lahr House</td>
<td>Williamsburg</td>
</tr>
</tbody>
</table>

#### 8.10 Minority managers.
A firm wants to understand the attitudes of its minority managers toward its system for assessing management performance. Below is a list of all the firm’s managers who are members of minority groups. Use the Simple Random Sample applet, other software, or Table B at line 139 to choose six to be interviewed in detail about the performance appraisal system.

<table>
<thead>
<tr>
<th>Abdulhamid</th>
<th>Duncan</th>
<th>Huang</th>
<th>Puri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarwal</td>
<td>Fernandez</td>
<td>Kim</td>
<td>Richards</td>
</tr>
<tr>
<td>Baxter</td>
<td>Fleming</td>
<td>Liao</td>
<td>Rodriguez</td>
</tr>
<tr>
<td>Bonds</td>
<td>Gates</td>
<td>Mourning</td>
<td>Santiago</td>
</tr>
<tr>
<td>Brown</td>
<td>Goel</td>
<td>Naber</td>
<td>Shen</td>
</tr>
<tr>
<td>Castillo</td>
<td>Gomez</td>
<td>Peters</td>
<td>Vega</td>
</tr>
<tr>
<td>Cross</td>
<td>Hernandez</td>
<td>Pliego</td>
<td>Wang</td>
</tr>
</tbody>
</table>

#### 8.11 Sampling the forest.
To gather data on a 1200-acre pine forest in Louisiana, the U.S. Forest Service laid a grid of 1410 equally spaced circular plots over a map of the forest. A ground survey visited a sample of 10% of these plots.

(a) How would you label the plots?
(b) Use Table B, beginning at line 105, to choose the first 5 plots in an SRS of 141 plots.
Other sampling designs

The general framework for statistical sampling is a probability sample.

**PROBABILITY SAMPLE**

A probability sample is a sample chosen by chance. We must know what samples are possible and what chance, or probability, each possible sample has.

Some probability sampling designs (such as an SRS) give each member of the population an equal chance to be selected. This may not be true in more elaborate sampling designs. In every case, however, the use of chance to select the sample is the essential principle of statistical sampling.

Designs for sampling from large populations spread out over a wide area are usually more complex than an SRS. For example, it is common to sample important groups within the population separately, then combine these samples. This is the idea of a stratified random sample.

**STRATIFIED RANDOM SAMPLE**

To select a stratified random sample, first classify the population into groups of similar individuals, called strata. Then choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Choose the strata based on facts known before the sample is taken. For example, a population of election districts might be divided into urban, suburban, and rural strata. A stratified design can produce more precise information than an SRS of the same size by taking advantage of the fact that individuals in the same stratum are similar to one another.

**EXAMPLE 8.8 Seat belt use in Hawaii**

Each state conducts an annual survey of seat belt use by drivers, following guidelines set by the federal government. The guidelines require probability samples. Seat belt use is observed at randomly chosen road locations at random times during daylight hours. The locations are not an SRS of all locations in the state but rather a stratified sample using the state’s counties as strata.

In Hawaii, the counties are the islands that make up the state’s territory. The seat belt survey sample consists of 135 road locations in the four most populated islands: 66 in Oahu, 24 in Maui, 23 in Hawaii, and 22 in Kauai. The sample sizes on the islands are proportional to the amount of road traffic.²

Seat belt surveys in larger states often use multistage samples. Counties are grouped into strata by population size. At the first stage, choose a stratified sample...
of counties that includes all of the most populated counties and a selection of smaller counties. The second stage selects locations at random within each county chosen at the first stage. These are also stratified samples, with locations grouped into strata by, for example, high, medium and low traffic volume.

Most large-scale sample surveys use multistage samples. The samples at individual stages may be SRSs but are often stratified. Analysis of data from sampling designs more complex than an SRS takes us beyond basic statistics. But the SRS is the building block of more elaborate designs, and analysis of other designs differs more in complexity of detail than in fundamental concepts.

**A P P L Y  Y O U R  K N O W L E D G E**

8.12 **A stratified sample.** A club has 30 student members and 10 faculty members.

The students are

- Abel
- Fisher
- Huber
- Miranda
- Reinmann
- Carson
- Ghosh
- Jimenez
- Moskowitz
- Santos
- Chen
- Griswold
- Jones
- Neyman
- Shaw
- David
- Hein
- Kim
- O’Brien
- Thompson
- Deming
- Hernandez
- Klotz
- Pearl
- Utts
- Elashoff
- Holland
- Liu
- Potter
- Varga

The faculty members are

- Andrews
- Fernandez
- Kim
- Moore
- West
- Besicovitch
- Gupta
- Lightman
- Vicario
- Yang

The club can send 4 students and 2 faculty members to a convention. It decides to choose those who will go by random selection. Use software or Table B to choose a stratified random sample of 4 students and 2 faculty members.

8.13 **Sampling by accountants.** Accountants use stratified samples during audits to verify a company’s records of such things as accounts receivable. The stratification is based on the dollar amount of the item and often includes 100% sampling of the largest items. One company reports 5000 accounts receivable. Of these, 100 are in amounts over $50,000; 500 are in amounts between $1000 and $50,000; and the remaining 4400 are in amounts under $1000. Using these groups as strata, you decide to verify all of the largest accounts and to sample 5% of the midsize accounts and 1% of the small accounts. How would you label the two strata from which you will sample? Use software or Table B, starting at line 115, to select only the first 5 accounts from each of these strata.

**Cautions about sample surveys**

Random selection eliminates bias in the choice of a sample from a list of the population. When the population consists of human beings, however, accurate information from a sample requires more than a good sampling design.
To begin, we need an accurate and complete list of the population. Because such a list is rarely available, most samples suffer from some degree of undercoverage. A sample survey of households, for example, will miss not only homeless people but prison inmates and students in dormitories. An opinion poll conducted by telephone will miss the 5% of American households without residential phones. The results of national sample surveys therefore have some bias if the people not covered—who most often are poor people—differ from the rest of the population.

A more serious source of bias in most sample surveys is nonresponse, which occurs when a selected individual cannot be contacted or refuses to cooperate. Nonresponse to sample surveys often reaches 50% or more, even with careful planning and several callbacks. Because nonresponse is higher in urban areas, most sample surveys substitute other people in the same area to avoid favoring rural areas in the final sample. If the people contacted differ from those who are rarely at home or who refuse to answer questions, some bias remains.

### UNDERCOVERAGE AND NONRESPONSE

- **Undercoverage** occurs when some groups in the population are left out of the process of choosing the sample.
- **Nonresponse** occurs when an individual chosen for the sample can’t be contacted or refuses to participate.

### EXAMPLE 8.9  How bad is nonresponse?

The Current Population Survey has the lowest nonresponse rate of any poll we know: only about 6% or 7% of the households in the sample don’t respond. People are more likely to respond to a government survey, and the CPS contacts its sample in person before doing later interviews by phone.

The University of Chicago’s General Social Survey (GSS) is the nation’s most important social science survey. (See Figure 8.4.) The GSS also contacts its sample in person, and it is run by a university. Despite these advantages, its most recent survey had a 30% rate of nonresponse.

What about opinion polls by news media and opinion-polling firms? We don’t know their rates of nonresponse because they won’t say. That itself is a bad sign. The Pew Research Center for the People and the Press imitated a careful telephone survey and published the results: out of 2879 households called, 1658 were never at home, refused, or would not finish the interview. That’s a nonresponse rate of 58%.

In addition, the behavior of the respondent or of the interviewer can cause response bias in sample results. People know that they should take the trouble to vote, for example, so many who didn’t vote in the last election will tell an interviewer that they did. The race or sex of the interviewer can influence responses to questions about race relations or attitudes toward feminism. Answers to questions that ask respondents to recall past events are often inaccurate because of faulty memory. For example, many people “telescope” events in the past, bringing them...
Cautions about sample surveys

The National Opinion Research Center
Pick List
Extract
Analyze
Homepage
Site Map

Introduction
⇒ About GSSDirs
⇒ GSS News
⇒ Credits
Codebook Indexes
⇒ Mnemonic
⇒ Sequential
⇒ Subject
⇒ Collections
⇒ GSS Publications
⇒ Questionnaires

Subject Index: D
Previous
Pick Page
Analyzing
Help

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Date Of Birth
Date Of Interview
Death penalty, see Capital Punishment
Deaths, see Homicide, Suicide, Trauma
Degrees, see Education
Democrats, see Political
Demonstrations
Dictionary Of Occupational Titles, see D.O.T. Codes
Disarmament
Dole, Bob, see Political
D.O.T. Codes
Draft, see Military
Drinking
Drug Use And Addiction

Figure 8.4 A small part of the subject index of the General Social Survey. The GSS has tracked opinions about a wide variety of issues since 1972.

forward in memory to more recent time periods. “Have you visited a dentist in the last 6 months?” will often draw a “Yes” from someone who last visited a dentist 8 months ago. Careful training of interviewers and careful supervision to avoid variation among the interviewers can reduce response bias. Good interviewing technique is another aspect of a well-done sample survey.

The wording of questions is the most important influence on the answers given to a sample survey. Confusing or leading questions can introduce strong bias, and even minor changes in wording can change a survey’s outcome. Here are some examples.

Example 8.10 Help the poor?

How do Americans feel about government help for the poor? Only 13% think we are spending too much on “assistance to the poor,” but 44% think we are spending too much on “welfare.”

Example 8.11 Independence for Scotland?

How do the Scots feel about the movement to become independent from England? Well, 51% would vote for “independence for Scotland,” but only 34% support “an independent Scotland separate from the United Kingdom.”

It seems that “assistance to the poor” and “independence” are nice, hopeful words. “Welfare” and “separate” are negative words. You can’t trust the results of a
sample survey until you have read the exact questions asked. The amount of nonresponse and the date of the survey are also important. Good statistical design is a part, but only a part, of a trustworthy survey.

**APPLY YOUR KNOWLEDGE**

8.14 Ring-no-answer. A common form of nonresponse in telephone surveys is “ring-no-answer.” That is, a call is made to an active number but no one answers. The Italian National Statistical Institute looked at nonresponse to a government survey of households in Italy during the periods January 1 to Easter and July 1 to August 31. All calls were made between 7 and 10 p.m., but 21.4% gave “ring-no-answer” in one period versus 41.5% “ring-no-answer” in the other period. Which period do you think had the higher rate of no answers? Why? Explain why a high rate of nonresponse makes sample results less reliable.

8.15 Question wording. In 2000, when the federal budget showed a large surplus, the Pew Research Center asked two questions of random samples of adults. Both questions stated that Social Security would be “fixed.” Here are the uses suggested for the remaining surplus:

- Should the money be used for a tax cut, or should it be used to fund new government programs?
- Should the money be used for a tax cut, or should it be spent on programs for education, the environment, health care, crime-fighting and military defense?

One of these questions drew 60% favoring a tax cut. The other drew only 22%. Which wording pulls respondents toward a tax cut? Why?

**Inference about the population**

Despite the many practical difficulties in carrying out a sample survey, using chance to choose a sample does eliminate bias in the actual selection of the sample from the list of available individuals. But it is unlikely that results from a sample are exactly the same as for the entire population. Sample results, like the official unemployment rate obtained from the monthly Current Population Survey, are only estimates of the truth about the population. If we select two samples at random from the same population, we will draw different individuals. So the sample results will almost certainly differ somewhat. Properly designed samples avoid systematic bias, but their results are rarely exactly correct and they vary from sample to sample.

How accurate is a sample result like the monthly unemployment rate? We can’t say for sure, because the result would be different if we took another sample. But the results of random sampling don’t change haphazardly from sample to sample. Because we deliberately use chance, the results obey the laws of probability that govern chance behavior. We can say how large an error we are likely to make in drawing conclusions about the population from a sample. Results from a sample survey usually come with a margin of error that sets bounds on the size of the likely
error. How to do this is part of the business of statistical inference. We will describe the reasoning in Chapter 14.

One point is worth making now: larger random samples give more accurate results than smaller samples. By taking a very large sample, you can be confident that the sample result is very close to the truth about the population. The Current Population Survey contacts about 60,000 households, so it estimates the national unemployment rate very accurately. Opinion polls that contact 1000 or 1500 people give less accurate results. Of course, only probability samples carry this guarantee. The AFA’s voluntary response sample on same-sex marriage is worthless even though 850,000 people clicked a response. Using a probability sampling design and taking care to deal with practical difficulties reduce bias in a sample. The size of the sample then determines how close to the population truth the sample result is likely to fall.

APPLY YOUR KNOWLEDGE

8.16 Ask more people. Just before a presidential election, a national opinion-polling firm increases the size of its weekly sample from the usual 1500 people to 4000 people. Why do you think the firm does this?

CHAPTER 8 SUMMARY

We can produce data intended to answer specific questions by observational studies or experiments. Sample surveys that select a part of a population of interest to represent the whole are one type of observational study. Experiments, unlike observational studies, actively impose some treatment on the subjects of the experiment.

Observational studies often fail to show that changes in an explanatory variable actually cause changes in a response variable, because the explanatory variable is confounded with lurking variables. Variables are confounded when their effects on a response can’t be distinguished from each other.

A sample survey selects a sample from the population of all individuals about which we desire information. We base conclusions about the population on data from the sample.

The design of a sample describes the method used to select the sample from the population. Probability sampling designs use chance to select a sample.

The basic probability sample is a simple random sample (SRS). An SRS gives every possible sample of a given size the same chance to be chosen.

Choose an SRS by labeling the members of the population and using a table of random digits to select the sample. Software can automate this process.

To choose a stratified random sample, classify the population into strata, groups of individuals that are similar in some way that is important to the response. Then choose a separate SRS from each stratum.
Failure to use probability sampling often results in bias, or systematic errors in the way the sample represents the population. Voluntary response samples, in which the respondents choose themselves, are particularly prone to large bias. In human populations, even probability samples can suffer from bias due to undercoverage or nonresponse, from response bias, or from misleading results due to poorly worded questions. Sample surveys must deal expertly with these potential problems in addition to using a probability sampling design.

CHECK YOUR SKILLS

8.17 The Nurses’ Health Study has interviewed a sample of more than 100,000 female registered nurses every two years since 1976. The study finds that “light-to-moderate drinkers had a significantly lower risk of death” than either nondrinkers or heavy drinkers. The Nurses’ Health Study is
(a) an observational study.
(b) an experiment.
(c) Can’t tell without more information.

8.18 How strong is the evidence from the Nurses’ Health Study (see the previous exercise) that moderate drinking lowers the risk of death?
(a) Quite strong because it comes from an experiment.
(b) Quite strong because it comes from a large random sample.
(c) Weak, because drinking habits are confounded with many other variables.

8.19 An opinion poll contacts 1161 adults and asks them, “Which political party do you think has better ideas for leading the country in the twenty-first century?” In all, 696 of the 1161 say, “The Democrats.” The sample in this setting is
(a) all 225 million adults in the United States.
(b) the 1161 people interviewed.
(c) the 696 people who chose the Democrats.

8.20 A committee on community relations in a college town plans to survey local businesses about the importance of students as customers. From telephone book listings, the committee chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the committee. The population for this study is
(a) all businesses in the college town.
(b) the 150 businesses chosen.
(c) the 73 businesses that returned the questionnaire.

8.21 The sample in the setting of the previous exercise is
(a) all businesses in the college town.
(b) the 150 businesses chosen.
(c) the 73 businesses that returned the questionnaire.

8.22 You can find the Excite Poll online at poll.excite.com. You simply click on a response to become part of the sample. The poll question for June 19, 2005, was “Do you prefer watching first-run movies at a movie theater, or waiting until they
are available on home video or pay-per-view?” In all, 8896 people responded, with only 13% (1118 people) saying they preferred theaters. You can conclude that
(a) American adults strongly prefer watching movies at home.
(b) the poll uses voluntary response, so the results tell us little about the population of all adults.
(c) the sample is too small to draw any conclusion.

8.23 You must choose an SRS of 10 of the 440 retail outlets in New York that sell your company’s products. How would you label this population in order to use Table B?
(a) 001, 002, 003, …, 439, 440
(b) 000, 001, 002, …, 439, 440
(c) 1, 2, …, 439, 440

8.24 You are using the table of random digits to choose a simple random sample of 6 students from a class of 30 students. You label the students 01 to 30 in alphabetical order. Go to line 133 of Table B. Your sample contains the students labeled
(a) 45, 74, 04, 18, 07, 65.
(b) 04, 18, 07, 13, 02, 07.
(c) 04, 18, 07, 13, 02, 05.

8.25 You want to choose an SRS of 5 of the 7200 salaried employees of a corporation. You label the employees 0001 to 7200 in alphabetical order. Using line 111 of Table B, your sample contains the employees labeled
(a) 6694, 5130, 0041, 2712, 3827.
(b) 6694, 0513, 0929, 7004, 1271.
(c) 8148, 6694, 8760, 5130, 9297.

8.26 A sample of households in a community is selected at random from the telephone directory. In this community, 4% of households have no telephone and another 35% have unlisted telephone numbers. The sample will certainly suffer from
(a) nonresponse.
(b) undercoverage.
(c) false responses.

CHAPTER 8 EXERCISES

In all exercises asking for an SRS, you may use Table B, the Simple Random Sample applet, or other software.

8.27 Alcohol and heart attacks. Many studies have found that people who drink alcohol in moderation have lower risk of heart attacks than either nondrinkers or heavy drinkers. Does alcohol consumption also improve survival after a heart attack? One study followed 1913 people who were hospitalized after severe heart attacks. In the year before their heart attacks, 47% of these people did not drink, 36% drank moderately, and 17% drank heavily. After four years, fewer of the moderate drinkers had died. Is this an observational study or an experiment? Why? What are the explanatory and response variables?
8.28 Reducing nonresponse. How can we reduce the rate of refusals in telephone surveys? Most people who answer at all listen to the interviewer’s introductory remarks and then decide whether to continue. One study made telephone calls to randomly selected households to ask opinions about the next election. In some calls, the interviewer gave her name, in others she identified the university she was representing, and in still others she identified both herself and the university. The study recorded what percent of each group of interviews was completed. Is this an observational study or an experiment? Why? What are the explanatory and response variables?

8.29 Safety of anesthetics. The National Halothane Study was a major investigation of the safety of anesthetics used in surgery. Records of over 850,000 operations performed in 34 major hospitals showed the following death rates for four common anesthetics:

<table>
<thead>
<tr>
<th>Anesthetic</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death rate</td>
<td>1.7%</td>
<td>1.7%</td>
<td>3.4%</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

There is a clear association between the anesthetic used and the death rate of patients. Anesthetic C appears dangerous.

(a) Explain why we call the National Halothane Study an observational study rather than an experiment, even though it compared the results of using different anesthetics in actual surgery.

(b) When the study looked at other variables that are confounded with a doctor’s choice of anesthetic, it found that Anesthetic C was not causing extra deaths. Suggest important lurking variables that are confounded with what anesthetic a patient receives.

8.30 Movie viewing. An opinion poll calls 2000 randomly chosen residential telephone numbers, then asks to speak with an adult member of the household. The interviewer asks, “How many movies have you watched in a movie theater in the past 12 months?”

(a) What population do you think the poll has in mind?

(b) In all, 1131 people respond. What is the rate (percent) of nonresponse?

(c) What source of response error is likely for the question asked?

8.31 The United States in world affairs. A Gallup Poll asked, “Do you think the U.S. should take the leading role in world affairs, take a major role but not the leading role, take a minor role, or take no role at all in world affairs?” Gallup’s report said, “Results are based on telephone interviews with 1,002 national adults, aged 18 and older, conducted Feb. 9–12, 2004.”

(a) What is the population for this sample survey? What was the sample?

(b) Gallup notes that the order of the four possible responses was rotated when the question was read over the phone. Why was this done?

8.32 Same-sex marriage. Example 8.5 reports an online poll in which 60% of the respondents favored making same-sex marriage legal. National random samples taken at the same time showed 55% to 60% of the respondents opposed to legalizing same-sex marriage. (The results varied a bit depending on the exact
question asked.) Explain briefly to someone who knows no statistics why the random samples report public opinion more reliably than the online poll.

8.33 **Ann Landers takes a sample.** Advice columnist Ann Landers once asked her female readers whether they would be content with affectionate treatment by men, with no sex ever. Over 90,000 women wrote in, with 72% answering “Yes.” Many of the letters described unfeeling treatment at the hands of men. Explain why this sample is certainly biased. What is the likely direction of the bias? That is, is 72% probably higher or lower than the truth about the population of all adult women?

8.34 **Seat belt use.** A study in El Paso, Texas, looked at seat belt use by drivers. Drivers were observed at randomly chosen convenience stores. After they left their cars, they were invited to answer questions that included questions about seat belt use. In all, 75% said they always used seat belts, yet only 61.5% were wearing seat belts when they pulled into the store parking lots. Explain the reason for the bias observed in responses to the survey. Do you expect bias in the same direction in most surveys about seat belt use?

8.35 **Do you trust the Internet?** You want to ask a sample of college students the question “How much do you trust information about health that you find on the Internet—a great deal, somewhat, not much, or not at all?” You try out this and other questions on a pilot group of 10 students chosen from your class. The class members are

<table>
<thead>
<tr>
<th>Anderson</th>
<th>Deng</th>
<th>Glaus</th>
<th>Nguyen</th>
<th>Samuels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arroyo</td>
<td>De Ramos</td>
<td>Helling</td>
<td>Palmiero</td>
<td>Shen</td>
</tr>
<tr>
<td>Batista</td>
<td>Drasin</td>
<td>Husain</td>
<td>Percival</td>
<td>Tse</td>
</tr>
<tr>
<td>Bell</td>
<td>Eckstein</td>
<td>Johnson</td>
<td>Prince</td>
<td>Velasco</td>
</tr>
<tr>
<td>Burke</td>
<td>Fernandez</td>
<td>Kim</td>
<td>Puri</td>
<td>Wallace</td>
</tr>
<tr>
<td>Cabrera</td>
<td>Fullmer</td>
<td>Molina</td>
<td>Richards</td>
<td>Washburn</td>
</tr>
<tr>
<td>Calloway</td>
<td>Gandhi</td>
<td>Morgan</td>
<td>Rider</td>
<td>Zabidi</td>
</tr>
<tr>
<td>Delluci</td>
<td>Garcia</td>
<td>Murphy</td>
<td>Rodriguez</td>
<td>Zhao</td>
</tr>
</tbody>
</table>

Choose an SRS of 10 students. If you use Table B, start at line 117.

8.36 **Telephone area codes.** There are approximately 371 active telephone area codes covering Canada, the United States, and some Caribbean areas. (More are created regularly.) You want to choose an SRS of 25 of these area codes for a study of available telephone numbers. Label the codes 001 to 371 and use the Simple Random Sample applet or other software to choose your sample. (If you use Table B, start at line 129 and choose only the first 5 codes in the sample.)

8.37 **Nonresponse.** Academic sample surveys, unlike commercial polls, often discuss nonresponse. A survey of drivers began by randomly sampling all listed residential telephone numbers in the United States. Of 45,956 calls to these numbers, 5029 were completed. What was the rate of nonresponse for this sample? (Only one call was made to each number. Nonresponse would be lower if more calls were made.)

8.38 **Running red lights.** The sample described in the previous exercise produced a list of 5024 licensed drivers. The investigators then chose an SRS of 880 of these drivers to answer questions about their driving habits.
(a) How would you assign labels to the 5024 drivers? Use Table B, starting at line 104, to choose the first 5 drivers in the sample.

(b) One question asked was “Recalling the last ten traffic lights you drove through, how many of them were red when you entered the intersections?” Of the 880 respondents, 171 admitted that at least one light had been red. A practical problem with this survey is that people may not give truthful answers. What is the likely direction of the bias: do you think more or fewer than 171 of the 880 respondents really ran a red light? Why?

8.39 **Sampling at a party.** At a party there are 30 students over age 21 and 20 students under age 21. You choose at random 3 of those over 21 and separately choose at random 2 of those under 21 to interview about attitudes toward alcohol. You have given every student at the party the same chance to be interviewed: what is that chance? Why is your sample not an SRS?

8.40 **Random digits.** In using Table B repeatedly to choose random samples, you should not always begin at the same place, such as line 101. Why not?

8.41 **Random digits.** Which of the following statements are true of a table of random digits, and which are false? Briefly explain your answers.

(a) There are exactly four 0s in each row of 40 digits.
(b) Each pair of digits has chance 1/100 of being 00.
(c) The digits 0000 can never appear as a group, because this pattern is not random.

8.42 **Sampling at a party.** At a large block party there are 290 men and 110 women. You want to ask opinions about how to improve the next party. To be sure that women’s opinions are adequately represented, you decide to choose a stratified random sample of 20 men and 20 women. Explain how you will assign labels to the names of the people at the party. Give the labels of the first 3 men and the first 3 women in your sample. If you use Table B, start at line 130.

8.43 **Sampling Amazon forests.** Stratified samples are widely used to study large areas of forest. Based on satellite images, a forest area in the Amazon basin is divided into 14 types. Foresters studied the four most commercially valuable types: alluvial climax forests of quality levels 1, 2, and 3, and mature secondary forest. They divided the area of each type into large parcels, chose parcels of each type at random, and counted tree species in a 20- by 25-meter rectangle randomly placed within each parcel selected. Here is some detail:

<table>
<thead>
<tr>
<th>Forest type</th>
<th>Total parcels</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Climax 1</td>
<td>36</td>
<td>4</td>
</tr>
<tr>
<td>Climax 2</td>
<td>72</td>
<td>7</td>
</tr>
<tr>
<td>Climax 3</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>Secondary</td>
<td>42</td>
<td>4</td>
</tr>
</tbody>
</table>

Choose the stratified sample of 18 parcels. Be sure to explain how you assigned labels to parcels. If you use Table B, start at line 162.

8.44 **Systematic random samples.** Systematic random samples are often used to choose a sample of apartments in a large building or dwelling units in a block at the last stage of a multistage sample. An example will illustrate the idea of a systematic
sample. Suppose that we must choose 4 addresses out of 100. Because \( \frac{100}{4} = 25 \), we can think of the list as four lists of 25 addresses. Choose 1 of the first 25 at random, using Table B. The sample contains this address and the addresses 25, 50, and 75 places down the list from it. If 13 is chosen, for example, then the systematic random sample consists of the addresses numbered 13, 38, 63, and 88.

(a) Use Table B to choose a systematic random sample of 5 addresses from a list of 200. Enter the table at line 120.

(b) Like an SRS, a systematic sample gives all individuals the same chance to be chosen. Explain why this is true, then explain carefully why a systematic sample is nonetheless not an SRS.

8.45 Random digit dialing. The list of individuals from which a sample is actually selected is called the sampling frame. Ideally, the frame should list every individual in the population, but in practice this is often difficult. A frame that leaves out part of the population is a common source of undercoverage.

(a) Suppose that a sample of households in a community is selected at random from the telephone directory. What households are omitted from this frame? What types of people do you think are likely to live in these households? These people will probably be underrepresented in the sample.

(b) It is usual in telephone surveys to use random digit dialing equipment that selects the last four digits of a telephone number at random after being given the exchange (the first three digits). Which of the households you mentioned in your answer to (a) will be included in the sampling frame by random digit dialing?

8.46 Wording survey questions. Comment on each of the following as a potential sample survey question. Is the question clear? Is it slanted toward a desired response?

(a) “Some cell phone users have developed brain cancer. Should all cell phones come with a warning label explaining the danger of using cell phones?”

(b) “Do you agree that a national system of health insurance should be favored because it would provide health insurance for everyone and would reduce administrative costs?”

(c) “In view of the negative externalities in parent labor force participation and pediatric evidence associating increased group size with morbidity of children in day care, do you support government subsidies for day care programs?”

8.47 Regulating guns. The National Gun Policy Survey asked respondents’ opinions about government regulation of firearms. A report from the survey says, “Participating households were identified through random digit dialing; the respondent in each household was selected by the most-recent-birthday method.”

(a) What is “random digit dialing?” Why is it a practical method for obtaining (almost) an SRS of households?

(b) The survey wants the opinion of an individual adult. Several adults may live in a household. In that case, the survey interviewed the adult with the most recent birthday. Why is this preferable to simply interviewing the person who answers the phone?

8.48 Your own bad questions. Write your own examples of bad sample survey questions.
(a) Write a biased question designed to get one answer rather than another.
(b) Write a question that is confusing, so that it is hard to answer.

8.49 Canada’s national health care. The Ministry of Health in the Canadian province of Ontario wants to know whether the national health care system is achieving its goals in the province. Much information about health care comes from patient records, but that source doesn’t allow us to compare people who use health services with those who don’t. So the Ministry of Health conducted the Ontario Health Survey, which interviewed a random sample of 61,239 people who live in Ontario.18

(a) What is the population for this sample survey? What is the sample?
(b) The survey found that 76% of males and 86% of females in the sample had visited a general practitioner at least once in the past year. Do you think these estimates are close to the truth about the entire population? Why?

8.50 Polling Hispanics. A New York Times News Service article on a poll concerned with the opinions of Hispanics includes this paragraph:

The poll was conducted by telephone from July 13 to 27, with 3,092 adults nationwide, 1,074 of whom described themselves as Hispanic. It has a margin of sampling error of plus or minus three percentage points for the entire poll and plus or minus four percentage points for Hispanics. Sample sizes for most Hispanic nationalities, like Cubans or Dominicans, were too small to break out the results separately.19

(a) Why is the “margin of sampling error” larger for Hispanics than for all 3092 respondents?
(b) Why would a very small sample size prevent a responsible news organization from breaking out results for Cubans separately?