JOHNER IMAGES/ALAMY

CHAPTER 3

Producing Data

Introduction

Reliable data are needed to make business decisions. Here are some examples where carefully collected data are essential.

- How does General Motors decide the numbers of vehicles of different colors that it will produce?
- How will Whole Foods choose a location for a new store?
- How does Monsanto decide how much it is willing to spend for a Super Bowl commercial?

In Chapters 1 and 2 we learned some basic tools of *data analysis.* We used graphs and numbers to describe data. When we do **exploratory data analysis,** we rely heavily on plotting the data. We look for patterns that suggest interesting conclusions or questions for further study. However, *exploratory analysis alone can rarely provide convincing evidence for its conclusions because striking patterns we find in data can arise from many sources.*

The validity of the conclusions that we draw from an analysis of data depends not only on the use of the best methods to perform the analysis but also on the quality of the data. Therefore, Section 3.1 begins this chapter with a short overview on sources of data. The two main sources for quality data are designed samples and designed experiments. We study these two sources in Sections 3.2 and 3.3, respectively.

Should an experiment or sample survey that could possibly provide interesting and important information always be performed? How can we safeguard the privacy of subjects in a sample survey? What constitutes the mistreatment of people or animals who are studied in an experiment? These are questions of **ethics.** In Section 3.4, we address ethical issues related to the design of studies and the analysis of data.

CHAPTER OUTLINE

exploratory data analysis

CAUTION

ethics

3.1 Sources of Data

There are many sources of data. Some data are very easy to collect, but they may not be very useful. Other data require careful planning and need professional staff to gather. These can be much more useful. Whatever the source, a good statistical analysis will start with a careful study of the source of the data. Here is one type of source.

Anecdotal data

It is tempting to simply draw conclusions from our own experience, making no use of more broadly representative data. An advertisement for a Pilates class says that men need this form of exercise even more than women. The ad describes the benefits that two men received from taking Pilates classes. A newspaper ad states that a particular brand of windows is "considered to be the best" and says that "now is the best time to replace your windows and doors." These types of stories, or *anecdotes,* sometimes provide quantitative data. However, this type of data does not give us a sound basis for drawing conclusions.

Anecdotal Evidence

Anecdotal evidence is based on haphazardly selected cases, which often come to our attention because they are striking in some way. These cases need not be representative of any larger group of cases.

APPLY YOUR KNOWLEDGE

3.1 Is this good market research? You and your friends are big fans of *True Detective,* an HBO police drama. To what extent do you think you can generalize your preference for this show to all students at your college?

3.2 Should you invest in stocks? You have just accepted a new job and are offered several options for your retirement account. One of these invests about 75% of your employer's contribution in stocks. You talk to a friend who joined the company several years ago who said that after he chose that option, the value of the stocks decreased substantially. He strongly recommended that you choose a different option. Comment on the value of your friend's advice.

3.3 Preference for a brand. Samantha is a serious runner. She and all her friends prefer drinking Gatorade Endurance to Heed prior to their long runs. Explain why Samantha's experience is not good evidence that most young people prefer Gatorade Endurance to Heed.

3.4 Reliability of a product. A friend has driven a Toyota Camry for more than 200,000 miles with only the usual service maintenance expenses. Explain why not all Camry owners can expect this kind of performance.

Available data

available data

Occasionally, data are collected for a particular purpose but can also serve as the basis for drawing sound conclusions about other research questions. We use the term **available data** for this type of data.

Available Data

Available data are data that were produced in the past for some other purpose but that may help answer a present question.

The library and the Internet can be good sources of available data. Because producing new data is expensive, we all use available data whenever possible. Here are two examples.

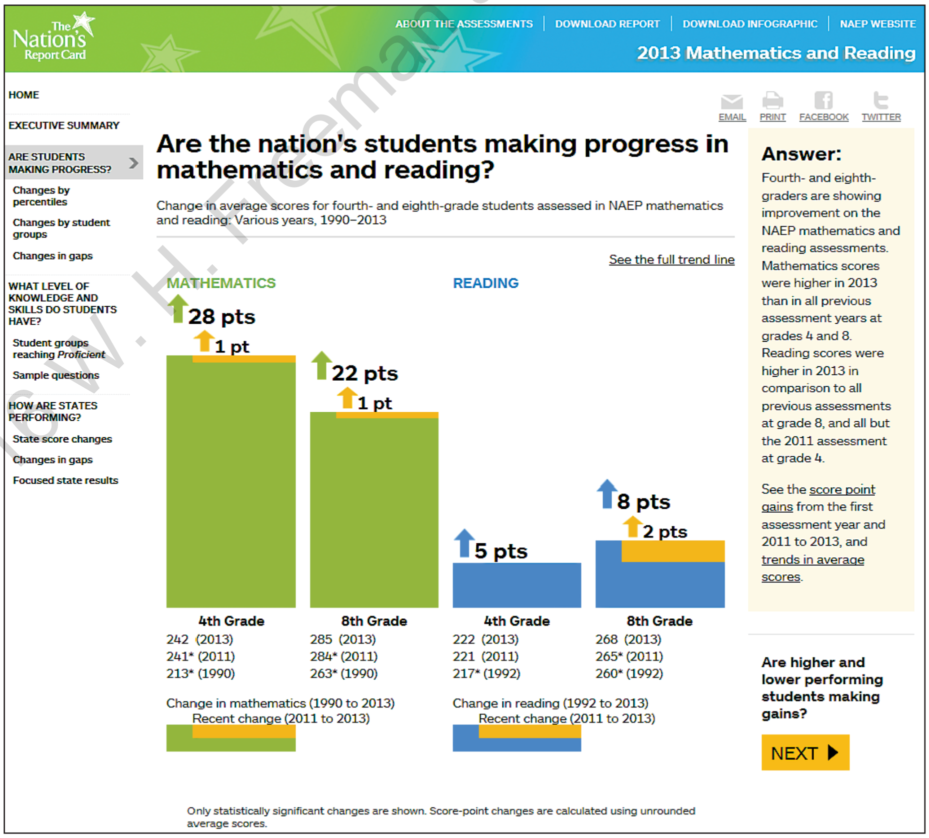
EXAMPLE 3.1 International Manufacturing Productivity

If you visit the U.S. Bureau of Labor Statistics website, **bls.gov**, you can find many interesting sets of data and statistical summaries. One recent study compared the average hourly manufacturing compensation costs of 34 countries. The study showed that Norway and Switzerland had the top two costs.[1]

EXAMPLE 3.2 Can Our Workforce Compete in a Global Economy?

In preparation to compete in the global economy, students need to improve their mathematics.[2] At the website of the National Center for Education Statistics, **nces.ed.gov/nationsreportcard**, you will find full details about the math skills of schoolchildren in the latest National Assessment of Educational Progress. Figure 3.1 shows one of the pages that reports on the increases in mathematics and reading scores.[3]

FIGURE 3.1 The websites of government statistical offices are prime sources of data. Here is a page from the National Assessment of Educational Progress, Example 3.2.



Source: The Nation's Report Card

Many nations have a single national statistical office, such as Statistics Canada (**statcan.gc.ca**) and Mexico's INEGI (**inegi.org.mx/default.aspx**). More than 70 different U.S. agencies collect data. You can reach most of them through the government's FedStats site (**fedstats.gov**).

APPLY YOUR KNOWLEDGE

3.5 Check out the Bureau of Labor Statistics website. Visit the Bureau of Labor Statistics website, **bls.gov**. Find a set of data that interests you. Explain how the data were collected and what questions the study was designed to answer.

Although available data can be very useful for many situations, we often find that clear answers to important questions require that data be produced to answer those specific questions. Are your customers likely to buy a product from a competitor if you raise your price? Is the expected return from a proposed advertising campaign sufficient to justify the cost? The validity of our conclusions from the analysis of data collected to address these issues rests on a foundation of carefully collected data. In this chapter, we learn how to produce trustworthy data and to judge the quality of data produced by others. The techniques for producing data that we study require no formulas, but they are among the most important ideas in statistics. Statistical designs for producing data rely on either *sampling* or *experiments.*

Sample surveys and experiments

How have the attitudes of Americans, on issues ranging from shopping online to satisfaction with work, changed over time? Sample surveys are the usual tool for answering questions like these. A **sample survey** collects data from a sample of cases that represent some larger population of cases.

sample survey

EXAMPLE 3.3 Confidence in Banks and Companies

One of the most important sample surveys is the General Social Survey (GSS) conducted by the NORC, a national organization for research and computing affiliated with the University of Chicago.[4] The GSS interviews about 3000 adult residents of the United States every second year. The survey includes questions about how much confidence people have in banks and companies.

sample

population

The GSS selects a **sample** of adults to represent the larger population of all English-speaking adults living in the United States. The idea of *sampling* is to study a part in order to gain information about the whole. Data are often produced by sampling a **population** of people or things. Opinion polls, for example, report the views of the entire country based on interviews with a sample of about 1000 people. Government reports on employment and unemployment are produced from a monthly sample of about 60,000 households. The quality of manufactured items is monitored by inspecting small samples each hour or each shift.

JOHN GRESS/REUTERS/CORBIS

APPLY YOUR KNOWLEDGE

3.6 Are Millennials loyal customers? A website claims that Millennial generation consumers are very loyal to the brands that they prefer. What additional information do you need to evaluate this claim?

census

In all our examples, the expense of examining every item in the population makes sampling a practical necessity. Timeliness is another reason for preferring a sample to a **census,** which is an attempt to contact every case in the entire population. We want information on current unemployment and public opinion next week, not next year. Moreover, a carefully conducted sample is often more accurate than a census. Accountants, for example, sample a firm's inventory to verify the accuracy of the records. Counting every item in a warehouse can be expensive and also inaccurate. Bored people might not count carefully.

If conclusions based on a sample are to be valid for the entire population, a sound design for selecting the sample is required. Sampling designs are the topic of Section 3.2.

A sample survey collects information about a population by selecting and measuring a sample from the population. The goal is a picture of the population, disturbed as little as possible by the act of gathering information. Sample surveys are one kind of *observational study.*

Observation versus Experiment

In an **observational study,** we observe cases and measure variables of interest but do not attempt to influence the responses.

In an **experiment,** we deliberately impose some treatment on cases and observe their responses.

APPLY YOUR KNOWLEDGE

3.7 Market share for energy drinks. A website reports that Red Bull is the top energy drink brand with sales of $2.9 billion in 2014.[5] Do you think that this report is based on an observational study or an experiment? Explain your answer.

3.8 An advertising agency chooses an ESPN television ad. An advertising agency developed two versions of an ad that will be shown during a major sporting event on EPSN but must choose only one to air. The agency recruited 100 college students and divided them into two groups of 50. Each group viewed one of the versions of the ad and then answered a collection of questions about their reactions to the ad. Is the advertising agency using an observational study or an experiment to help make its decision? Give reasons for your answer.

intervention
treatment
experiment

An observational study, even one based on a statistical sample, is a poor way to determine what will happen if we change something. The best way to see the effects of a change is to do an **intervention**—where we actually impose the change. The change imposed is called a **treatment.** When our goal is to understand cause and effect, experiments are the only source of fully convincing data. In an **experiment,** a treatment is imposed and the responses are recorded. Experiments usually require some sort of randomization.

We begin the discussion of statistical designs for data collection in Section 3.2 with the principles underlying the design of samples. We then move to the design of experiments in Section 3.3.

SECTION 3.1 Summary

- **Anecdotal data** come from stories or reports about cases that do not necessarily represent a larger group of cases.
- **Available data** are data that were produced for some other purpose but that may help answer a question of interest.
- A **sample survey** collects data from a sample of cases that represent some larger population of cases.
- A **census** collects data from all cases in the population of interest.
- In an **experiment,** a **treatment** is imposed and the responses are recorded.

SECTION 3.1 Exercises

For Exercises 3.1 to 3.4, see page 124; for 3.5, see page 126; for 3.6, see page 126; and for 3.7 and 3.8, see page 127.

In several of the following exercises, you are asked to identify the type of data that is described. Possible answers include anecdotal data, available data, observational data that are from sample surveys, observational data that are not from sample surveys, and experiments. It is possible for some data to be classified in more than one category.

3.9 A dissatisfied customer. You like to eat tuna sandwiches. Recently you noticed that there does not seem to be as much tuna as you expected when you opened the can. Identify the type of data that this represents, and describe how it can or cannot be used to reach a conclusion about the amount of tuna in the cans.

3.10 Claims settled for $3,300,000! According to a story in *Consumer Reports,* three major producers of canned tuna agreed to pay $3,300,000 to settle claims in California that the amount of tuna in their cans was less than the amount printed on the label of the cans.[6] What kind of data do you think was used in this situation to convince the producers to pay this amount of money to settle the claims? Explain your answer fully.

3.11 Marketing milk. An advertising campaign was developed to promote the consumption of milk by adolescents. Part of the campaign was based on a study conducted to determine the effect of additional milk in the diet of adolescents over a period of 18 months. A control group received no extra milk. Growth rates of total body bone mineral content (TBBMC) over the study period were calculated for each subject. Data for the control group were used to examine the relationship between growth rate of TBBMC and age.

(a) How would you classify the data used to evaluate the effect of the additional milk in the diet? Explain your answer.

(b) How would you classify the control group data on growth rate of TBBMC and age for the study of this relationship? Explain your answer.

(c) Can you classify the growth rate of TBBMC and age variables as explanatory or response? If so, which is the explanatory variable? Give reasons for your answer.

3.12 Satisfaction with allocation of concert tickets. Your college sponsored a concert that sold out.

(a) After the concert, an article in the student newspaper reported interviews with three students who were unable to get tickets and were very upset with that fact. What kind of data does this represent? Explain your answer.

(b) A week later, the student organization that sponsored the concert set up a website where students could rank their satisfaction with the way that the tickets were allocated using a 5-point scale with values "very satisfied," "satisfied," "neither satisfied nor unsatisfied," "dissatisfied," and "very dissatisfied." The website was open to any students who chose to provide their opinion. How would you classify these data? Give reasons for your answer.

(c) Suppose that the website in part (b) was changed so that only a sample of students from the college were invited by text message to respond, and those who did not respond within three days were sent an additional text message reminding them to respond. How would your answer to part (b) change, if at all?

(d) Write a short summary contrasting different types of data using your answers to parts (a), (b), and (c) of this exercise.

3.13 Gender and consumer choices. Men and women differ in their choices for many product categories. Are there gender differences in preferences for health insurance plans as well? A market researcher interviews a large sample of consumers, both men and women. She asks each consumer which of two health plans he or she prefers. Is this study an experiment? Why or why not? What are the explanatory and response variables?

3.14 Is the product effective? An educational software company wants to compare the effectiveness of its computer animation for teaching about supply, demand, and market clearing with that of a textbook presentation. The company tests the economic knowledge of 50 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?

3.15 Does job training work? A state institutes a job-training program for manufacturing workers who lose their jobs. After five years, the state reviews how well the program works. Critics claim that because the state's unemployment rate for manufacturing workers was 6% when the program began and 10% five years later, the program is ineffective. Explain why higher unemployment does not necessarily mean that the training program failed. In particular, identify some lurking variables (see page 118 in Chapter 2) whose effect on unemployment may be confounded with the effect of the training program.

3.16 Are there treatments? Refer to Exercises 3.9 through 3.15. For any of these that involve an experiment, describe the treatment that is used.

3.2 Designing Samples

Samsung and O2 want to know how much time smartphone users spend on their smartphones. An automaker hires a market research firm to learn what percent of adults aged 18 to 35 recall seeing television advertisements for a new sport utility vehicle. Government economists inquire about average household income. In all these cases, we want to gather information about a large group of people. We will not, as in an experiment, impose a treatment in order to observe the response. Also, time, cost, and inconvenience forbid contacting every person. In such cases, we gather information about only part of the group—a *sample*—in order to draw conclusions about the whole. **Sample surveys** are an important kind of observational study.

sample survey

Population and Sample

The entire group of cases that we want to study is called the **population.**

A **sample** is a subset of the population for which we collect data.

Notice that "population" is defined in terms of our desire for knowledge. If we wish to draw conclusions about all U.S. college students, that group is our population—even if only local students are available for questioning. The sample is the part from which we draw conclusions about the whole. The **design** of a sample survey refers to the method used to choose the sample from the population.

sample design

EXAMPLE 3.4 Can We Compete Globally?

A lack of reading skills has been cited as one factor that limits our ability to compete in the global economy.[7] Various efforts have been made to improve this situation. One of these is the Reading Recovery (RR) program. RR has specially trained teachers work one-on-one with at-risk first-grade students to help them learn to read. A study was designed to examine the relationship between the RR teachers' beliefs about their ability to motivate students and the progress of the students whom they teach.[8] The National Data Evaluation Center (NDEC) website (**ndec.us**) says that there are 6112 RR teachers. The researchers send a questionnaire to a random sample of 200 of these. The population consists of all 6112 RR teachers, and the sample is the 200 that were randomly selected.

Unfortunately, our idealized framework of population and sample does not exactly correspond to the situations that we face in many cases. In Example 3.4, the list of teachers was prepared at a particular time in the past. It is very likely that some of the teachers on the list are no longer working as RR teachers today. New teachers have been trained in RR methods and are not on the list. A list of items to be sampled is often called a **sampling frame.** For our example, we view this list as the population. We may have out-of-date addresses for some who are still working as RR teachers, and some teachers may choose not to respond to our survey questions.

sampling frame

In reporting the results of a sample survey, it is important to include all details regarding the procedures used. The proportion of the original sample who actually provide usable data is called the **response rate** and should be reported for all surveys. If only 150 of the teachers who were sent questionnaires provided usable data, the response rate would be 150/200, or 75%. Follow-up mailings or phone calls to those who do not initially respond can help increase the response rate.

response rate

APPLY YOUR KNOWLEDGE

3.17 Taxes and forestland usage. A study was designed to assess the impact of taxes on forestland usage in part of the Upper Wabash River Watershed in Indiana.[9] A survey was sent to 772 forest owners from this region, and 348 were returned. Consider the population, the sample, and the response rate for this study. Describe these based on the information given, and indicate any additional information that you would need to give a complete answer.

3.18 Job satisfaction. A research team wanted to examine the relationship between employee participation in decision making and job satisfaction in a company. They are planning to randomly select 300 employees from a list of 2500 employees in the company. The Job Descriptive Index (JDI) will be used to measure job satisfaction, and the Conway Adaptation of the Alutto-Belasco Decisional Participation Scale will be used to measure decision participation. Describe the population and the sample for this study. Can you determine the response rate? Explain your answer.

Poor sample designs can produce misleading conclusions. Here is an example.

EXAMPLE 3.5 Sampling Product in a Steel Mill

A mill produces large coils of thin steel for use in manufacturing home appliances. The quality engineer wants to submit a sample of 5-centimeter squares to detailed laboratory examination. She asks a technician to cut a sample of 10 such squares. Wanting to provide "good" pieces of steel, the technician carefully avoids the visible defects in the coil material when cutting the sample. The laboratory results are wonderful, but the customers complain about the material they are receiving.

In Example 3.5, the samples were selected in a manner that guaranteed that they would not be representative of the entire population. This sampling scheme displays *bias,* or systematic error, in favoring some parts of the population over others. Online opinion polls are particularly vulnerable to bias because the sample who respond are not representative of the population at large. Online polls use *voluntary response samples,* a particularly common form of biased sample.

Voluntary Response Sample

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

The remedy for bias in choosing a sample is to allow impersonal chance to do the choosing so that there is neither favoritism by the sampler nor voluntary response.

Random selection of a sample eliminates bias by giving all cases an equal chance to be chosen.

Voluntary response is one common type of bad sample design. Another is **convenience sampling,** which chooses the cases easiest to reach. Here is an example of convenience sampling.

convenience sampling

EXAMPLE 3.6 Interviewing Customers at the Mall

Manufacturers and advertising agencies often use interviews at shopping malls to gather information about the habits of consumers and the effectiveness of ads. A sample of mall customers is fast and cheap. But people contacted at shopping malls are not representative of the entire U.S. population. They are richer, for example, and more likely to be teenagers or retired. Moreover, mall interviewers tend to select neat, safe-looking subjects from the stream of customers. Decisions based on mall interviews may not reflect the preferences of all consumers.

Both voluntary response samples and convenience samples produce samples that are almost guaranteed not to represent the entire population. These sampling methods display *bias* in favoring some parts of the population over others.

Bias

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

Big data involves extracting useful information from large and complex data sets. There are exciting developments in this field and opportunities for new uses of data are widespread. Some have suggested that there are potential biases in the results obtained from some big data sets.[10] Here is an example:

TIMOTHY LENNEY

EXAMPLE 3.7 Bias and Big Data

A study used Twitter and Foursquare data on coffee, food, nightlife, and shopping activity to describe the disruptive effects of Hurricane Sandy.[11] However, the data are dominated by tweets and smartphone activity from Manhattan. Relatively little data are from areas such as Breezy Point, where the effects of the hurricane were most severe.

APPLY YOUR KNOWLEDGE

3.19 What is the population? For each of the following sampling situations, identify the population as exactly as possible. That is, indicate what kind of cases the population consists of and exactly which cases fall in the population. If the information given is not sufficient, complete the description of the population in a reasonable way.

(a) Each week, the Gallup Poll questions a sample of about 1500 adult U.S. residents to determine national opinion on a wide variety of issues.

(b) The 2000 census tried to gather basic information from every household in the United States. Also, a "long form" requesting additional information was sent to a sample of about 17% of households.

(c) A machinery manufacturer purchases voltage regulators from a supplier. There are reports that variation in the output voltage of the regulators is affecting the performance of the finished products. To assess the quality of the supplier's production, the manufacturer sends a sample of five regulators from the last shipment to a laboratory for study.

3.20 Market segmentation and movie ratings. You wonder if that new "blockbuster" movie is really any good. Some of your friends like the movie, but you decide to check the Internet Movie Database (**imdb.com**) to see others' ratings. You find that 2497 people chose to rate this movie, with an average rating of only 3.7 out of 10. You are surprised that most of your friends liked the movie, while many people gave low ratings to the movie online. Are you convinced that a majority of those who saw the movie would give it a low rating? What type of sample are your friends? What type of sample are the raters on the Internet Movie Database? Discuss this example in terms of market segmentation (see, for example, **businessplans.org/Segment.html.**)

Simple random samples

The simplest sampling design amounts to placing names in a hat (the population) and drawing out a handful (the sample). This is *simple random sampling.*

Simple Random Sample

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

We select an SRS by labeling all the cases in the population and using software or a table of random digits to select a sample of the desired size. Notice that an SRS not only gives each case an equal chance to be chosen (thus avoiding bias in the choice), but gives every possible sample an equal chance to be chosen. There are other random sampling designs that give each case, but not each sample, an equal chance. One such design, *systematic random sampling*, is described later in Exercise 3.36 (pages 141–142).

Thinking about random digits helps you to understand randomization even if you will use software in practice. Table B at the back of the book is a table of random digits.

Random Digits

A **table of random digits** is a list of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 that has the following properties:

1. The digit in any position in the list has the same chance of being any one of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

2. The digits in different positions are independent in the sense that the value of one has no influence on the value of any other.

You can think of Table B as the result of asking an assistant (or a computer) to mix the digits 0 to 9 in a hat, draw one, then replace the digit drawn, mix again, draw a second digit, and so on. The assistant's mixing and drawing saves us the work of mixing and drawing when we need to randomize. 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 digits having the properties 1 and 2 described in the preceding box.

Our goal is to use random digits to select random samples. We need the following facts about random digits, which are consequences of the basic properties 1 and 2:

- Any *pair* of random digits has the same chance of being any of the 100 possible pairs: 00, 01, 02, . . . , 98, 99.

- Any *triple* of random digits has the same chance of being any of the 1000 possible triples: 000, 001, 002, . . . , 998, 999.
- . . . and so on for groups of four or more random digits.

EXAMPLE 3.8 Brands

BRANDS

A brand is a symbol or an image that is associated with a company. An effective brand identifies the company and its products. Using a variety of measures, dollar values for brands can be calculated. In Exercise 1.53 (page 36), you examined the distribution of the values of the top 100 brands.

Suppose that you want to write a research report on some of the characteristics of the companies in this elite group. You decide to look carefully at the websites of 10 companies from the list. One way to select the companies is to use a simple random sample. Here are some details about how to do this using Table B.

We start with a list of the companies with the top 100 brands. This is given in the data file BRANDS. Next, we need to label the companies. In the data file, they are listed with their ranks, 1 to 100. Let's assign the labels 01 to 99 to the first 99 companies and 00 to the company with rank 100. With these labels, we can use Table B to select the SRS.

Let's start with line 156 of Table B. This line has the entries 55494 67690 88131 81800 11188 28552 25752 21953. These are grouped in sets of five digits, but we need to use sets of two digits for our randomization. Here is line 156 of Table B in sets of two digits: 55 49 46 76 90 88 13 18 18 00 11 18 82 85 52 25 75 22 19 53.

Using these random digits, we select Kraft (55), Accenture (49), Fox (46), Starbucks (76), Ericsson (90), Chase (88), Oracle (13), Disney (18; we skip the second 18 because we have already selected Disney to be in our SRS), Estee Lauder (00; recoded from rank 100), and BMW (11).

Most statistical software will select an SRS for you, eliminating the need for Table B. The *Simple Random Sample* applet on the text website is another convenient way to automate this task.

Excel and other spreadsheet software can do the job. There are four steps:

1. Create a data set with all the elements of the population in the first column.

2. Assign a random number to each element of the population; put these in the second column.

3. Sort the data set by the random number column.

4. The simple random sample is obtained by taking elements in order from the sorted list until the desired sample size is reached.

We illustrate the procedure with a simplified version of Example 3.8.

EXAMPLE 3.9 Select a Random Sample

Figure 3.2(a) gives the spreadsheet with the company names in column B. Only the first 12 of the 100 companies in the top 100 brands list are shown.

The random numbers generated by the RAND() function are given in the next column in Figure 3.2(b). The sorted data set is given in Figure 3.2(c). The 10 brands were selected for our random sample are Danone, Disney, Boeing, Home Depot, Nescafe, Mastercard, Gucci, Nintendo, Apple, and Credit Suisse.

(a)

	A	B
1	Rank	Brand
2	1	Apple
3	2	Microsoft
4	3	Coca-Cola
5	4	IBM
6	5	Google
7	6	McDonald's
8	7	General Electric
9	8	Intel
10	9	Samsung
11	10	Louis Vuitton
12	11	BMW
13	12	Cisco

(b)

	A	B	C
1	Rank	Brand	Random
2	1	Apple	0.066778
3	2	Microsoft	0.764104
4	3	Coca-Cola	0.201496
5	4	IBM	0.149092
6	5	Google	0.213753
7	6	McDonald's	0.425806
8	7	General Electric	0.177453
9	8	Intel	0.147609
10	9	Samsung	0.923272
11	10	Louis Vuitton	0.906979
12	11	BMW	0.363267
13	12	Cisco	0.457902

(c)

	A	B	C
1	Rank	Brand	Random
2	42	Danone	0.001657
3	17	Disney	0.003221
4	85	Boeing	0.012790
5	48	Home Depot	0.041523
6	27	Nescafe	0.049226
7	67	Mastercard	0.054817
8	38	Gucci	0.060661
9	63	Nintendo	0.060992
10	1	Apple	0.066778
11	94	Credit Suisse	0.078149
12	33	Amazon.Com	0.081160
13	84	Dell	0.083890

FIGURE 3.2 Selection of a simple random sample of brands using Excel, Example 3.9: (a) labels; (b) random numbers; (c) randomly sorted labels.

APPLY YOUR KNOWLEDGE

3.21 Ringtones for cell phones. You decide to change the ringtones for your cell phone by choosing two from a list of the 10 most popular ringtones.[12] Here is the list:

Fancy	Happy	Turn Down for What	Rude	Problem
Bottoms Up	All of Me	Crise	Beachin'	Wiggle

Select your two ringtones using a simple random sample.

3.22 Listen to three songs. The walk to your statistics class takes about 10 minutes, about the amount of time needed to listen to three songs on your iPod. You decide to take a simple random sample of songs from the top 10 songs listed on the Billboard Top Heatseekers Songs.[13] Here is the list:

Studio	Habits (Stay High)	Leave the Night On	I'm Ready
Ready Set Roll	All About That Bass	Riptide	Cool Kids
v.3005	Hope You Get Lonely Tonight		

Select the three songs for your iPod using a simple random sample.

Stratified samples

The general framework for designs that use chance to choose a sample 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 sample.*

Stratified Random Sample

To select a **stratified random sample,** first divide the population into groups of similar cases, 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 exact information than an SRS of the same size by taking advantage of the fact that cases in the same stratum are similar to one another. Think of the extreme case in which all cases in each stratum are identical: just one case from each stratum is then enough to completely describe the population.

EXAMPLE 3.10 Fraud against Insurance Companies

A dentist is suspected of defrauding insurance companies by describing some dental procedures incorrectly on claim forms and overcharging for them. An investigation begins by examining a sample of his bills for the past three years. Because there are five suspicious types of procedures, the investigators take a stratified sample. That is, they randomly select bills for each of the five types of procedures separately.

Multistage samples

Another common means of restricting random selection is to choose the sample in stages. This is common practice for national samples of households or people. For example, data on employment and unemployment are gathered by the government's Current Population Survey, which conducts interviews in about 60,000 households each month. The cost of sending interviewers to the widely scattered households in an SRS would be too high. Moreover, the government wants data broken down by states and large cities. The Current Population Survey, therefore, uses a **multistage sampling design.** The final sample consists of clusters of nearby households that an interviewer can easily visit. Most opinion polls and other national samples are also multistage, though interviewing in most national samples today is done by telephone rather than in person, eliminating the economic need for clustering. The Current Population Survey sampling design is roughly as follows:[14]

multistage sample

Stage 1. Divide the United States into 2007 geographical areas called primary sampling units, or PSUs. PSUs do not cross state lines. Select a sample of 754 PSUs. This sample includes the 428 PSUs with the largest populations and a stratified sample of 326 of the others.

Stage 2. Divide each PSU selected into smaller areas called "blocks." Stratify the blocks using ethnic and other information, and take a stratified sample of the blocks in each PSU.

Stage 3. Sort the housing units in each block into clusters of four nearby units. Interview the households in a probability sample of these clusters.

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.

APPLY YOUR KNOWLEDGE

DATA WSHOP

3.23 Who goes to the market research workshop? A small advertising firm has 30 junior associates and 10 senior associates. The junior associates are

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

The senior associates are

Andrews	Fernandez	Kim	Moore	West
Besicovitch	Gupta	Lightman	Vicario	Yang

The firm will send four junior associates and two senior associates to a workshop on current trends in market research. It decides to choose those who will go by random selection. Use Table B to choose a stratified random sample of four junior associates and two senior associates. Start at line 141 to choose your sample.

3.24 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 Table B, starting at line 125, to select *only the first five* 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. Sample surveys of large human populations, however, require much 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 as well. An opinion poll conducted by telephone will miss the 6% of American households without residential phones. Thus, the results of national sample surveys 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 case 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 a case chosen for the sample cannot be contacted or does not cooperate.

EXAMPLE 3.11 Nonresponse in the Current Population Survey

How bad is nonresponse? The Current Population Survey (CPS) has the lowest nonresponse rate of any poll we know: only about 4% of the households in the CPS sample refuse to take part, and another 3% or 4% can't be contacted. People are more likely to respond to a government survey such as the CPS, and the CPS contacts its sample in person before doing later interviews by phone.

The General Social Survey (Figure 3.3) is the nation's most important social science research survey. 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.[15]

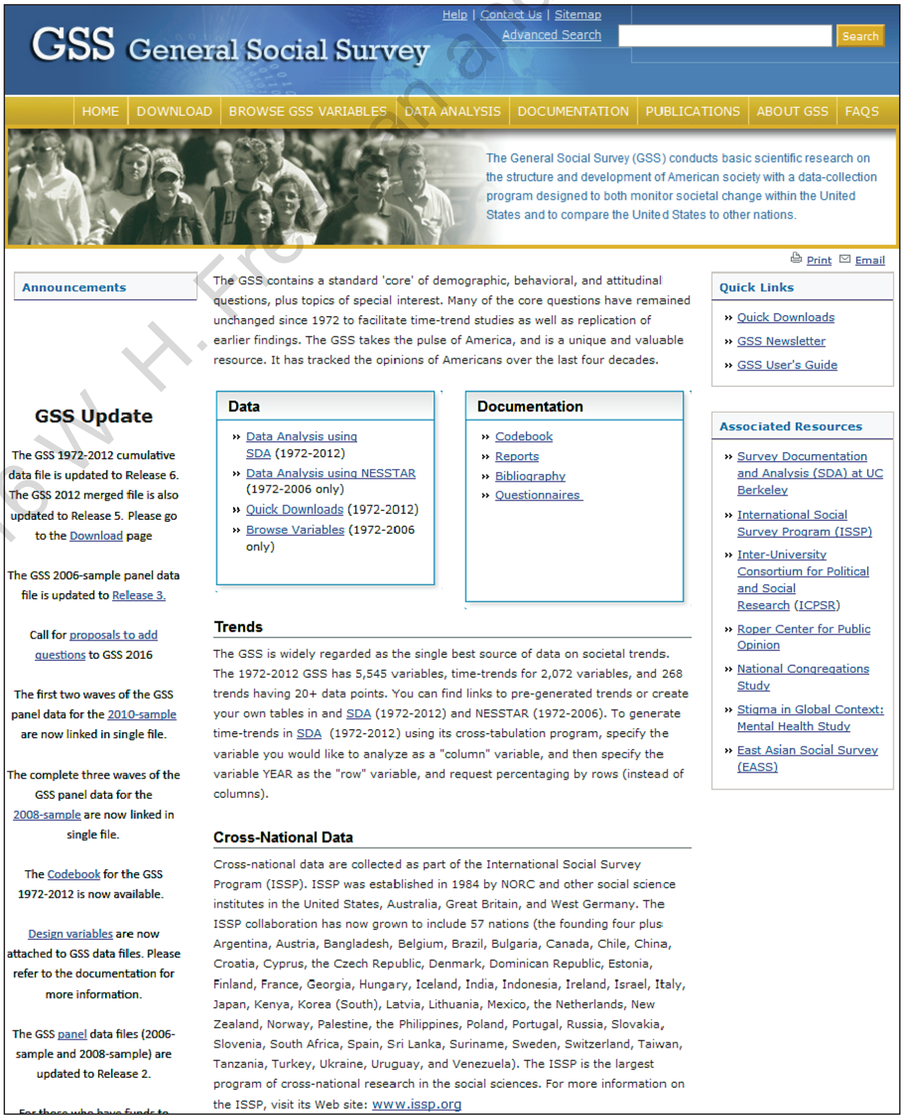


FIGURE 3.3 The General Social Survey (GSS) assesses attitudes on a variety of topics, Example 3.11.

What about polls done by the media and by market research and opinion-polling firms? We don't know their rates of nonresponse because they won't say. That in itself is a bad sign.

EXAMPLE 3.12 Change in Nonresponse in Pew Surveys

The Pew Research Center conducts research using surveys on a variety of issues, attitudes, and trends.[16] A study by the center examined the decline in the response rates to their surveys over time. The changes are dramatic, and there is a consistent pattern over time. Here are some data from the report:[17]

Year	1997	2000	2003	2006	2009	2012
Nonresponse rate	64%	72%	75%	79%	85%	91%

The center is devising alternative methods that show some promise of improving the response rates of their surveys.

Most sample surveys, and almost all opinion polls, are now carried out by telephone or online. This and other details of the interview method can affect the results. When presented with several options for a reply—such as completely agree, mostly agree, mostly disagree, and completely disagree—people tend to be a little more likely to respond to the first one or two options presented.

response bias

The behavior of the respondent or of the interviewer can cause **response bias** in sample results. Respondents may lie, especially if asked about illegal or unpopular behavior. The race or gender 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.

wording of questions

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 3.13 The Form of the Question Is Important

In response to the question "Are you heterosexual, homosexual, or bisexual?" in a social science research survey, one woman answered, "It's just me and my husband, so bisexual." The issue is serious, even if the example seems silly: reporting about sexual behavior is difficult because people understand and misunderstand sexual terms in many ways.

APPLY YOUR KNOWLEDGE

3.25 Random digit dialing. The list of cases from which a sample is actually selected is called the sampling frame. Ideally, the frame should include every case 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 that you mentioned in your answer to part (a) will be included in the sampling frame by random digit dialing?

The statistical design of sample surveys is a science, but this science is only part of the art of sampling. Because of nonresponse, response bias, and the difficulty of posing clear and neutral questions, you should hesitate to fully trust reports about complicated issues based on surveys of large human populations. *Insist on knowing the exact questions asked, the rate of nonresponse, and the date and method of the survey before you trust a poll result.*

BEYOND THE BASICS: Capture-Recapture Sampling

capture-recapture sampling

Pacific salmon return to reproduce in the river where they were hatched three or four years earlier. How many salmon made it back this year? The answer will help determine quotas for commercial fishing on the west coast of Canada and the United States. Biologists estimate the size of animal populations with a special kind of repeated sampling, called **capture-recapture sampling**. More recently, capture-recapture methods have been used on human populations as well.

EXAMPLE 3.14 Sampling for a Major Industry in British Columbia

The old method of counting returning salmon involved placing a "counting fence" in a stream and counting all the fish caught by the fence. This is expensive and difficult. For example, fences are often damaged by high water.

Repeat sampling using small nets is more practical. During this year's spawning run in the Chase River in British Columbia, Canada, you net 200 coho salmon, tag the fish, and release them. Later in the week, your nets capture 120 coho salmon in the river, of which 12 have tags.

The proportion of your second sample that have tags should estimate the proportion in the entire population of returning salmon that are tagged. So if N is the unknown number of coho salmon in the Chase River this year, we should have approximately

$$\text{proportion tagged in sample} = \text{proportion tagged in population}$$

$$\frac{12}{120} = \frac{200}{N}$$

Solve for N to estimate that the total number of salmon in this year's spawning run in the Chase River is approximately

$$N = 200 \times \frac{120}{12} = 2000$$

The capture-recapture idea extends the use of a sample proportion to estimate a population proportion. The idea works well if both samples are SRSs from the population and the population remains unchanged between samples. In practice, complications arise. For example, some tagged fish might be caught by bears or otherwise die between the first and second samples.

Variations on capture-recapture samples are widely used in wildlife studies and are now finding other applications. One way to estimate the census undercount in a district is to consider the census as "capturing and marking" the households that respond. Census workers then visit the district, take an SRS of households, and see how many of those counted by the census show up in the sample. Capture-recapture estimates the total count of households in the district. As with estimating wildlife populations, there are many practical pitfalls. Our final word is as before: the real world is less orderly than statistics textbooks imply.

SECTION 3.2 Summary

- A sample survey selects a **sample** from the **population** that is the object of our study. We base conclusions about the population on data collected from the sample.
- The **design** of a sample refers to the method used to select the sample from the population. **Probability sampling designs** use impersonal 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,** divide the population into **strata,** or groups of cases that are similar in some way that is important to the response. Then choose a separate SRS from each stratum, and combine them to form the full sample.
- **Multistage samples** select successively smaller groups within the population in stages, resulting in a sample consisting of clusters of cases. Each stage may employ an SRS, a stratified sample, or another type of sample.
- 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** due to the behavior of the interviewer or the respondent, or from misleading results due to **poorly worded questions.**

SECTION 3.2 Exercises

For Exercises 3.17 and 3.18 see page 130; for 3.19 and 3.20, see pages 131–132; for 3.21 and 3.22, see page 134; for 3.23 and 3.24, see page 136; and for 3.25, see page 138.

3.26 What's wrong? Explain what is wrong in each of the following statements.
(a) A simple random sample is the only way to randomly select cases from a population.
(b) Random digits cannot be used to select a sample from a population that has more than 100 cases.
(c) The population consists of all cases selected in a simple random sample.

3.27 What's wrong? Explain what is wrong with each of the following random selection procedures, and explain how you would do the randomization correctly.
(a) To determine the reading level of an introductory statistics text, you evaluate all of the written material in the third chapter.
(b) You want to sample student opinions about a proposed change in procedures for changing majors. You hand out questionnaires to 100 students as they arrive for class at 7:30 A.M.
(c) A population of subjects is put in alphabetical order, and a simple random sample of size 10 is taken by selecting the first 10 subjects in the list.

3.28 Importance of students as customers. 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 120 businesses at random. Of these, 54 return the questionnaire mailed by the committee. What is the population for this sample survey? What is the sample? What is the rate (percent) of nonresponse?

3.29 Popularity of news personalities can affect market share. A Gallup Poll conducted telephone interviews with 1001 U.S. adults aged 18. One of the questions asked whether the respondents had a favorable or an unfavorable opinion of 17 news personalities. Diane Sawyer received the highest rating, with 80% of the respondents giving her a favorable rating.[18]
(a) What is the population for this sample survey? What was the sample size?

(b) The report on the survey states that 8% of the respondents either never heard of Sawyer or had no opinion about her. When they included only those who provided an opinion, Sawyer's approval percent rose to 88%, and she was still at the top of the list. Charles Gibson, on the other hand, was ranked eighth on the original list, with a 55% favorable rating. When only those providing an opinion were counted, his rank rose to second, with 87% approving. Discuss the advantages and disadvantages of the two different ways of reporting the approval percent. State which one you prefer and why.

3.30 Identify the populations. For each of the following sampling situations, identify the population as exactly as possible. That is, indicate what kind of cases the population consists of and exactly which cases fall in the population. If the information given is not complete, complete the description of the population in a reasonable way.
(a) A college has changed its core curriculum and wants to obtain detailed feedback information from the students during each of the first 12 weeks of the coming semester. Each week, a random sample of five students will be selected to be interviewed.
(b) The American Community Survey (ACS) replaced the census "long form" starting with the 2010 census. The main part of the ACS contacts 250,000 addresses by mail each month, with follow-up by phone and in person if there is no response. Each household answers questions about their housing, economic, and social status.
(c) 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?"

3.31 Interview potential customers. You have been hired by a company that is planning to build a new apartment complex for students in a college town. They want you to collect information about preferences of potential customers for their complex. Most of the college students who live in apartments live in one of 33 complexes. You decide to select six apartment complexes at random for in-depth interviews with residents. Select a simple random sample of six of the following apartment complexes. If you use Table B, start at line 107. RESID

Ashley Oaks	Burberry	Del-Lynn
Bay Pointe	Cambridge	Fairington
Beau Jardin	Chauncey Village	Fairway Knolls
Bluffs	Country Squire	Fowler
Brandon Place	Country View	Franklin Park
Briarwood	Country Villa	Georgetown
Brownstone	Crestview	Greenacres
Lahr House	Peppermill	Salem Courthouse
Mayfair Village	Pheasant Run	Village Manor
Nobb Hill	Richfield	Waterford Court
Pemberly Courts	Sagamore Ridge	Williamsburg

(*Continued*)

3.32 Using GIS to identify mint field conditions. A Geographic Information System (GIS) is to be used to distinguish different conditions in mint fields. Ground observations will be used to classify regions of each field as either healthy mint, diseased mint, or weed-infested mint. The GIS divides mint-growing areas into regions called pixels. An experimental area contains 200 pixels. For a random sample of 30 pixels, ground measurements will be made to determine the status of the mint, and these observations will be compared with information obtained by the GIS. Select the random sample. If you use Table B, start at line 152 and choose only the first six pixels in the sample.

APPLET **3.33 Select a simple random sample.** After you have labeled the cases in a population, the *Simple Random Sample* applet automates the task of choosing an SRS. Use the applet to choose the sample in the previous exercise.

APPLET **3.34 Select a simple random sample.** There are approximately 446 active telephone area codes covering Canada, the United States, and some Caribbean areas.[19] (More are created regularly.) You want to choose an SRS of 30 of these area codes for a study of available telephone numbers. Use software or the *Simple Random Sample* applet to choose your sample. ACODES

3.35 Repeated use of Table B. In using Table B repeatedly to choose samples, you should not always begin at the same place, such as line 101. Why not?

3.36 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 four addresses out of 100. Because 100/4 = 25, we can think of the list as four lists of 25 addresses. Choose one 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) A study of dating among college students wanted a sample of 200 of the 9000 single male students on campus. The sample consisted of every 45th name from

a list of the 9000 students. Explain why the survey chooses every 45th name.
(b) Use software or Table B at line 135 to choose the starting point for this systematic sample.

3.37 Systematic random samples versus simple random samples. The previous exercise introduces systematic random samples. Explain carefully why a systematic random sample *does* give every case the same chance to be chosen but is *not* a simple random sample.

3.38 Random digit telephone dialing for market research. A market research firm in California uses random digit dialing to choose telephone numbers at random. Numbers are selected separately within each California area code. The size of the sample in each area code is proportional to the population living there.
(a) What is the name for this kind of sampling design?
(b) California area codes, in rough order from north to south, are

209	213	310	323	408	415	510	530	559	562
619	626	650	661	707	714	760	805	818	831
858	909	916	925	949					

Another California survey does not call numbers in all area codes but starts with an SRS of eight area codes. Choose such an SRS. If you use Table B, start at line 132. CACODES

3.39 Select employees for an awards committee. A department has 30 hourly workers and 10 salaried workers. The hourly workers are

Abel	Fisher	Huber	Moran	Reinmann
Carson	Golomb	Jimenez	Moskowitz	Santos
Chen	Griswold	Jones	Neyman	Shaw
David	Hein	Kiefer	O'Brien	Thompson
Deming	Hernandez	Klotz	Pearl	Utts
Elashoff	Holland	Liu	Potter	Vlasic

and the salaried workers are

Andrews	Fernandez	Kim	Moore	Rabinowitz
Besicovitch	Gupta	Lightman	Phillips	Yang

The committee will have seven hourly workers and three salaried workers. Random selection will be used to select the committee members. Select a stratified random sample of seven hourly workers and three salaried workers. CMEMB

3.40 When do you ask? When observations are taken over time, it is important to check for patterns that may be important for the interpretation of the data. In Section 1.2 (page 19), we learned to use a time plot for this purpose. Describe and discuss a sample survey question where you would expect to have variation over time (answers would be different at different times) for the following situations:
(a) Data are taken at each hour of the day from 8 A.M. to 6 P.M.
(b) Date are taken on each of the seven days of the week.
(c) Data are taken during each of the 12 months of the year.

3.41 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 escalating environmental degradation and incipient resource depletion, would you favor economic incentives for recycling of resource-intensive consumer goods?"

3.3 Designing Experiments

A study is an experiment when we actually do something to people, animals, or objects in order to observe the response. Here is the basic vocabulary of experiments.

> **Experimental Units, Subjects, Treatment**
> The cases on which the experiment is done are the **experimental units.** When the units are human beings, they are called **subjects.** A specific experimental condition applied to the units is called a **treatment.**

Because the purpose of an experiment is to reveal the response of one variable to changes in other variables, the distinction between explanatory and response variables is important. The explanatory variables in an experiment are often called **factors.** Many experiments study the joint effects of several factors. In such an experiment, each treatment is formed by combining a specific value (often called a **level**) of each of the factors.

factors

level of a factor

EXAMPLE 3.15 Is the Cost Justified?

The increased costs for teacher salaries and facilities associated with smaller class sizes can be substantial. Are smaller classes really better? We might do an observational study that compares students who happened to be in smaller and larger classes in their early school years. Small classes are expensive, so they are more common in schools that serve richer communities. Students in small classes tend to also have other advantages: their schools have more resources, their parents are better educated, and so on. The size of the classes is **confounded** with other characteristics of the students, making it impossible to isolate the effects of small classes.

confounded

The Tennessee STAR program was an experiment on the effects of class size. It has been called "one of the most important educational investigations ever carried out." The *subjects* were 6385 students who were beginning kindergarten. Each student was assigned to one of three *treatments:* regular class (22 to 25 students) with one teacher, regular class with a teacher and a full-time teacher's aide, and small class (13 to 17 students) with one teacher. These treatments are levels of a single *factor:* the type of class. The students stayed in the same type of class for four years, then all returned to regular classes. In later years, students from the small classes had higher scores on standard tests, were less likely to fail a grade, had better high school grades, and so on. The benefits of small classes were greatest for minority students.[20]

Example 3.15 illustrates the big advantage of experiments over observational studies. **In principle, experiments can give good evidence for causation.** In an experiment, we study the specific factors we are interested in, while controlling the effects of lurking variables. All the students in the Tennessee STAR program followed the usual curriculum at their schools. Because students were assigned to different class types within their schools, school resources and family backgrounds were not confounded with class type. The only systematic difference was the type of class. When students from the small classes did better than those in the other two types, we can be confident that class size made the difference.

EXAMPLE 3.16 Effects of TV Advertising

What are the effects of repeated exposure to an advertising message? The answer may depend both on the length of the ad and on how often it is repeated. An experiment investigates this question using undergraduate students as subjects. All subjects view a 40-minute television program that includes ads for a digital camera. Some subjects see a 30-second commercial; others, a 90-second version. The same commercial is repeated one, three, or five times during the program. After viewing, all of the subjects answer questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it. These are the response variables.[21]

This experiment has two factors: length of the commercial, with two levels; and repetitions, with three levels. All possible combinations of the 2×3 factor levels form six treatment combinations. Figure 3.4 shows the layout of these treatments.

FIGURE 3.4 The treatments in the experimental design of Example 3.16. Combinations of levels of the two factors form six treatments.

Factor A Length	Factor B Repetitions: 1 time	3 times	5 times
30 seconds	1	2	3
90 seconds	4	5	6

Experimentation allows us to study the effects of the specific treatments we are interested in. Moreover, we can control the environment of the subjects to hold constant the factors that are of no interest to us, such as the specific product advertised in Example 3.16. In one sense, the ideal case is a laboratory experiment in which we control all lurking variables and so see only the effect of the treatments on the response. On the other hand, the effects of being in an artificial environment such as a laboratory may also affect the outcomes. *The balance between control and realism is an important consideration in the design of experiments.*

Another advantage of experiments is that we can study the combined effects of several factors simultaneously. The interaction of several factors can produce effects that could not be predicted from looking at the effect of each factor alone. Perhaps longer commercials increase interest in a product, and more commercials also increase interest, but if we make a commercial longer *and* show it more often, viewers get annoyed and their interest in the product drops. The two-factor experiment in Example 3.16 will help us find out.

APPLY YOUR KNOWLEDGE

3.42 Radiation and storage time for food products. Storing food for long periods of time is a major challenge for those planning for human space travel beyond the moon. One problem is that exposure to radiation decreases the length of time that food can be stored. One experiment examined the effects of nine different levels of radiation on a particular type of fat, or lipid.[22] The amount of oxidation of the lipid is the measure of the extent of the damage due to the radiation. Three samples are exposed to each radiation level. Give the experimental units, the treatments, and the response variable. Describe the factor and its levels. There are many different types of lipids. To what extent do you think the results of this experiment can be generalized to other lipids?

3.43 Can they use the Web? A course in computer graphics technology requires students to learn multiview drawing concepts. This topic is traditionally taught using supplementary material printed on paper. The instructor of the course believes that a web-based interactive drawing program will be more effective in increasing the drawing skills of the students.[23] The 50 students who are enrolled in the course will be randomly assigned to either the paper-based instruction or the web-based instruction. A standardized drawing test will be given before and after the instruction. Explain why this study is an experiment, and give the experimental units, the treatments, and the response variable. Describe the factor and its levels. To what extent do you think the results of this experiment can be generalized to other settings?

3.44 Is the packaging convenient for the customer? A manufacturer of food products uses package liners that are sealed by applying heated jaws after the package is filled. The customer peels the sealed pieces apart to open the package. What effect does the temperature of the jaws have on the force needed to peel the liner? To answer this question, engineers prepare 20 package liners. They seal five liners at each of four different temperatures: 250°F, 275°F, 300°F, and 325°F. Then they measure the force needed to peel each seal.

(a) What are the experimental units studied?
(b) There is one factor (explanatory variable). What is it, and what are its levels?
(c) What is the response variable?

Comparative experiments

Many experiments have a simple design with only a single treatment, which is applied to all experimental units. The design of such an experiment can be outlined as

Treatment ⟶ **Observe response**

EXAMPLE 3.17 Increase the Sales Force

A company may increase its sales force in the hope that sales will increase. The company compares sales before the increase with sales after the increase. Sales are up, so the manager who suggested the change gets a bonus.

Increase the sales force ⟶ **Observe sales**

The sales experiment of Exercise 3.17 was poorly designed to evaluate the effect of increasing the sales force. Perhaps sales increased because of seasonal variation in demand or other factors affecting the business.

placebo effect

In medical settings, an improvement in condition is sometimes due to a phenomenon called the **placebo effect.** In medicine, a placebo is a dummy or fake treatment, such as a sugar pill. Many participants, regardless of treatment, respond favorably to personal attention or to the expectation that the treatment will help them.

comparative experiment

For the sales force study, we don't know whether the increase in sales was due to increasing the sales force or to other factors. The experiment gave inconclusive results because the effect of increasing the sales force was confounded with other factors that could have had an effect on sales. The best way to avoid confounding is to do a **comparative experiment.** Think about a study where the sales force is increased in half of the regions where the product is sold and is not changed in the other regions. A comparison of sales from the two sets of regions would provide an evaluation of the effect of the increasing the sales force.

control group
treatment group

In medical settings, it is standard practice to randomly assign patients to either a **control group** or a **treatment group.** All patients are treated the same in every way except that the treatment group receives the treatment that is being evaluated. In the setting of our comparative sales experiment, we would randomly divide the regions into two groups. One group will have the sales force increased and the other group will not.

REMINDER
bias, p. 131

Uncontrolled experiments in medicine and the behavioral sciences can be dominated by such influences as the details of the experimental arrangement, the selection of subjects, and the placebo effect. The result is often *bias.*

An uncontrolled study of a new medical therapy, for example, is biased in favor of finding the treatment effective because of the placebo effect. It should not surprise you to learn that uncontrolled studies in medicine give new therapies a much higher success rate than proper comparative experiments do. Well-designed experiments usually compare several treatments.

APPLY YOUR KNOWLEDGE

3.45 Does using statistical software improve exam scores? An instructor in an elementary statistics course wants to know if using a new statistical software package will improve students' final-exam scores. He asks for volunteers, and approximately half of the class agrees to work with the new software. He compares the final-exam scores of the students who used the new software with the scores of those who did not. Discuss possible sources of bias in this study.

Randomized comparative experiments

experiment design

The **design of an experiment** first describes the response variables, the factors (explanatory variables), and the layout of the treatments, with *comparison* as the leading principle. The second aspect of design is the rule used to assign the subjects to the treatments. Comparison of the effects of several treatments is valid only when all treatments are applied to similar groups of subjects. If one corn variety is planted on more fertile ground, or if one cancer drug is given to less seriously ill patients, comparisons among treatments are biased. How can we assign cases to treatments in a way that is fair to all the treatments?

randomization

randomized comparative experiment

Our answer is the same as in sampling: let impersonal chance make the assignment. The use of chance to divide subjects into groups is called **randomization.** Groups formed by randomization don't depend on any characteristic of the subjects or on the judgment of the experimenter. An experiment that uses both comparison and randomization is a **randomized comparative experiment.** Here is an example.

EXAMPLE 3.18 Testing a Breakfast Food

A food company assesses the nutritional quality of a new "instant breakfast" product by feeding it to newly weaned male white rats. The response variable is a rat's weight gain over a 28-day period. A control group of rats eats a standard diet but otherwise receives exactly the same treatment as the experimental group.

This experiment has one factor (the diet) with two levels. The researchers use 30 rats for the experiment and so divide them into two groups of 15. To do this in an unbiased fashion, put the cage numbers of the 30 rats in a hat, mix them up, and draw 15. These rats form the experimental group and the remaining 15 make up the control group. *Each group is an SRS of the available rats.* Figure 3.5 outlines the design of this experiment.

FIGURE 3.5 Outline of a randomized comparative experiment, Example 3.18.

APPLY YOUR KNOWLEDGE

3.46 Diagram the food storage experiment. Refer to Exercise 3.42 (page 144). Draw a diagram similar to Figure 3.5 that describes the food for space travel experiment.

3.47 Diagram the Web use. Refer to Exercise 3.43 (page 144). Draw a diagram similar to Figure 3.5 that describes the computer graphics drawing experiment.

Completely randomized designs

The design in Figure 3.5 combines comparison and randomization to arrive at the simplest statistical design for an experiment. This "flowchart" outline presents all the essentials: randomization, the sizes of the groups and which treatment they receive, and the response variable. There are, as we will see later, statistical reasons for generally using treatment groups that are approximately equal in size. We call designs like that in Figure 3.5 *completely randomized.*

Completely Randomized Design

In a **completely randomized** experimental design, all the subjects are allocated at random among all the treatments.

Completely randomized designs can compare any number of treatments. Here is an example that compares three treatments.

EXAMPLE 3.19 Utility Companies and Energy Conservation

Many utility companies have introduced programs to encourage energy conservation among their customers. An electric company considers placing electronic meters in households to show what the cost would be if the electricity use at that moment continued for a month. Will these meters reduce electricity use? Would cheaper methods work almost as well? The company decides to design an experiment.

One cheaper approach is to give customers a chart and information about monitoring their electricity use. The experiment compares these two approaches (meter, chart) and also a control. The control group of customers receives information about energy conservation but no help in monitoring electricity use. The response variable is total electricity used in a year. The company finds 60 single-family residences in the same city willing to participate, so it assigns 20 residences at random to each of the three treatments. Figure 3.6 outlines the design.

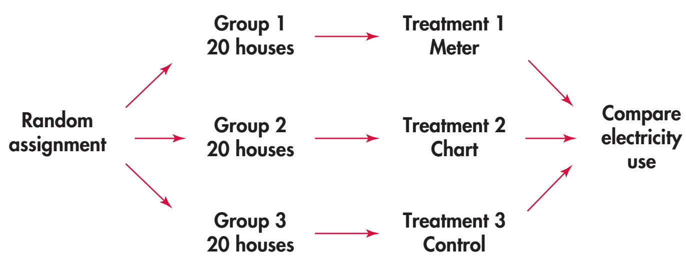


FIGURE 3.6 Outline of a completely randomized design comparing three treatments, Example 3.19.

How to randomize

The idea of randomization is to assign experimental units to treatments by drawing names from a hat. In practice, experimenters use software to carry out randomization. In Example 3.19, we have 60 residences that need to be randomly assigned to three treatments. Most statistical software will be able to do the randomization required.

We prefer to use software for randomizing but if you do not have that option available to you, a *table of random digits,* such as Table B can be used. Using software, the method is similar to what we used to select an SRS in Example 3.9 (page 133). Here are the steps needed:

Step 1: Label. Give each experimental unit a unique label. For privacy reasons, we might want to use a numerical label and a keep a file that identifies the experimental units with the number in a separate place.

Step 2: Use the computer. Once we have the labels, we create a data file with the labels and generate a random number for each label. In Excel, this can be done with the RAND() function. Finally, we sort the entire data set based on the random numbers. Groups are formed by selecting units in order from the sorted list.

EXAMPLE 3.20 Do the Randomization for the Utility Company Experiment Using Excel

In the utility company experiment of Example 3.19, we must assign 60 residences to three treatments. First we generate the labels. Let's use numerical labels and keep a separate file that gives the residence address for each number. So for Step 1, we will use these labels, 1 to 60:

1, 2, 3, . . . , 59, 60

To illustrate Step 2, we will show several Excel files. To see what we are doing, it will be easier if we reduce the number of residences to be randomized. So, let's randomize 12 residences to the three treatments. Our labels are

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

For the first part of Step 2, we create an Excel file with the numbers 1 to 12 in the first column. This file is shown in Figure 3.7(a). Next, we use the RAND() function in Excel to generate 12 random numbers in the second column. The result is shown in Figure 3.7(b). We then sort the file based in the random numbers. We create a third column with the following treatments: "Meter" for the first four, "Chart" for the next four, and "Control" for the last four. The result is displayed in Figure 3.7(c).

Excel

	A	B	C
1	Label		
2	1		
3	2		
4	3		
5	4		
6	5		
7	6		
8	7		
9	8		
10	9		
11	10		
12	11		
13	12		

Sheet1

(a)

Excel

	A	B	C
1	Label	Random	
2	1	0.3853602	
3	2	0.0735648	
4	3	0.8955801	
5	4	0.2396214	
6	5	0.9765729	
7	6	0.7951346	
8	7	0.7646003	
9	8	0.4595461	
10	9	0.1309937	
11	10	0.3216289	
12	11	0.7165543	
13	12	0.9886189	

Sheet1

(b)

Excel

	A	B	C
1	Label	Random	Treatment
2	2	0.0735648	Meter
3	9	0.1309937	Meter
4	4	0.2396214	Meter
5	10	0.3216289	Meter
6	1	0.3853602	Chart
7	8	0.4595461	Chart
8	11	0.7165543	Chart
9	7	0.7646003	Chart
10	6	0.7951346	Control
11	3	0.8955801	Control
12	5	0.9765729	Control
13	12	0.9886189	Control

Sheet1

(c)

FIGURE 3.7 Randomization of 12 experimental units to three treatments using Excel, Example 3.20: (a) labels; (b) random numbers; (c) randomly sorted labels with treatment assignments.

If software is not available, you can use the random digits in Table B to do the randomization. The method is similar to the one we used to select an SRS in Example 3.8 (page 133). Here are the steps that you need:

Step 1: Label. Give each experimental unit a numerical label. Each label must contain the same number of digits. So, for example, if you are randomizing 10 experimental units, you could use the labels, 0, 1, . . . , 8, 9; or 01, 02, . . . , 10. Note that with the first choice you need only one digit, but for the second choice, you need two.

Step 2: Table. Start anywhere in Table B and read digits in groups corresponding to one-digit or two-digit groups. (You really do not want to use Table B for more than 100 experimental units. Software is needed here.)

EXAMPLE 3.21 Do the Randomization for the Utility Company Experiment Using Random Digits

As we did in Example 3.20, we will illustrate the method by randomizing 12 residences to three treatments. For Step 1, we assign the 12 residences the following labels:

01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12

Compare these labels with the ones we used in Example 3.20. Here, we need the same number of digits for each label, so we put a zero as the first digit for the first nine labels.

For Step 2, we will use Table B starting at line 118. Here are the table entries for that line:

73190 32533 04470 29669 84407 90785 65956 86382

To make our work a little easier, we rewrite these digits in pairs:

73 19 03 25 33 04 47 02 96 69 84 40 79 07 85 65 95 68 63 82

We now select the labels for the first treatment, "Meter." Reading pairs of digits from left to write and ignoring pairs that do not correspond to any of our labels, we see the labels 03, 04, 02, and 07. The corresponding residences will receive the "Meter" treatment. We will continue the process to find four labels to be assigned to the "Chart" treatment. We continue to the next line in Table B, where we do not find any labels between 01 and 12. On line 120, we have the label 04. This label has already been assigned to a treatment so we ignore it. Line 121 has two labels between 01 and 12: 07, which has already been assigned to a treatment, and 10, which we assign to "Chart." On the next line, we have 05, 09, and 08 which we also assign to "Chart." The remaining four labels are assigned to the "Control" treatment. In summary, 02, 03, 04, and 07 are assigned to "Meter," 05, 08, 09, and 10 are assigned to "Chart," and 01, 06, 11, and 12 are assigned to "Control."

CAUTION

As Example 3.21 illustrates, randomization requires two steps: assign labels to the experimental units and then use Table B to select labels at random. *Be sure that all labels are the same length so that all have the same chance to be chosen.* You can read digits from Table B in any order—along a row, down a column, and so on—because the table has no order. As an easy standard practice, we recommend reading along rows. In Example 3.21, we needed 180 random digits from four and a half lines (118 to 121 and half of 122) to complete the randomization. If we wanted to reduce this amount, we could use more than one label for each residence. For example, we could use labels 01, 21, 41, 61, and 81 for the first residence; 02, 22, 42, 62, and 82 for the second residence; and so forth.

Examples 3.18 and 3.19 describe completely randomized designs that compare levels of a single factor. In Example 3.18, the factor is the diet fed to the rats. In Example 3.19, it is the method used to encourage energy conservation. Completely randomized designs can have more than one factor. The advertising experiment of Example 3.16 has two factors: the length and the number of repetitions of a television commercial. Their combinations form the six treatments outlined in Figure 3.4 (page 144). A completely randomized design assigns subjects at random to these six treatments. Once the layout of treatments is set, the randomization needed for a completely randomized design is tedious but straightforward.

APPLY YOUR KNOWLEDGE

CCARE

3.48 Does child care help recruit employees? Will providing child care for employees make a company more attractive to women? You are designing an experiment to answer this question. You prepare recruiting material for two fictitious companies, both in similar businesses in the same location. Company A's brochure does not mention child care. There are two versions of Company B's brochure. One is identical to Company A's brochure. The other is also the same, but a description of the company's onsite child care facility is included. Your subjects are 40 women who are college seniors seeking employment. Each subject will read recruiting material for Company A and one of the versions of the recruiting material for Company B. You will give each version of Company B's brochure to half the women. After reading the material for both companies, each subject chooses the one she would prefer to work for. You expect that a higher percent of those who read the description that includes child care will choose Company B.

(a) Outline an appropriate design for the experiment.
(b) The names of the subjects appear below. Use software or Table B, beginning at line 112, to do the randomization required by your design. List the subjects who will read the version that mentions child care.

Abrams	Danielson	Gutierrez	Lippman	Rosen
Adamson	Durr	Howard	Martinez	Sugiwara
Afifi	Edwards	Hwang	McNeill	Thompson
Brown	Fluharty	Iselin	Morse	Travers
Cansico	Garcia	Janle	Ng	Turing
Chen	Gerson	Kaplan	Quinones	Ullmann
Cortez	Green	Kim	Rivera	Williams
Curzakis	Gupta	Lattimore	Roberts	Wong

3.49 Sealing food packages. Use a diagram to describe a completely randomized experimental design for the package liner experiment of Exercise 3.44 (page 145). (Show the size of the groups, the treatment each group receives, and the response variable. Figures 3.5 and 3.6 are models to follow.) Use software or Table B, starting at line 140, to do the randomization required by your design.

The logic of randomized comparative experiments

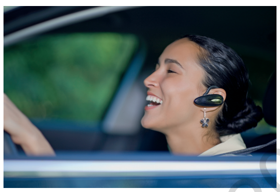
Randomized comparative experiments are designed to give good evidence that differences in the treatments actually *cause* the differences we see in the response. The logic is as follows:

- Random assignment of subjects forms groups that should be similar in all respects before the treatments are applied.
- Comparative design ensures that influences other than the experimental treatments operate equally on all groups.
- Therefore, differences in average response must be due either to the treatments or to the play of chance in the random assignment of subjects to the treatments.

That “either-or” deserves more thought. In Example 3.18 (page 146), we cannot say that *any* difference in the average weight gains of rats fed the two diets must be caused by a difference between the diets. There would be some difference even if both groups received the same diet because the natural variability among rats means that some grow faster than others. If chance assigns the faster-growing rats to one group or the other, this creates a chance difference between the groups. We would not trust an experiment with just one rat in each group, for example. The results would depend on which group got lucky and received the faster-growing rat. If we assign many rats to each diet, however, the effects of chance will average out, and there will be little difference in the average weight gains in the two groups unless the diets themselves cause a difference. “Use enough subjects to reduce chance variation” is the third big idea of statistical design of experiments.

Principles of Experimental Design

1. **Compare** two or more treatments. This will control the effects of lurking variables on the response.
2. **Randomize**—use chance to assign subjects to treatments.
3. **Replicate** each treatment on enough subjects to reduce chance variation in the results.


JED SHARE/KAORU SHARE/ GETTY IMAGES

EXAMPLE 3.22 Cell Phones and Driving

Does talking on a hands-free cell phone distract drivers? Undergraduate students “drove” in a high-fidelity driving simulator equipped with a hands-free cell phone. The car ahead brakes: how quickly does the subject respond? Twenty students (the control group) simply drove. Another 20 (the experimental group) talked on the cell phone while driving. The simulator gave the same driving conditions to both groups.[24]

This experimental design has good control because the only difference in the conditions for the two groups is the use of the cell phone. Students are randomized to the two groups, so we satisfy the second principle. Based on past experience with the simulators, the length of the drive and the number of subjects were judged to provide sufficient information to make the comparison. (We learn more about choosing sample sizes for experiments in starting Chapter 7.)

We hope to see a difference in the responses so large that it is unlikely to happen just because of chance variation. We can use the laws of probability, which give a mathematical description of chance behavior, to learn if the treatment effects are larger than we would expect to see if only chance were operating. If they are, we call them *statistically significant*.

statistically significant

Statistical Significance

An observed effect so large that it would rarely occur by chance is called **statistically significant.**

If we observe statistically significant differences among the groups in a comparative randomized experiment, we have good evidence that the treatments actually caused these differences. You will often see the phrase "statistically significant" in reports of investigations in many fields of study. The great advantage of randomized comparative experiments is that they can produce data that give good evidence for a cause-and-effect relationship between the explanatory and response variables. We know that, in general, a strong association does not imply causation. A statistically significant association in data from a well-designed experiment does imply causation.

APPLY YOUR KNOWLEDGE

3.50 Utility companies. Example 3.19 (page 147) describes an experiment to learn whether providing households with electronic meters or charts will reduce their electricity consumption. An executive of the utility company objects to including a control group. He says, "It would be simpler to just compare electricity use last year (before the meter or chart was provided) with consumption in the same period this year. If households use less electricity this year, the meter or chart must be working." Explain clearly why this design is inferior to that in Example 3.19.

3.51 Statistical significance. The financial aid office of a university asks a sample of students about their employment and earnings. The report says that "for academic year earnings, a significant difference was found between the sexes, with men earning more on the average. No significant difference was found between the earnings of black and white students." Explain the meaning of "a significant difference" and "no significant difference" in plain language.

Completely randomized designs can compare any number of treatments. The treatments can be formed by levels of a single factor or by more than one factor. Here is an example with two factors.

EXAMPLE 3.23 Randomization for the TV Commercial Experiment

Figure 3.4 (page 144) displays six treatments formed by the two factors in an experiment on response to a TV commercial. Suppose that we have 150 students who are willing to serve as subjects. We must assign 25 students at random to each group. Figure 3.8 outlines the completely randomized design.

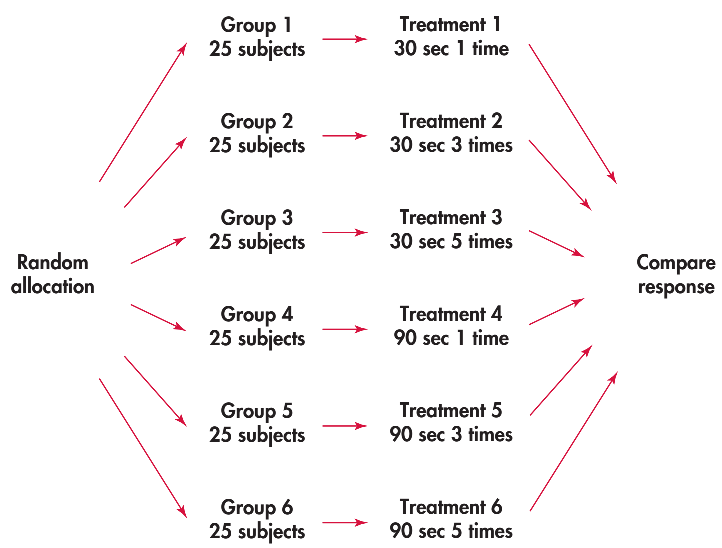


FIGURE 3.8 Outline of a completely randomized design for comparing six treatments, Example 3.23.

To carry out the random assignment, label the 150 students 001 to 150. (Three digits are needed to label 150 subjects.) Enter Table B and read three-digit groups until you have selected 25 students to receive Treatment 1 (a 30-second ad shown once). If you start at line 140, the first few labels for Treatment 1 subjects are 129, 048, and 003.

Continue in Table B to select 25 more students to receive Treatment 2 (a 30-second ad shown three times). Then select another 25 for Treatment 3 and so on until you have assigned 125 of the 150 students to Treatments 1 through 5. The 25 students who remain get Treatment 6. The randomization is straightforward but very tedious to do by hand. We recommend software such as the *Simple Random Sample* applet. Exercise 3.62 (page 158) shows how to use the applet to do the randomization for this example.

APPLET

APPLY YOUR KNOWLEDGE

3.52 Do the randomization. Use computer software to carry out the randomization in Example 3.23.

Cautions about experimentation

The logic of a randomized comparative experiment depends on our ability to treat all the subjects identically in every way except for the actual treatments being compared. Good experiments therefore require careful attention to details.

CAUTION

Many—perhaps most—experiments have some weaknesses in detail. The environment of an experiment can influence the outcomes in unexpected ways. Although experiments are the gold standard for evidence of cause and effect, really convincing evidence usually requires that a number of studies in different places with different details produce similar results. The most serious potential weakness of experiments is **lack of realism.** The subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study. Here are two examples.

lack of realism

EXAMPLE 3.24 Layoffs and Feeling Bad

How do layoffs at a workplace affect the workers who remain on the job? Psychologists asked student subjects to proofread text for extra course credit, then "let go" some of the workers (who were actually accomplices of the experimenters). Some subjects were told that those let go had performed poorly (Treatment 1). Others were told that not all could be kept and that it was just luck that they were kept and others let go (Treatment 2). We can't be sure that the reactions of the students are the same as those of workers who survive a layoff in which other workers lose their jobs. Many behavioral science experiments use student subjects in a campus setting. Do the conclusions apply to the real world?

EXAMPLE 3.25 Does the Regulation Make the Product Safer?

Do those high center brake lights, required on all cars sold in the United States since 1986, really reduce rear-end collisions? Randomized comparative experiments with fleets of rental and business cars, done before the lights were required, showed that the third brake light reduced rear-end collisions by as much as 50%. Unfortunately, requiring the third light in all cars led to only a 5% drop.

What happened? Most cars did not have the extra brake light when the experiments were carried out, so it caught the eye of following drivers. Now that almost all cars have the third light, they no longer capture attention.

Lack of realism can limit our ability to apply the conclusions of an experiment to the settings of greatest interest. Most experimenters want to generalize their conclusions to some setting wider than that of the actual experiment. Statistical analysis of the original experiment cannot tell us how far the results will generalize. Nonetheless, the randomized comparative experiment, because of its ability to give convincing evidence for causation, is one of the most important ideas in statistics.

APPLY YOUR KNOWLEDGE

3.53 Managers and stress. Some companies employ consultants to train their managers in meditation in the hope that this practice will relieve stress and make the managers more effective on the job. An experiment that claimed to show that meditation reduces anxiety proceeded as follows. The experimenter interviewed the subjects and rated their level of anxiety. Then the subjects were randomly assigned to two groups. The experimenter taught one group how to meditate, and they meditated daily for a month. The other group was simply told to relax more. At the end of the month, the experimenter interviewed all the subjects again and rated their anxiety level. The meditation group now had less anxiety. Psychologists said that the results were suspect because the ratings were not blind—that is, the experimenter knew which treatment each subject received. Explain what this means and how lack of blindness could bias the reported results.

3.54 Frustration and teamwork. A psychologist wants to study the effects of failure and frustration on the relationships among members of a work team. She forms a team of students, brings them to the psychology laboratory, and has them play a game that requires teamwork. The game is rigged so that they lose regularly. The psychologist observes the students through a one-way window and notes the changes in their behavior during an evening of game playing. Why is it doubtful that the findings of this study tell us much about the effect of working for months developing a new product that never works right and is finally abandoned by your company?

Matched pairs designs

Completely randomized designs are the simplest statistical designs for experiments. They illustrate clearly the principles of control, randomization, and replication of treatments on a number of subjects. However, completely randomized designs are often inferior to more elaborate statistical designs. In particular, matching the subjects in various ways can produce more precise results than simple randomization.

matched pairs design

One common design that combines matching with randomization is the **matched pairs design.** A matched pairs design compares just two treatments. Choose pairs of subjects that are as closely matched as possible. Assign one of the treatments to each subject in a pair by tossing a coin or reading odd and even digits from Table B. Sometimes, each "pair" in a matched pairs design consists of just one subject, who gets both treatments one after the other. Each subject serves as his or her own control. The *order* of the treatments can influence the subject's response, so we randomize the order for each subject, again by a coin toss.

EXAMPLE 3.26 Matched Pairs for the Cell Phone Experiment

Example 3.22 (page 151) describes an experiment on the effects of talking on a cell phone while driving. The experiment compared two treatments: driving in a simulator and driving in a simulator while talking on a hands-free cell phone. The response variable is the time the driver takes to apply the brake when the car in front brakes suddenly. In Example 3.22, 40 student subjects were assigned at random, 20 students to each treatment. Subjects differ in driving skill and reaction times. The completely randomized design relies on chance to create two similar groups of subjects.

In fact, the experimenters used a matched pairs design in which all subjects drove under both conditions. They compared each subject's reaction times with and without the phone. If all subjects drove first with the phone and then without it, the effect of talking on the cell phone would be confounded with the fact that this is the first run in the simulator. The proper procedure requires that all subjects first be trained in using the simulator, that the *order* in which a subject drives with and without the phone be random, and that the two drives be on separate days to reduce the chance that the results of the second treatment will be affected by the first treatment.

The completely randomized design uses chance to decide which 20 subjects will drive with the cell phone. The other 20 drive without it. The matched pairs design uses chance to decide which 20 subjects will drive first with and then without the cell phone. The other 20 drive first without and then with the phone.

Block designs

Matched pairs designs apply the principles of comparison of treatments, randomization, and replication. However, the randomization is not complete—we do not randomly assign all the subjects at once to the two treatments. Instead, we only randomize within each matched pair. This allows matching to reduce the effect of variation among the subjects. Matched pairs are an example of *block designs.*

Block Design

A **block** is a group of subjects that are known before the experiment to be similar in some way expected to affect the response to the treatments. In a **block design,** the random assignment of individuals to treatments is carried out separately within each block.

A block design combines the idea of creating equivalent treatment groups by matching with the principle of forming treatment groups at random. Here is a typical example of a block design.

EXAMPLE 3.27 Men, Women, and Advertising

An experiment to compare the effectiveness of three television commercials for the same product will want to look separately at the reactions of men and women, as well as assess the overall response to the ads.

A completely randomized design considers all subjects, both men and women, as a single pool. The randomization assigns subjects to three treatment groups without regard to their gender. This ignores the differences between men and women. A better design considers women and men separately. Randomly assign the women to three groups, one to view each commercial. Then separately assign the men at random to three groups. Figure 3.9 outlines this improved design.

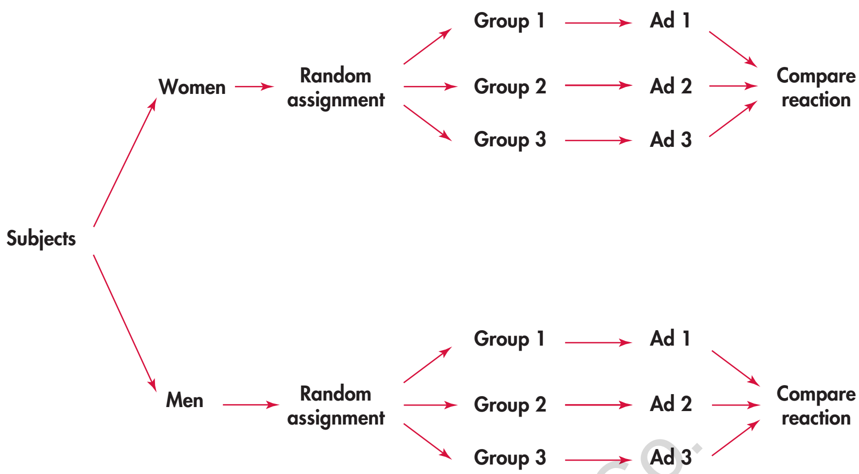


FIGURE 3.9 Outline of a block design, Example 3.27.

A block is a group of subjects formed before an experiment starts. We reserve the word "treatment" for a condition that we impose on the subjects. We don't speak of six treatments in Example 3.29 even though we can compare the responses of six groups of subjects formed by the two blocks (men, women) and the three commercials. Block designs are similar to stratified samples. Blocks and strata both group similar individuals together. We use two different names only because the idea developed separately for sampling and experiments.

Blocks are another form of *control*. They control the effects of some outside variables by bringing those variables into the experiment to form the blocks. The advantages of block designs are the same as the advantages of stratified samples. Blocks allow us to draw separate conclusions about each block—for example, about men and women in the advertising study in Example 3.27. Blocking also allows more precise overall conclusions because the systematic differences between men and women can be removed when we study the overall effects of the three commercials.

The idea of blocking is an important additional principle of statistical design of experiments. A wise experimenter will form blocks based on the most important unavoidable sources of variability among the experimental subjects. Randomization will then average out the effects of the remaining variation and allow an unbiased comparison of the treatments.

Like the design of samples, the design of complex experiments is a job for experts. Now that we have seen a bit of what is involved, we will usually just act as if most experiments were completely randomized.

APPLY YOUR KNOWLEDGE

3.55 Does charting help investors? Some investment advisers believe that charts of past trends in the prices of securities can help predict future prices. Most economists disagree. In an experiment to examine the effects of using charts, business students trade (hypothetically) a foreign currency at computer screens. There are 20 student subjects available, named for convenience A, B, C, . . . , T. Their goal is to make as much money as possible, and the best performances are rewarded with small prizes. The student traders have the price history of the foreign currency in dollars in their computers. They may or may not also have software that highlights trends. Describe two designs for this experiment—a completely randomized design and a matched pairs design in which each student serves as his or her own control. In both cases, carry out the randomization required by the design.

SECTION 3.3 Summary

- In an experiment, we impose one or more **treatments** on the **experimental units** or **subjects.** Each treatment is a combination of levels of the explanatory variables, which we call **factors.**
- The **design** of an experiment describes the choice of treatments and the manner in which the subjects are assigned to the treatments.
- The basic principles of statistical design of experiments are **control, randomization,** and **replication.**
- The simplest form of control is **comparison.** Experiments should compare two or more treatments in order to avoid **confounding** the effect of a treatment with other influences, such as lurking variables.
- **Randomization** uses chance to assign subjects to the treatments. Randomization creates treatment groups that are similar (except for chance variation) before the treatments are applied. Randomization and comparison together prevent **bias,** or systematic favoritism, in experiments.
- You can carry out randomization by giving numerical labels to the subjects and using a **table of random digits** to choose treatment groups.
- **Replication** of each treatment on many subjects reduces the role of chance variation and makes the experiment more sensitive to differences among the treatments.
- Good experiments require attention to detail as well as good statistical design. **Lack of realism** in an experiment can prevent us from generalizing its results.
- In addition to comparison, a second form of control is to restrict randomization by forming **blocks** of subjects that are similar in some way that is important to the response. Randomization is then carried out separately within each block.
- **Matched pairs** are a common form of blocking for comparing just two treatments. In some matched pairs designs, each subject receives both treatments in a random order. In others, the subjects are matched in pairs as closely as possible, and one subject in each pair receives each treatment.

SECTION 3.3 Exercises

For Exercises 3.42 to 3.44, see pages 144–145; for 3.45, see page 146; for 3.46 and 3.47, see page 147; for 3.48 and 3.49, see page 150; for 3.50 and 3.51, see page 152; for 3.52, see page 153; for 3.53 and 3.54, see page 154; and for 3.55, see page 156.

3.56 What is needed? Explain what is deficient in each of the following proposed experiments, and explain how you would improve the experiment.
(a) Two forms of a lab exercise are to be compared. There are 10 rows in the classroom. Students who sit in the first five rows of the class are given the first form, and students who sit in the last five rows are given the second form.
(b) The effectiveness of a leadership program for high school students is evaluated by examining the change in scores on a standardized test of leadership skills.
(c) An innovative method for teaching introductory biology courses is examined by using the traditional method in the fall zoology course and the new method in the spring botany course.

3.57 What is wrong? Explain what is wrong with each of the following randomization procedures, and describe how you would do the randomization correctly.
(a) A list of 50 subjects is entered into a computer file and then sorted by last name. The subjects are assigned to five

treatments by taking the first 10 subjects for Treatment 1, the next 10 subjects for Treatment 2, and so forth.
(b) Eight subjects are to be assigned to two treatments, four to each. For each subject, a coin is tossed. If the coin comes up heads, the subject is assigned to the first treatment; if the coin comes up tails, the subject is assigned to the second treatment.
(c) An experiment will assign 80 rats to four different treatment conditions. The rats arrive from the supplier in batches of 20, and the treatment lasts two weeks. The first batch of 20 rats is randomly assigned to one of the four treatments, and data for these rats are collected. After a one-week break, another batch of 20 rats arrives and is assigned to one of the three remaining treatments. The process continues until the last batch of rats is given the treatment that has not been assigned to the three previous batches.

3.58 Evaluate a new method for training new employees. A new method for training new employees is to be evaluated by randomly assigning new employees to either the current training program or the new method. A questionnaire will be used to evaluate the satisfaction of the new employees with the training. Explain how this experiment should be done in a double-blind fashion.

3.59 Can you change attitudes of workers about teamwork? You will conduct an experiment designed to change attitudes of workers about teamwork. Discuss some variables that you might use if you were to use a block design for this experiment.

3.60 An experiment for a new product. Compost tea is rich in microorganisms that help plants grow. It is made by soaking compost in water.[25] Design a comparative experiment that will provide evidence about whether or not compost tea works for a particular type of plant that interests you. Be sure to provide all details regarding your experiment, including the response variable or variables that you will measure. Assuming that the experiment shows positive results, write a short description about how you would use the results in a marketing campaign for compost tea.

3.61 Marketing your training materials. Water quality of streams and lakes is an issue of concern to the public. Although trained professionals typically are used to take reliable measurements, many volunteer groups are gathering and distributing information based on data that they collect.[26] You are part of a team to train volunteers to collect accurate water quality data. Design an experiment to evaluate the effectiveness of the training. Write a summary of your proposed design to present to your team. Be sure to include all the details that they will need to evaluate your proposal. How would you use the results of the experiment to market your training materials?

APPLET **3.62 Randomly assign the subjects.** You can use the *Simple Random Sample* applet to choose a treatment group at random once you have labeled the subjects. Example 3.22 (page 151) describes an experiment in which 20 students are chosen from a group of 40 for the treatment group in a study of the effect of cell phones on driving. Use the applet to choose the 20 students for the experimental group. Which students did you choose? The remaining 20 students make up the control group.

APPLET **3.63 Randomly assign the subjects.** The *Simple Random Sample* applet allows you to randomly assign experimental units to more than two groups without difficulty. Example 3.23 (page 152) describes a randomized comparative experiment in which 150 students are randomly assigned to six groups of 25.
(a) Use the applet to randomly choose 25 out of 150 students to form the first group. Which students are in this group?
(b) The "population hopper" now contains the 125 students that were not chosen, in scrambled order. Click "Sample" again to choose 25 of these remaining students to make up the second group. Which students were chosen?
(c) Click "Sample" three more times to choose the third, fourth, and fifth groups. Don't take the time to write down these groups. Check that there are only 25 students remaining in the "population hopper." These subjects get Treatment 6. Which students are they?

3.64 Random digits. Table B is a table of random digits. Which of the following statements are true of a table of random digits, and which are false? Explain your answers.
(a) Each pair of digits has chance 1/100 of being 50.
(b) There are exactly four 0s in each row of 40 digits.
(c) The digits 9999 can never appear as a group, because this pattern is not random.

3.65 I'll have a Mocha Light. Here's the opening of a press release: "Starbucks Corp. on Monday said it would roll out a line of blended coffee drinks intended to tap into the growing popularity of reduced-calorie and reduced-fat menu choices for Americans." You wonder if Starbucks customers like the new "Mocha Frappuccino Light" as well as the regular version of this drink.
(a) Describe a matched pairs design to answer this question. Be sure to include proper blinding of your subjects.
(b) You have 30 regular Starbucks customers on hand. Use software or Table B at line 151 to do the randomization that your design requires.

3.66 Price cuts on athletic shoes. Stores advertise price reductions to attract customers. What type of price cut is most attractive? Market researchers prepared ads for athletic shoes announcing different levels of discounts (20%, 40%, or 60%). The student subjects who read the ads were also given "inside information" about the fraction of shoes on sale (50% or 100%). Each subject then rated the attractiveness of the sale on a scale of 1 to 7.
(a) There are two factors. Make a sketch like Figure 3.4 (page 144) that displays the treatments formed by all combinations of levels of the factors.
(b) Outline a completely randomized design using 50 student subjects. Use software or Table B at line 121 to choose the subjects for the first treatment.

3.67 Effects of price promotions. A researcher is studying the effect of price promotions on consumers' expectations. She makes up a history of the store price of a hypothetical brand of laundry detergent for the past year. Students in a marketing course view the price history on a computer. Some students see a steady price, while others see regular promotions that temporarily cut the price. Then the students are asked what price they would expect to pay for the detergent.
(a) Is this study an experiment? Explain your answer.
(b) What are the explanatory and response variables?

3.68 Aspirin and heart attacks. "Nearly five decades of research now link aspirin to the prevention of stroke and heart attacks." So says the Bayer Aspirin website, **bayeraspirin.com**. The most important evidence for this claim comes from the Physicians' Health Study, a large medical experiment involving 22,000 male physicians. One group of about 11,000 physicians took an aspirin every second day, while the rest took a placebo. After several years, the study found that subjects in the aspirin group had significantly fewer heart attacks than subjects in the placebo group.
(a) Identify the experimental subjects, the factor and its levels, and the response variable in the Physicians' Health Study.
(b) Use a diagram to outline a completely randomized design for the Physicians' Health Study.
(c) What does it mean to say that the aspirin group had "significantly fewer heart attacks"?

3.69 Marketing to children. If children are given more choices within a class of products, will they tend to prefer that product to a competing product that offers fewer choices? Marketers want to know. An experiment prepared three sets of beverages. Set 1 contained two milk drinks and two fruit drinks. Set 2 had two fruit drinks and four milk drinks. Set 3 contained four fruit drinks but only two milk drinks. The researchers divided 120 children aged 4 to 12 years into three groups at random. They offered each group one of the sets. As each child chose a beverage to drink from the set presented, the researchers noted whether the choice was a milk drink or a fruit drink.
(a) What are the experimental subjects?
(b) What is the factor and what are its levels? What is the response variable?
(c) Use a diagram to outline a completely randomized design for the study.
(d) Explain how you would assign labels to the subjects. Use software to do the randomization or Table B at line 145 to choose the first five subjects assigned to the first treatment.

3.70 Effects of TV advertising. You decide to use a completely randomized design in the two-factor experiment on response to advertising described in Example 3.16 (page 143). The 30 students named below will serve as subjects. Outline the design. Then use software or Table B at line 110 to randomly assign the subjects to the six treatments. TVADS

Alomar	Denman	Han	Liang	Padilla	Valasco
Asihiro	Durr	Howard	Maldonado	Plochman	Vaughn
Bennett	Edwards	Hruska	Marsden	Rosen	Wei
Chao	Fleming	James	O'Brian	Trujillo	Willis
Clemente	George	Kaplan	Ogle	Tullock	Zhang

3.71 Temperature and work performance. An expert on worker performance is interested in the effect of room temperature on the performance of tasks requiring manual dexterity. She chooses temperatures of 20ºC (68ºF) and 30ºC (86ºF) as treatments. The response variable is the number of correct insertions, during a 30-minute period, in a peg-and-hole apparatus that requires the use of both hands simultaneously. Each subject is trained on the apparatus and is then asked to make as many insertions as possible in 30 minutes of continuous effort.
(a) Outline a completely randomized design to compare dexterity at 20ºC and 30ºC. Twenty subjects are available.
(b) Because people differ greatly in dexterity, the wide variation in individual scores may hide the systematic effect of temperature unless there are many subjects in each group. Describe in detail the design of a matched pairs experiment in which each subject serves as his or her own control.

3.4 Data Ethics

The production and use of data often involve ethical questions. We won't discuss the telemarketer who begins a telephone sales pitch with "I'm conducting a survey." Such deception is clearly unethical. It enrages legitimate survey organizations, which find the public less willing to talk with them. Neither will we discuss those few researchers who, in the pursuit of professional advancement, publish fake data. There is no ethical question here—faking data to advance your career is just wrong. But just how honest must researchers be about real, unfaked data? Here is an example that suggests the answer is "More honest than they often are."

EXAMPLE 3.28 Provide All the Critical Information

Papers reporting scientific research are supposed to be short, with no extra baggage. Brevity, however, can allow researchers to avoid complete honesty about their data. Did they choose their subjects in a biased way? Did they report data on only some of their subjects? Did they try several statistical analyses and report only the ones that looked best? The statistician John Bailar screened more than 4000 medical papers in more than a decade as consultant to the *New England Journal of Medicine.* He says, "When it came to the statistical review, it was often clear that critical information was lacking, and the gaps nearly always had the practical effect of making the authors' conclusions look stronger than they should have."[27] The situation is no doubt worse in fields that screen published work less carefully.

The most complex issues of data ethics arise when we collect data from people. The ethical difficulties are more severe for experiments that impose some treatment on people than for sample surveys that simply gather information. Trials of new medical treatments, for example, can do harm as well as good to their subjects. Here are some basic standards of data ethics that must be obeyed by any study that gathers data from human subjects, whether sample survey or experiment.

Basic Data Ethics

The organization that carries out the study must have an **institutional review board** that reviews all planned studies in advance in order to protect the subjects from possible harm.

All subjects in a study must give their **informed consent** before data are collected.

All subject data must be kept **confidential.** Only statistical summaries for groups of subjects may be made public.

The law requires that studies carried out or funded by the federal government obey these principles.[28] But neither the law nor the consensus of experts is completely clear about the details of their application.

Institutional review boards

The purpose of an institutional review board is not to decide whether a proposed study will produce valuable information or whether it is statistically sound. The board's purpose is, in the words of one university's board, "to protect the rights and welfare of human subjects (including patients) recruited to participate in research activities." The board reviews the plan of the study and can require changes. It reviews the consent form to ensure that subjects are informed about the nature of the study and about any potential risks. Once research begins, the board monitors the study's progress at least once a year.

The most pressing issue concerning institutional review boards is whether their workload has become so large that their effectiveness in protecting subjects drops. When the government temporarily stopped human-subject research at Duke University Medical Center in 1999 due to inadequate protection of subjects, more than 2000 studies were going on. That's a lot of review work. There are shorter review procedures for projects that involve only minimal risks to subjects, such as most sample surveys. When a board is overloaded, there is a temptation to put more proposals in the minimal-risk category to speed the work.

APPLY YOUR KNOWLEDGE

The exercises in this section on ethics are designed to help you think about the issues that we are discussing and to formulate some opinions. In general, there are no wrong or right answers but you need to give reasons for your answers.

3.72 Who should be on an institutional review board? Government regulations require that institutional review boards consist of at least five people, including at least one scientist, one nonscientist, and one person from outside the institution. Most boards are larger, but many contain just one outsider.

(a) Why should review boards contain people who are not scientists?
(b) Do you think that one outside member is enough? How would you choose that member? (For example, would you prefer a medical doctor? A member of the clergy? An activist for patients' rights?)

3.73 Do these proposals involve minimal risk? You are a member of your college's institutional review board. You must decide whether several research proposals qualify for lighter review because they involve only minimal risk to subjects. Federal regulations say that "minimal risk" means the risks are no greater than "those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests." That's vague. Which of these do you think qualifies as "minimal risk"?

(a) Draw a drop of blood by pricking a finger in order to measure blood sugar.
(b) Draw blood from the arm for a full set of blood tests.
(c) Insert a tube that remains in the arm, so that blood can be drawn regularly.

Informed consent

Both words in the phrase "informed consent" are important, and both can be controversial. Subjects must be *informed* in advance about the nature of a study and any risk of harm it may bring. In the case of a sample survey, physical harm is not possible. The subjects should be told what kinds of questions the survey will ask and about how much of their time it will take. Experimenters must tell subjects the nature and purpose of the study and outline possible risks. Subjects must then *consent* in writing.

EXAMPLE 3.29 Who Can Give Informed Consent?

Are there some subjects who can't give informed consent? It was once common, for example, to test new vaccines on prison inmates who gave their consent in return for good-behavior credit. Now, we worry that prisoners are not really free to refuse, and the law forbids almost all medical research in prisons.

Children can't give fully informed consent, so the usual procedure is to ask their parents. A study of new ways to teach reading is about to start at a local elementary school, so the study team sends consent forms home to parents. Many parents don't return the forms. Can their children take part in the study because the parents did not say "No," or should we allow only children whose parents returned the form and said "Yes"?

What about research into new medical treatments for people with mental disorders? What about studies of new ways to help emergency room patients who may be unconscious? In most cases, there is not time to get the consent of the family. Does the principle of informed consent bar realistic trials of new treatments for unconscious patients?

These are questions without clear answers. Reasonable people differ strongly on all of them. There is nothing simple about informed consent.[29]

The difficulties of informed consent do not vanish even for capable subjects. Some researchers, especially in medical trials, regard consent as a barrier to getting patients to participate in research. They may not explain all possible risks; they may not point out that there are other therapies that might be better than those being studied; they may be too optimistic in talking with patients even when the consent form has all the right details. On the other hand, mentioning every possible risk leads to very long consent forms that really are barriers. "They are like rental car contracts," one lawyer said. Some subjects don't read forms that run five or six printed pages. Others are frightened by the large number of possible (but unlikely) disasters that might happen and so refuse to participate. Of course, unlikely disasters sometimes happen. When they do, lawsuits follow—and the consent forms become yet longer and more detailed.

Confidentiality

Ethical problems do not disappear once a study has been cleared by the review board, has obtained consent from its subjects, and has actually collected data about the subjects. It is important to protect the subjects' privacy by keeping all data about subjects confidential. The report of an opinion poll may say what percent of the 1200 respondents felt that legal immigration should be reduced. It may not report what *you* said about this or any other issue.

anonymity

Confidentiality is not the same as **anonymity.** Anonymity means that subjects are anonymous—their names are not known even to the director of the study. Anonymity is rare in statistical studies. Even where it is possible (mainly in surveys conducted by mail), anonymity prevents any follow-up to improve nonresponse or inform subjects of results.

Any breach of confidentiality is a serious violation of data ethics. The best practice is to separate the identity of the subjects from the rest of the data at once. Sample surveys, for example, use the identification only to check on who did or did not respond. In an era of advanced technology, however, it is no longer enough to be sure that each set of data protects people's privacy. The government, for example, maintains a vast amount of information about citizens in many separate databases—census responses, tax returns, Social Security information, data from surveys such as the Current Population Survey, and so on. Many of these databases can be searched by computers for statistical studies. A clever computer search of several databases might be able, by combining information, to identify you and learn a great deal about you even if your name and other identification have been removed from the data available for search. A colleague from Germany once remarked that "female full professor of statistics with PhD from the United States" was enough to identify her among all the 83 million residents of Germany. Privacy and confidentiality of data are hot issues among statisticians in the computer age.

EXAMPLE 3.30 Data Collected by the Government

Citizens are required to give information to the government. Think of tax returns and Social Security contributions. The government needs these data for administrative purposes—to see if we paid the right amount of tax and how large a Social Security benefit we are owed when we retire. Some people feel that people should be able to forbid any other use of their data, even with all identification removed. This would prevent using government records to study, say, the ages, incomes, and household sizes of Social Security recipients. Such a study could well be vital to debates on reforming Social Security.

APPLY YOUR KNOWLEDGE

3.74 Should we allow this personal information to be collected? In which of the following circumstances would you allow collecting personal information without the subjects' consent?

(a) A government agency takes a random sample of income tax returns to obtain information on the average income of people in different occupations. Only the incomes and occupations are recorded from the returns, not the names.
(b) A social psychologist attends public meetings of a religious group to study the behavior patterns of members.
(c) A social psychologist pretends to be converted to membership in a religious group and attends private meetings to study the behavior patterns of members.

3.75 How can we obtain informed consent? A researcher suspects that traditional religious beliefs tend to be associated with an authoritarian personality. She prepares a questionnaire that measures authoritarian tendencies and also asks many religious questions. Write a description of the purpose of this research to be read by subjects in order to obtain their informed consent. You must balance the conflicting goals of not deceiving the subjects as to what the questionnaire will tell about them and of not biasing the sample by scaring off religious people.

Clinical trials

Clinical trials are experiments that study the effectiveness of medical treatments on actual patients. Medical treatments can harm as well as heal, so clinical trials spotlight the ethical problems of experiments with human subjects. Here are the starting points for a discussion:

- Randomized comparative experiments are the only way to see the true effects of new treatments. Without them, risky treatments that are no more effective than placebos will become common.
- Clinical trials produce great benefits, but most of these benefits go to future patients. The trials also pose risks, and these risks are borne by the subjects of the trial. So we must balance future benefits against present risks.
- Both medical ethics and international human rights standards say that "the interests of the subject must always prevail over the interests of science and society."

The quoted words are from the 1964 Helsinki Declaration of the World Medical Association, the most respected international standard. The most outrageous examples of unethical experiments are those that ignore the interests of the subjects.

EXAMPLE 3.31 The Tuskegee Study

In the 1930s, syphilis was common among black men in the rural South, a group that had almost no access to medical care. The Public Health Service Tuskegee study recruited 399 poor black sharecroppers with syphilis and 201 others without the disease in order to observe how syphilis progressed when no treatment was given. Beginning in 1943, penicillin became available to treat syphilis. The study subjects were not treated. In fact, the Public Health Service prevented any treatment until word leaked out and forced an end to the study in the 1970s.

The Tuskegee study is an extreme example of investigators following their own interests and ignoring the well-being of their subjects. A 1996 review said, "It has come to symbolize racism in medicine, ethical misconduct in human research, paternalism by physicians, and government abuse of vulnerable people." In 1997, President Clinton formally apologized to the surviving participants in a White House ceremony.[30]

Because "the interests of the subject must always prevail," medical treatments can be tested in clinical trials only when there is reason to hope that they will help the patients who are subjects in the trials. Future benefits aren't enough to justify experiments with human subjects. Of course, if there is already strong evidence that a treatment works and is safe, it is unethical *not* to give it. Here are the words of Dr. Charles Hennekens of the Harvard Medical School, who directed the large clinical trial that showed that aspirin reduces the risk of heart attacks:

> *There's a delicate balance between when to do or not do a randomized trial. On the one hand, there must be sufficient belief in the agent's potential to justify exposing half the subjects to it. On the other hand, there must be sufficient doubt about its efficacy to justify withholding it from the other half of subjects who might be assigned to placebos.*[31]

Why is it ethical to give a control group of patients a placebo? Well, we know that placebos often work. Moreover, placebos have no harmful side effects. So in the state of balanced doubt described by Dr. Hennekens, the placebo group may be getting a better treatment than the drug group. If we *knew* which treatment was better, we would give it to everyone. When we don't know, it is ethical to try both and compare them.

The idea of using a control or a placebo is a fundamental principle to be considered in designing experiments. In many situations, deciding what to use as an appropriate control requires some careful thought. The choice of the control can have a substantial impact on the conclusions drawn from an experiment. Here is an example.

EXAMPLE 3.32 Was the Claim Misleading?

The manufacturer of a breakfast cereal designed for children claims that eating this cereal has been clinically shown to improve attentiveness by nearly 20%. The study used two groups of children who were tested before and after breakfast. One group received the cereal for breakfast, while breakfast for the control group was water. The results of tests taken three hours after breakfast were used to make the claim.

The Federal Trade Commission investigated the marketing of this product. They charged that the claim was false and violated federal law. The charges were settled, and the company agreed to not use misleading claims in its advertising.[32]

It is not sufficient to obtain appropriate controls. The data from all groups must be collected and analyzed in the same way. Here is an example of this type of flawed design.

EXAMPLE 3.33 The Product Doesn't Work!

Two scientists published a paper claiming to have developed a very exciting new method to detect ovarian cancer using blood samples. The potential market for such a procedure is substantial, and there is no specific screening test currently available. When other scientists were unable to reproduce the results in different labs, the original work was examined more carefully. The original study used blood samples from women with ovarian cancer and from healthy controls. The blood samples were all analyzed using a mass spectrometer. The control samples were analyzed on one day and the cancer samples were analyzed on the next day. This design was flawed because it could not control for changes over time in the measuring instrument.[33]

APPLY YOUR KNOWLEDGE

3.76 Should the treatments be given to everyone? Effective drugs for treating AIDS are very expensive, so most African nations cannot afford to give them to large numbers of people. Yet AIDS is more common in parts of Africa than anywhere else. Several clinical trials being conducted in Africa are looking at ways to prevent pregnant mothers infected with HIV from passing the infection to their unborn children, a major source of HIV infections in Africa. Some people say these trials are unethical because they do not give effective AIDS drugs to their subjects, as would be required in rich nations. Others reply that the trials are looking for treatments that can work in the real world in Africa and that they promise benefits at least to the children of their subjects. What do you think?

3.77 Is this study ethical? Researchers on aging proposed to investigate the effect of supplemental health services on the quality of life of older people. Eligible patients of a large medical clinic were to be randomly assigned to treatment and control groups. The treatment group would be offered hearing aids, dentures, transportation, and other services not available without charge to the control group. The review board felt that providing these services to some but not other persons in the same institution raised ethical questions. Do you agree?

Behavioral and social science experiments

When we move from medicine to the behavioral and social sciences, the direct risks to experimental subjects are less acute, but so are the possible benefits to the subjects. Consider, for example, the experiments conducted by psychologists in their study of human behavior.

EXAMPLE 3.34 Personal Space

Psychologists observe that people have a "personal space" and are uneasy if others come too close to them. We don't like strangers to sit at our table in a coffee shop if other tables are available, and we see people move apart in elevators if there is room to do so. Americans tend to require more personal space than people in most other cultures. Can violations of personal space have physical, as well as emotional, effects?

Investigators set up shop in a men's public restroom. They blocked off urinals to force men walking in to use either a urinal next to an experimenter (treatment group) or a urinal separated from the experimenter (control group). Another experimenter, using a periscope from a toilet stall, measured how long the subject took to start urinating and how long he continued.[34]

This personal space experiment illustrates the difficulties facing those who plan and review behavioral studies:

- There is no risk of harm to the subjects, although they would certainly object to being watched through a periscope. Even when physical harm is unlikely, are there other types of harm that need to be considered? Emotional harm? Undignified situations? Invasion of privacy?
- What about informed consent? The subjects did not even know they were participating in an experiment. Many behavioral experiments rely on hiding the true purpose of the study. The subjects would change their behavior if told in advance what the investigators were studying. Subjects are asked to consent on the basis of vague information. They receive full information only after the experiment.

The "Ethical Principles" of the American Psychological Association require consent unless a study merely observes behavior in a public place. They allow deception only when it is necessary to the study, does not hide information that might influence a subject's willingness to participate, and is explained to subjects as soon as possible. The personal space study (from the 1970s) does not meet current ethical standards.

We see that the basic requirement for informed consent is understood differently in medicine and psychology. Here is an example of another setting with yet another interpretation of what is ethical. The subjects get no information and give no consent. They don't even know that an experiment may be sending them to jail for the night.

EXAMPLE 3.35 Reducing Domestic Violence

How should police respond to domestic-violence calls? In the past, the usual practice was to remove the offender and order the offender to stay out of the household overnight. Police were reluctant to make arrests because the victims rarely pressed charges. Women's groups argued that arresting offenders would help prevent future violence even if no charges were filed. Is there evidence that arrest will reduce future offenses? That's a question that experiments have tried to answer.

A typical domestic-violence experiment compares two treatments: arrest the suspect and hold the suspect overnight or warn and release the suspect. When police officers reach the scene of a domestic-violence call, they calm the participants and investigate. Weapons or death threats require an arrest. If the facts permit an arrest but do not require it, an officer radios headquarters for instructions. The person on duty opens the next envelope in a file prepared in advance by a statistician. The envelopes contain the treatments in random order. The police either make an arrest or warn and release, depending on the contents of the envelope. The researchers then watch police records and visit the victim to see if the domestic violence reoccurs.

Such experiments show that arresting domestic-violence suspects does reduce their future violent behavior.[35] As a result of this evidence, arrest has become the common police response to domestic violence.

The domestic-violence experiments shed light on an important issue of public policy. Because there is no informed consent, the ethical rules that govern clinical trials and most social science studies would forbid these experiments. They were cleared by review boards because, in the words of one domestic-violence researcher, "These people became subjects by committing acts that allow the police to arrest them. You don't need consent to arrest someone."

SECTION 3.4 Summary

- The purpose of an **institutional review board** is to protect the rights and welfare of the human subjects in a study. Institutional review boards review **informed consent** forms that subjects will sign before participating in a study.
- Information about subjects in a study must be kept **confidential,** but statistical summaries of groups of subjects may be made public.
- **Clinical trials** are experiments that study the effectiveness of medical treatments on actual patients.
- Some studies in the **behavioral** and **social sciences** are observational, while others are designed experiments.

SECTION 3.4 Exercises

For Exercises 3.72 and 3.73, see page 161; for 3.74 and 3.75, see page 163; and for 3.76 and 3.77, see page 165.

Most of these exercises pose issues for discussion. There are no right or wrong answers, but there are more and less thoughtful answers.

3.78 How should the samples been analyzed? Refer to the ovarian cancer diagnostic test study in Example 3.33 (page 165). Describe how you would process the samples through the mass spectrometer.

3.79 The Vytorin controversy. Vytorin is a combination pill designed to lower cholesterol. The combination consists of a relatively inexpensive and widely used drug, Zocor, and a newer drug called Zetia. Early study results suggested that Vytorin was no more effective than Zetia. Critics claimed that the makers of the drugs tried to change the response variable for the study, and two congressional panels investigated why there was a two-year delay in the release of the results. Use the Web to search for more information about this controversy, and write a report of what you find. Include an evaluation in the framework of ethical use of experiments and data. A good place to start your search would be to look for the phrase "Vytorin's shortcomings."

3.80 Facebook and academic performance. *First Monday* is a peer-reviewed journal on the Internet. It recently published two articles concerning Facebook and academic performance. Visit its website, **firstmonday.org**, and look at the first three articles in Volume 14, Number 5–4, May 2009. Identify the key controversial issues that involve the use of statistics in these articles, and write a report summarizing the facts as you see them. Be sure to include your opinions regarding ethical issues related to this work.

3.81 Anonymity and confidentiality in mail surveys. Some common practices may appear to offer anonymity while actually delivering only confidentiality. Market researchers often use mail surveys that do not ask the respondent's identity but contain hidden codes on the questionnaire that identify the respondent. A false claim of anonymity is clearly unethical. If only confidentiality is promised, is it also unethical to say nothing about the identifying code, perhaps causing respondents to believe their replies are anonymous?

3.82 Studying your blood. Long ago, doctors drew a blood specimen from you when you were treated for anemia. Unknown to you, the sample was stored. Now researchers plan to use stored samples from you and many other people to look for genetic factors that may influence anemia. It is no longer possible to ask your consent. Modern technology can read your entire genetic makeup from the blood sample.
(a) Do you think it violates the principle of informed consent to use your blood sample if your name is on it but you were not told that it might be saved and studied later?
(b) Suppose that your identity is not attached. The blood sample is known only to come from (say) "a 20-year-old white female being treated for anemia." Is it now ethical to use the sample for research?
(c) Perhaps we should use biological materials such as blood samples only from patients who have agreed to allow the material to be stored for later use in research. It isn't possible to say in advance what kind of research, so this falls short of the usual standard for informed consent. Is it acceptable, given complete confidentiality and the fact that using the sample can't physically harm the patient?

3.83 Anonymous? Confidential? One of the most important nongovernment surveys in the United States is the National Opinion Research Center's General Social Survey. The GSS regularly monitors public opinion on a wide variety of political and social issues. Interviews are conducted in person in the subject's home. Are a subject's responses to GSS questions anonymous, confidential, or both? Explain your answer.

3.84 Anonymous? Confidential? Texas A&M, like many universities, offers free screening for HIV, the virus that causes AIDS. The announcement says, "Persons who sign up for the HIV screening will be assigned a number so that they do not have to give their name." They can learn the results of the test by telephone, still without giving their name. Does this practice offer *anonymity* or just *confidentiality?*

3.85 Political polls. Candidates for public office hire polling organizations to take sample surveys to find out what the voters think about the issues. What information should the pollsters be required to disclose?
(a) What does the standard of informed consent require the pollsters to tell potential respondents?
(b) Should polling organizations be required to give respondents the name and address of the organization that carries out the poll?
(c) The polling organization usually has a professional name such as "Samples Incorporated," so respondents don't know that the poll is being paid for by a political party or candidate. Would revealing the sponsor to respondents bias the poll? Should the sponsor always be announced whenever poll results are made public?

3.86 Making poll results public. Some people think that the law should require that all political poll results be made public. Otherwise, the possessors of poll results can use the information to their own advantage. They can act on the information, release only selected parts of it, or time the release for best effect. A candidate's organization replies that it is paying for the poll in order to gain information for its own use, not to amuse the public. Do you favor requiring complete disclosure of political poll results? What about other private surveys, such as market research surveys of consumer tastes?

3.87 Student subjects. Students taking Psychology 001 are required to serve as experimental subjects. Students in Psychology 002 are not required to serve, but they are given extra credit if they do so. Students in Psychology 003 are required either to sign up as subjects or to write a term paper. Serving as an experimental subject may be educational, but current ethical standards frown on using "dependent subjects" such as prisoners or charity medical patients. Students are certainly somewhat dependent on their teachers. Do you object to any of these course policies? If so, which ones, and why?

3.88 How many have HIV? Researchers from Yale, working with medical teams in Tanzania, wanted to know how common infection with HIV (the virus that causes AIDS) is among pregnant women in that African country. To do this, they planned to test blood samples drawn from pregnant women.

Yale's institutional review board insisted that the researchers get the informed consent of each woman and tell her the results of the test. This is the usual procedure in developed nations. The Tanzanian government did not want to tell the women why blood was drawn or tell them the test results. The government feared panic if many people turned out to have an incurable disease for which the country's medical system could not provide care. The study was canceled. Do you think that Yale was right to apply its usual standards for protecting subjects?

3.89 AIDS trials in Africa. One of the most important goals of AIDS research is to find a vaccine that will protect against HIV infection. Because AIDS is so common in parts of Africa, that is the easiest place to test a vaccine. It is likely, however, that a vaccine would be so expensive that it could not (at least at first) be widely used in Africa. Is it ethical to test in Africa if the benefits go mainly to rich countries? The treatment group of subjects would get the vaccine, and the placebo group would later be given the vaccine if it proved effective. So the actual subjects would benefit and the future benefits then would go elsewhere. What do you think?

3.90 Asking teens about sex. The Centers for Disease Control and Prevention, in a survey of teenagers, asked the subjects if they were sexually active. Those who said "Yes" were then asked, "How old were you when you had sexual intercourse for the first time?" Should consent of parents be required to ask minors about sex, drugs, and other such issues, or is consent of the minors themselves enough? Give reasons for your opinion.

3.91 Deceiving subjects. Students sign up to be subjects in a psychology experiment. When they arrive, they are told that interviews are running late and are taken to a waiting room. The experimenters then stage a theft of a valuable object left in the waiting room. Some subjects are alone with the thief, and others are in pairs—these are the treatments being compared.

Will the subject report the theft? The students had agreed to take part in an unspecified study, and the true nature of the experiment is explained to them afterward. Do you think this study is ethical?

3.92 Deceiving subjects. A psychologist conducts the following experiment: she measures the attitude of subjects toward cheating, then has them play a game rigged so that winning without cheating is impossible. The computer that organizes the game also records—unknown to the subjects—whether or not they cheat. Then attitude toward cheating is retested.

Subjects who cheat tend to change their attitudes to find cheating more acceptable. Those who resist the temptation to cheat tend to condemn cheating more strongly on the second test of attitude. These results confirm the psychologist's theory.

This experiment tempts subjects to cheat. The subjects are led to believe that they can cheat secretly when, in fact, they are observed. Is this experiment ethically objectionable? Explain your position.

3.93 What is wrong? Explain what is wrong in each of the following scenarios.
(a) Clinical trials are always ethical as long as they randomly assign patients to the treatments.
(b) The job of an institutional review board is complete when they decide to allow a study to be conducted.
(c) A treatment that has no risk of physical harm to subjects is always ethical.

CHAPTER 3 Review Exercises

3.94 Online behavioral advertising. The Federal Trade Commission Staff Report "Self-Regulatory Principles for Online Behavioral Advertising" defines behavioral advertising as "the tracking of a consumer's online activities over time—including the searches the consumer has conducted, the webpages visited and the content viewed—to deliver advertising targeted to the individual consumer's interests." The report suggests four governing concepts for online behavioral advertising:
1. Transparency and control: when companies collect information from consumers for advertising, they should tell the consumers about how the data will be collected, and consumers should be given a choice about whether to allow the data to be collected.
2. Security and data retention: data should be kept secure and should be retained only as long as needed.
3. Privacy: before data are used in a way that differs from how the companies originally said they would use the information, companies should obtain consent from consumers.
4. Sensitive data: consent should be obtained before using any sensitive data.[36]
Write a report discussing your opinions concerning online behavioral advertising and the four governing concepts. Pay particular attention to issues related to the ethical collection and use of statistical data.

3.95 Confidentiality at NORC. The National Opinion Research Center conducts a large number of surveys and has established procedures for protecting the confidentiality of their survey participants. For its Survey of Consumer Finances, it provides a pledge to participants regarding confidentiality. This pledge is available at **scf.norc.org/Confidentiality.html.** Review the pledge and summarize its key parts. Do you think that the pledge adequately addresses issues related to the ethical collection and use of data? Explain your answer.

3.96 What's wrong? Explain what is wrong in each of the following statements. Give reasons for your answers.
(a) A simple random sample was used to assign a group of 30 subjects to three treatments.
(b) It is better to use a table of random numbers to select a simple sample than it is to use a computer.
(c) Matched pairs designs and block designs are complicated and should be avoided if possible.

3.97 Price promotions and consumer behavior. A researcher is studying the effect of price promotions on consumer behavior. Subjects are asked to choose between purchasing a single case of a soft drink for $4.00 or three cases of the same soft drink for $10.00. Is this study an experiment? Why? What are the explanatory and response variables?

3.98 What type of study? What is the best way to answer each of the following questions: an experiment, a sample survey, or an observational study that is not a sample survey? Explain your choices.
(a) Are people generally satisfied with the service they receive from a customer call center?
(b) Do new employees learn basic facts about your company better in a workshop or using an online set of materials?

(c) How long do your customers have to wait to resolve a problem with a new purchase?

3.99 Choose the type of study. Give an example of a question about your customers, their behavior, or their opinions that would best be answered by
(a) a sample survey.
(b) an observational study that is not a sample survey.
(c) an experiment.

3.100 Compare Pizza Hut with Domino's. Do consumers prefer pizza from Pizza Hut or from Domino's? Discuss how you might make this a blind test in which neither source of the pizza is identified. Do you think that your blinding will be successful for all subjects? Describe briefly the design of a matched pairs experiment to investigate this question. How will you use randomization?

3.101 Coupons and customer expectations. A researcher studying the effect of coupons on consumers' expectations makes up two different series of ads for a hypothetical brand of cola for the past year. Students in a family science course view one or the other sequence of ads on a computer. Some students see a sequence of ads with no coupon offered on the cola, while others see regular coupon offerings that effectively lower the price of the cola temporarily. Next, the students are asked what price they would expect to pay for the cola.
(a) Is this study an experiment? Why?
(b) What are the explanatory and response variables?

3.102 Can you remember how many? An opinion poll calls 2200 randomly chosen residential telephone numbers, and 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, 1435 people respond. What is the rate (percent) of nonresponse?
(c) For the question asked, what source of response error is likely present?
(d) Write a variation on this question that would reduce the associated response error.

3.103 Marketing a dietary supplement. Your company produces a dietary supplement that contains a significant amount of calcium as one of its ingredients. The company would like to be able to market this fact successfully to one of the target groups for the supplement: men with high blood pressure. To this end, you must design an experiment to demonstrate that added calcium in the diet reduces blood pressure. You have available 30 men with high blood pressure who are willing to serve as subjects. CALSUPP
(a) Outline an appropriate design for the experiment, taking the placebo effect into account.
(b) The names of the subjects appear below. Do the randomization required by your design, and list the subjects to whom you will give the drug. (If you use Table B, enter the table at line 136.)

Alomar	Denman	Han	Liang	Rosen
Asihiro	Durr	Howard	Maldonado	Solomon
Bikalis	Farouk	Imrani	Moore	Townsend
Chen	Fratianna	James	O'Brian	Tullock
Cranston	Green	Krushchev	Plochman	Willis
Curtis	Guillen	Lawless	Rodriguez	Zhang

3.104 A hot fund. A large mutual funds group assigns a young securities analyst to manage its small biotechnology stock fund. The fund's share value increases an impressive 43% during the first year under the new manager. Explain why this performance does not necessarily establish the manager's ability.

3.105 Employee meditation. You see a news report of an experiment that claims to show that a meditation technique increased job satisfaction of employees. The experimenter interviewed the employees and assessed their levels of job satisfaction. The subjects then learned how to meditate and did so regularly for a month. The experimenter reinterviewed them at the end of the month and assessed their job satisfaction levels again.
(a) There was no control group in this experiment. Why is this a blunder? What lurking variables might be confounded with the effect of meditation?
(b) The experimenter who diagnosed the effect of the treatment knew that the subjects had been meditating. Explain how this knowledge could bias the experimental conclusions.
(c) Briefly discuss a proper experimental design, with controls and blind diagnosis, to assess the effect of meditation on job satisfaction.

3.106 Executives and exercise. A study of the relationship between physical fitness and leadership uses as subjects middle-aged executives who have volunteered for an exercise program. The executives are divided into a low-fitness group and a high-fitness group on the basis of a physical examination. All subjects then take a psychological test designed to measure leadership, and the results for the two groups are compared. Is this an observational study or an experiment? Explain your answer.

3.107 Does the new product taste better? Before a new variety of frozen muffins is put on the market, it is subjected to extensive taste testing. People are asked to taste the new muffin and a competing brand and to say which they prefer. (Both muffins are unidentified in the test.) Is this an observational study or an experiment? Why?

3.108 Questions about attitudes. Write two questions about an attitude that concerns you for use in a sample survey. Make the first question so that it is biased in one direction, and make the second question biased in the opposite direction. Explain why your questions are biased, and then write a third question that has little or no bias.

3.109 Will the regulation make the product safer? Canada requires that cars be equipped with "daytime running lights," headlights that automatically come on at a low level when the car is started. Some manufacturers are now equipping cars sold in the United States with running lights. Will running lights reduce accidents by making cars more visible?
(a) Briefly discuss the design of an experiment to help answer this question. In particular, what response variables will you examine?
(b) Example 3.25 (pages 153–154) discusses center brake lights. What cautions do you draw from that example that apply to an experiment on the effects of running lights?

3.110 Learning about markets. Your economics professor wonders if playing market games online will help students understand how markets set prices. You suggest an experiment: have some students use the online games, while others discuss markets in recitation sections. The course has two lectures, at 8:30 A.M. and 2:30 P.M. There are 11 recitation sections attached to each lecture. The students are already assigned to recitations. For practical reasons, all students in each recitation must follow the same program.
(a) The professor says, "Let's just have the 8:30 group do online work in recitation and the 2:30 group do discussion." Why is this a bad idea?
(b) Outline the design of an experiment with the 22 recitation sections as cases. Carry out your randomization, and include in your outline the recitation numbers assigned to each treatment.

3.111 How much do students earn? A university's financial aid office wants to know how much it can expect students to earn from summer employment. This information will be used to set the level of financial aid. The population contains 3478 students who have completed at least one year of study but have not yet graduated. The university will send a questionnaire to an SRS of 100 of these students, drawn from an alphabetized list.
(a) Describe how you will label the students in order to select the sample.
(b) Use Table B, beginning at line 120, to select the first eight students in the sample.

3.112 Attitudes toward collective bargaining. A labor organization wants to study the attitudes of college faculty members toward collective bargaining. These attitudes appear to be different depending on the type of college. The American Association of University Professors classifies colleges as follows:
Class I. Offer doctorate degrees and award at least 15 per year.
Class IIA. Award degrees above the bachelor's but are not in Class I.
Class IIB. Award no degrees beyond the bachelor's.
Class III. Two-year colleges.
Discuss the design of a sample of faculty from colleges in your state, with total sample size about 200.

3.113 Student attitudes concerning labor practices. You want to investigate the attitudes of students at your school about the labor practices of factories that make college-brand apparel. You have a grant that will pay the costs of contacting about 500 students.
(a) Specify the exact population for your study. For example, will you include part-time students?
(b) Describe your sample design. Will you use a stratified sample?
(c) Briefly discuss the practical difficulties that you anticipate. For example, how will you contact the students in your sample?

3.114 Treating drunk drivers. Once a person has been convicted of drunk driving, one purpose of court-mandated treatment or punishment is to prevent future offenses of the same kind. Suggest three different treatments that a court might require. Then outline the design of an experiment to compare their effectiveness. Be sure to specify the response variables you will measure.

3.115 Experiments and surveys for business. Write a short report describing the differences and similarities between experiments and surveys that would be used in business. Include a discussion of the advantages and disadvantages of each.

3.116 The product should not be discolored. Few people want to eat discolored french fries. Potatoes are kept refrigerated before being cut for french fries to prevent spoiling and preserve flavor. But immediate processing of cold potatoes causes discoloring due to complex chemical reactions. The potatoes must, therefore, be brought to room temperature before processing. Fast-food chains and other sellers of french fries must understand potato behavior. Design an experiment in which tasters will rate the color and flavor of french fries prepared from several groups of potatoes. The potatoes will be freshly harvested, stored for a month at room temperature, or stored for a month refrigerated. They will then be sliced and cooked either immediately or after an hour at room temperature.

(a) What are the factors and their levels, the treatments, and the response variables?

(b) Describe and outline the design of this experiment.

(c) It is efficient to have each taster rate fries from all treatments. How will you use randomization in presenting fries to the tasters?

3.117 Quality of service. Statistical studies can often help service providers assess the quality of their service. The U.S. Postal Service is one such provider of services. We wonder if the number of days a letter takes to reach another city is affected by the time of day it is mailed and whether or not the zip code is used. Describe briefly the design of a two-factor experiment to investigate this question. Be sure to specify the treatments exactly and to tell how you will handle lurking variables such as the day of the week on which the letter is mailed.

3.118 Mac versus PC. Many people hold very strong opinions about the superiority of the computer they use. Design an experiment to compare customer satisfaction with the Mac versus the PC. Consider whether or not you will include subjects who routinely use both types of computers and whether or not you will block on the type of computer currently being used. Write a summary of your design, including your reasons for the choices you make. Be sure to include the question or questions that you will use to measure customer satisfaction.

3.119 Design your own experiment. The previous two exercises illustrate the use of statistically designed experiments to answer questions of interest to consumers as well as to businesses. Select a question of interest to you that an experiment might answer, and briefly discuss the design of an appropriate experiment.

3.120 Randomization for testing a breakfast food. To demonstrate how randomization reduces confounding, return to the breakfast food testing experiment described in Example 3.18 (page 146). Label the 30 rats 01 to 30. Suppose that, unknown to the experimenter, the 10 rats labeled 01 to 10 have a genetic defect that will cause them to grow more slowly than normal rats. If the experimenter simply puts rats 01 to 15 in the experimental group and rats 16 to 30 in the control group, this lurking variable will bias the experiment against the new food product.

Use software or Table B to assign 15 rats at random to the experimental group as in Example 3.20. Record how many of the 10 rats with genetic defects are placed in the experimental group and how many are in the control group. Repeat the randomization using different lines in Table B until you have done five random assignments. What is the mean number of genetically defective rats in experimental and control groups in your five repetitions?

3.121 Two ways to ask sensitive questions. Sample survey questions are usually read from a computer screen. In a computer-aided personal interview (CAPI), the interviewer reads the questions and enters the responses. In a computer-aided self interview (CASI), the interviewer stands aside and the respondent reads the questions and enters responses. One method almost always shows a higher percent of subjects admitting use of illegal drugs. Which method? Explain why.

3.122 Your institutional review board. Your college or university has an institutional review board that screens all studies that use human subjects. Get a copy of the document that describes this board (you can probably find it online).

(a) According to this document, what are the duties of the board?

(b) How are members of the board chosen? How many members are not scientists? How many members are not employees of the college? Do these members have some special expertise, or are they simply members of the "general public"?