Producing Data: Experiments

A study is an experiment when we actually do something to people, animals, or objects in order to observe the response. 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 essential.

Experiments

Here is the basic vocabulary of experiments.

SUBJECTS, FACTORS, TREATMENTS

The individuals studied in an experiment are often called subjects, particularly when they are people.

The explanatory variables in an experiment are often called factors.

A treatment is any specific experimental condition applied to the subjects.

If an experiment has several factors, a treatment is a combination of specific values of each factor.
EXAMPLE 9.1 Effects of good day care

Does day care help low-income children stay in school and hold good jobs later in life? The Carolina Abecedarian Project (the name suggests the ABCs) has followed a group of children since 1972. The subjects are 111 people who in 1972 were healthy but low-income black infants in Chapel Hill, North Carolina. All the infants received nutritional supplements and help from social workers. Half, chosen at random, were also placed in an intensive preschool program. The experiment compares these two treatments. There is a single factor, “preschool, yes or no.” There are many response variables, recorded over more than 20 years, including academic test scores, college attendance, and employment.¹

EXAMPLE 9.2 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 investigated this question using undergraduate students as subjects. All subjects viewed a 40-minute television program that included ads for a digital camera. Some subjects saw a 30-second commercial; others, a 90-second version. The same commercial was shown either 1, 3, or 5 times during the program.

This experiment has two factors: length of the commercial, with 2 values, and repetitions, with 3 values. The 6 combinations of one value of each factor form 6 treatments. Figure 9.1 shows the layout of the treatments. After viewing, all of the subjects answered questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it. These are the response variables.²

Examples 9.1 and 9.2 illustrate the advantages of experiments over observational studies. In an experiment, we can study the effects of the specific treatments we are interested in. By assigning subjects to treatments, we can avoid confounding. If, for example, we simply compare children whose parents did and did not choose an intensive preschool program, we may find that children in the program come from richer and better-educated parents. Example 9.1 avoids that. Moreover,

![Figure 9.1](image_url)
we can control the environment of the subjects to hold constant factors that are of no interest to us, such as the specific product advertised in Example 9.2.

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 both 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 9.2 will help us find out.

APPLY YOUR KNOWLEDGE

9.1 Internet telephone calls. You can use your computer to make long-distance telephone calls over the Internet. How will the cost affect the use of this service? A university plans an experiment to find out. It will offer voice over Internet service to all 350 students in one of its dormitories. Some students will pay a low flat rate. Others will pay higher rates at peak periods and very low rates off-peak. The university is interested in how the payment plan affects the amount and time of use. What are the subjects, the factors, the treatments, and the response variables in this experiment?

9.2 Growing in the shade. Ability to grow in shade may help pines found in the dry forests of Arizona resist drought. How well do these pines grow in shade? Investigators planted pine seedlings in a greenhouse in either full light, light reduced to 25% of normal by shade cloth, or light reduced to 5% of normal. At the end of the study, they dried the young trees and weighed them. What are the individuals, the treatments, and the response variable in this experiment?

9.3 Improving adolescents’ habits. Most American adolescents don’t eat well and don’t exercise enough. Can middle schools increase physical activity among their students? Can they persuade students to eat better? Investigators designed a “physical activity intervention” to increase activity in physical education classes and during leisure periods throughout the school day. They also designed a “nutrition intervention” that improved school lunches and offered ideas for healthy home-packed lunches. Each participating school was randomly assigned to one of the interventions, both interventions, or no intervention. The investigators observed physical activity and lunchtime consumption of fat. Identify the individuals, the factors, and the response variables in this experiment. Use a diagram like that in Figure 9.1 to display the treatments.

How to experiment badly

Experiments are the preferred method for examining the effect of one variable on another. By imposing the specific treatment of interest and controlling other influences, we can pin down cause and effect. Statistical designs are often essential for effective experiments, just as they are for sampling. To see why, let’s start with an example of a bad design.
CHAPTER 9 • Producing Data: Experiments

Online vs. classroom → CAUSE? GMAT score after course

Students’ ages and backgrounds

FIGURE 9.2 Confounding. We can’t distinguish the effect of the treatment from the effects of lurking variables.

EXAMPLE 9.3 An uncontrolled experiment

A college regularly offers a review course to prepare candidates for the Graduate Management Admission Test (GMAT), which is required by most graduate business schools. This year, it offers only an online version of the course. The average GMAT score of students in the online course is 10% higher than the longtime average for those who took the classroom review course. Is the online course more effective?

This experiment has a very simple design. A group of subjects (the students) were exposed to a treatment (the online course), and the outcome (GMAT scores) was observed. Here is the design:

Subjects → Online course → GMAT scores

A closer look at the GMAT review course showed that the students in the online review course were quite different from the students who in past years took the classroom course. In particular, they were older and more likely to be employed. An online course appeals to these mature people, but we can’t compare their performance with that of the undergraduates who previously dominated the course. The online course might even be less effective than the classroom version. The effect of online versus in-class instruction is confounded with the effect of lurking variables. Figure 9.2 shows the confounding in picture form. As a result of confounding, the experiment is biased in favor of the online course.

Most laboratory experiments use a design like that in Example 9.3:

Subjects → Treatment → Measure response

In the controlled environment of the laboratory, simple designs often work well. Field experiments and experiments with human subjects are exposed to more variable conditions and deal with more variable subjects. A simple design often yields worthless results because of confounding with lurking variables.

APPLY YOUR KNOWLEDGE

9.4 Reducing unemployment. Will cash bonuses speed the return to work of unemployed people? A state department of employment security notes that last
Randomized comparative experiments

The remedy for the confounding in Example 9.3 is to do a comparative experiment in which some students are taught in the classroom and other, similar students take the course online. The first group is called a control group. Most well-designed experiments compare two or more treatments. Part of the design of an experiment is a description of the factors (explanatory variables) and the layout of the treatments, with comparison as the leading principle.

Comparison alone isn’t enough to produce results we can trust. If the treatments are given to groups that differ markedly when the experiment begins, bias will result. For example, if we allow students to elect online or classroom instruction, students who are older and employed are likely to sign up for the online course. Personal choice will bias our results in the same way that volunteers bias the results of online opinion polls. The solution to the problem of bias is the same for experiments and for samples: use impersonal chance to select the groups.

Randomized comparative experiments

An experiment that uses both comparison of two or more treatments and chance assignment of subjects to treatments is a randomized comparative experiment.

Example 9.4 On-campus versus online

The college decides to compare the progress of 25 on-campus students taught in the classroom with that of 25 students taught the same material online. Select the students who will be taught online by taking a simple random sample of size 25 from the 50 available subjects. The remaining 25 students form the control group. They will receive classroom instruction. The result is a randomized comparative experiment with two groups. Figure 9.3 outlines the design in graphical form.

The selection procedure is exactly the same as it is for sampling: label and table. **Step 1. Label** the 50 students 01 to 50. **Step 2. Table.** Go to the table of random digits and read successive two-digit groups. The first 25 labels encountered select the online group. As usual, ignore repeated labels and groups of digits not used as labels. For example, if you begin at line 125 in Table B, the first five students chosen are those labeled 21, 49, 37, 18, and 44. Software such as the Simple Random Sample applet makes it particularly easy to choose treatment groups at random.

Golfing at random

Random drawings give everyone the same chance to be chosen, so they offer a fair way to decide who gets a scarce good—like a round of golf. Lots of golfers want to play the famous Old Course at St. Andrews, Scotland. Some can reserve in advance, at considerable expense. Most must hope that chance favors them in the daily random drawing for tee times. At the height of the summer season, only 1 in 6 wins the right to pay $200 for a round.
The design in Figure 9.3 is comparative because it compares two treatments (the two instructional settings). It is randomized because the subjects are assigned to the treatments by chance. 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 about equal in size. We call designs like that in Figure 9.3 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 9.5 Conserving energy**

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 meters reduce electricity use? Would cheaper methods work almost as well? The company decides to conduct 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 9.4 outlines the design.

To carry out the random assignment, label the 60 households 01 to 60. Enter Table B (or use software) to select an SRS of 20 to receive the meters. Continue in Table B, selecting 20 more to receive charts. The remaining 20 form the control group.
Randomized comparative experiments

Random assignment

Group 1
20 houses

Group 2
20 houses

Group 3
20 houses

Treatment 1
Meter

Treatment 2
Chart

Treatment 3
Control

Compare electricity use

FIGURE 9.4 Outline of a completely randomized design comparing three energy-saving programs, for Example 9.5.

Examples 9.4 and 9.5 describe completely randomized designs that compare values of a single factor. In Example 9.4, the factor is the type of instruction. In Example 9.5, it is the method used to encourage energy conservation. Completely randomized designs can have more than one factor. The advertising experiment of Example 9.2 has two factors: the length and the number of repetitions of a television commercial. Their combinations form the six treatments outlined in Figure 9.1. 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

9.5 Does ginkgo improve memory? The law allows marketers of herbs and other natural substances to make health claims that are not supported by evidence. Brands of ginkgo extract claim to “improve memory and concentration.” A randomized comparative experiment found no evidence for such effects.3 The subjects were 230 healthy people over 60 years old. They were randomly assigned to ginkgo or a placebo pill (a dummy pill that looks and tastes the same). All the subjects took a battery of tests for learning and memory before treatment started and again after six weeks.

(a) Following the model of Figure 9.3, outline the design of this experiment.
(b) Use the Simple Random Sample applet, other software, or Table B to assign half the subjects to the ginkgo group. If you use software, report the first 20 members of the ginkgo group (in the applet’s “Sample bin”) and the first 20 members of the placebo group (those left in the “Population hopper”). If you use Table B, start at line 103 and choose only the first 5 members of the ginkgo group.

9.6 Can tea prevent cataracts? Eye cataracts are responsible for over 40% of blindness around the world. Can drinking tea regularly slow the growth of
cataracts? We can’t experiment on people, so we use rats as subjects. Researchers injected 18 young rats with a substance that causes cataracts. One group of the rats also received black tea extract; a second group received green tea extract; and a third got a placebo, a substance with no effect on the body. The response variable was the growth of cataracts over the next six weeks. Yes, both tea extracts did slow cataract growth.4

(a) Following the model of Figures 9.3 and 9.4, outline the design of this experiment.

(b) The Simple Random Sample applet allows you to randomly assign subjects to more than two groups. Use the applet to choose an SRS of 6 of the 18 rats to form the first group. Which rats are in this group? The “Population hopper” now contains the 12 remaining rats, in scrambled order. Click “Sample” again to choose an SRS of 6 of these to make up the second group. Which rats were chosen? The 6 rats remaining in the “Population hopper” form the third group.

9.7 Growing in the shade. You have 45 pine seedlings available for the experiment described in Exercise 9.2. Outline the design of this experiment. Use software or Table B to randomly assign seedlings to the three treatment groups.

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. Exercise 9.48 uses the Simple Random Sample applet to demonstrate this.

• 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 9.4, we cannot say that any difference between the average GMAT scores of students enrolled online and in the classroom must be caused by a difference in the effectiveness of the two types of instruction. There would be some difference even if both groups received the same instruction, because of variation among students in background and study habits. Chance assigns students to one group or the other, and this creates a chance difference between the groups. We would not trust an experiment with just one student in each group, for example. The results would depend too much on which
group got lucky and received the stronger student. If we assign many subjects to each group, however, the effects of chance will average out and there will be little difference in the average responses in the two groups unless the treatments 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**

The basic principles of statistical design of experiments are

1. **Control** the effects of lurking variables on the response, most simply by comparing two or more treatments.
2. **Randomize**—use impersonal chance to assign subjects to treatments.
3. **Use enough subjects** in each group to reduce chance variation in the results.

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**.

**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 randomized comparative 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**

9.8 **Conserving energy.** Example 9.5 describes an experiment to learn whether providing households with electronic meters or charts will reduce their electricity consumption. An executive of the electric 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 9.5.

9.9 Exercise and heart attacks. Does regular exercise reduce the risk of a heart attack? Here are two ways to study this question. Explain clearly why the second design will produce more trustworthy data.

1. A researcher finds 2000 men over 40 who exercise regularly and have not had heart attacks. She matches each with a similar man who does not exercise regularly, and she follows both groups for 5 years.
2. Another researcher finds 4000 men over 40 who have not had heart attacks and are willing to participate in a study. She assigns 2000 of the men to a regular program of supervised exercise. The other 2000 continue their usual habits. The researcher follows both groups for 5 years.

9.10 The Monday effect. Puzzling but true: stocks tend to go down on Mondays. There is no convincing explanation for this fact. A study looked at this "Monday effect" in more detail, using data on the daily returns of stocks over a 30-year period. Here are some of the findings:

To summarize, our results indicate that the well-known Monday effect is caused largely by the Mondays of the last two weeks of the month. The mean Monday return of the first three weeks of the month is, in general, not significantly different from zero and is generally significantly higher than the mean Monday return of the last two weeks. Our finding seems to make it more difficult to explain the Monday effect.5

A friend thinks that “significantly” in this article has its plain English meaning, roughly “I think this is important.” Explain in simple language what “significantly higher” and “not significantly different from zero” tell us here.

Scratch my furry ears
Rats and rabbits, specially bred to be uniform in their inherited characteristics, are the subjects in many experiments. Animals, like people, are quite sensitive to how they are treated. This can create opportunities for hidden bias. For example, human affection can change the cholesterol level of rabbits. Choose some rabbits at random and regularly remove them from their cages to have their heads scratched by friendly people. Leave other rabbits unloved. All the rabbits eat the same diet, but the rabbits that receive affection have lower cholesterol.

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.

The experiment on the effects of ginkgo on memory (Exercise 9.5) is a typical medical experiment. All of the subjects took the same tests and received the same medical attention. All of them took a pill every day, ginkgo in the treatment group and a placebo in the control group. A placebo is a dummy treatment. Many patients respond favorably to any treatment, even a placebo, perhaps because they trust the doctor. The response to a dummy treatment is called the placebo effect. If the control group did not take any pills, the effect of ginkgo in the treatment group would be confounded with the placebo effect, the effect of simply taking pills.

In addition, the study was double-blind. The subjects didn’t know whether they were taking ginkgo or a placebo. Neither did the investigators who worked with them. The double-blind method avoids unconscious bias by, for example, a doctor who is convinced that a new medical treatment must be better than a placebo.
In many medical studies, only the statistician who does the randomization knows which treatment each patient is receiving.

DOUBLE-BLIND EXPERIMENTS

In a double-blind experiment, neither the subjects nor the people who interact with them know which treatment each subject is receiving.

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.

EXAMPLE 9.6 Response to advertising

The study of television advertising in Example 9.2 showed a 40-minute videotape to students who knew an experiment was going on. We can't be sure that the results apply to everyday television viewers. Many behavioral science experiments use as subjects students or other volunteers who know they are subjects in an experiment. That's not a realistic setting.

EXAMPLE 9.7 Center brake lights

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%. Alas, 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 an 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

9.11 Dealing with pain. Health care providers are giving more attention to relieving the pain of cancer patients. An article in the journal Cancer surveyed a number of studies and concluded that controlled-release morphine tablets, which release the painkiller gradually over time, are more effective than giving standard morphine when the patient needs it. The “methods” section of the article begins: “Only those published studies that were controlled (i.e., randomized, double blind, and comparative), repeated-dose studies with CR morphine tablets in cancer pain
patients were considered for this review.” Explain the terms in parentheses to someone who knows nothing about medical experiments.

9.12 Does meditation reduce anxiety? 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. Explain what this means and how lack of blindness could bias the reported results.

Matched pairs and other block designs

Completely randomized designs are the simplest statistical designs for experiments. They illustrate clearly the principles of control, randomization, and adequate 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.

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. Use chance to decide which subject in a pair gets the first treatment. The other subject in that pair gets the other treatment. That is, the random assignment of subjects to treatments is done within each matched pair, not for all subjects at once. 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.

Example 9.8 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 react? Let’s compare two designs for this experiment. There are 40 student subjects available.

In a completely randomized design, all 40 subjects are assigned at random, 20 to simply drive and the other 20 to talk on the cell phone while driving. In the matched pairs design that was actually used, all subjects drive both with and without using the cell phone. The two drives are on separate days to reduce carryover effects. The order of the two treatments is assigned at random: 20 subjects are chosen to drive first with the phone, and the remaining 20 drive first without the phone.

Some subjects naturally react faster than others. The completely randomized design relies on chance to distribute the faster subjects roughly evenly between the two groups. The matched pairs design compares each subject’s reaction time with and without the cell phone. This makes it easier to see the effects of using the phone.
Matched pairs designs use the principles of comparison of treatments and randomization. However, the randomization is not complete—we do not randomly assign all the subjects at once to the two treatments. Instead, we randomize only within each matched pair. This allows matching to reduce the effect of variation among the subjects. Matched pairs are one kind of block design, with each pair forming a block.

**BLOCK DESIGN**

A block is a group of individuals that are known before the experiment to be similar in some way that is 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. 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. Here are some typical examples of block designs.

**EXAMPLE 9.9 Men, women, and advertising**

Women and men respond differently to advertising. An experiment to compare the effectiveness of three advertisements 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 sex. 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 advertisement. Then separately assign the men at random to three groups. Figure 9.5 outlines this improved design.

**FIGURE 9.5 Outline of a block design, for Example 9.9. The blocks consist of male and female subjects. The treatments are three advertisements for the same product.**
EXAMPLE 9.10 Comparing welfare policies

A social policy experiment will assess the effect on family income of several proposed new welfare systems and compare them with the present welfare system. Because the future income of a family is strongly related to its present income, the families who agree to participate are divided into blocks of similar income levels. The families in each block are then allocated at random among the welfare systems.

A block design allows us to draw separate conclusions about each block, for example, about men and women in Example 9.9. 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 advertisements. 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 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 concentrate for the most part on completely randomized experiments.

APPLY YOUR KNOWLEDGE

9.13 Comparing hand strength. Is the right hand generally stronger than the left in right-handed people? You can crudely measure hand strength by placing a bathroom scale on a shelf with the end protruding, then squeezing the scale between the thumb below and the four fingers above it. The reading of the scale shows the force exerted. Describe the design of a matched pairs experiment to compare the strength of the right and left hands, using 10 right-handed people as subjects. (You need not actually do the randomization.)

9.14 How long did I work? A psychologist wants to know if the difficulty of a task influences our estimate of how long we spend working at it. She designs two sets of mazes that subjects can work through on a computer. One set has easy mazes and the other has hard mazes. Subjects work until told to stop (after 6 minutes, but subjects do not know this). They are then asked to estimate how long they worked. The psychologist has 30 students available to serve as subjects.

(a) Describe the design of a completely randomized experiment to learn the effect of difficulty on estimated time.

(b) Describe the design of a matched pairs experiment using the same 30 subjects.

9.15 Technology for teaching statistics. The Brigham Young University statistics department is performing randomized comparative experiments to compare teaching methods. Response variables include students’ final-exam scores and a measure of their attitude toward statistics. One study compares two levels of technology for large lectures: standard (overhead projectors and chalk) and multimedia. The individuals in the study are the 8 lectures in a basic statistics course. There are four instructors, each of whom teaches two lectures. Because
the lecturers differ, their lectures form four blocks. Suppose the lectures and lecturers are as follows:

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Lecturer</th>
<th>Lecture</th>
<th>Lecturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hilton</td>
<td>5</td>
<td>Tolley</td>
</tr>
<tr>
<td>2</td>
<td>Christensen</td>
<td>6</td>
<td>Hilton</td>
</tr>
<tr>
<td>3</td>
<td>Hadfield</td>
<td>7</td>
<td>Tolley</td>
</tr>
<tr>
<td>4</td>
<td>Hadfield</td>
<td>8</td>
<td>Christensen</td>
</tr>
</tbody>
</table>

Outline a block design and do the randomization that your design requires.

**CHAPTER 9 SUMMARY**

In an experiment, we impose one or more treatments on individuals, often called subjects. Each treatment is a combination of values 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 and randomization to combat bias and using enough subjects to reduce chance variation.

The simplest form of control is comparison. Experiments should compare two or more treatments in order to avoid confounding of 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 using software or by giving numerical labels to the subjects and using a table of random digits to choose treatment groups.

Applying each treatment to 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. Many behavioral and medical experiments are double-blind. Some give a placebo to a control group. 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 individuals 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 each subject in a pair receives one of the treatments.
CHECK YOUR SKILLS

9.16 A study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same sex, age, and race who did not have brain cancer, then asked about use of cell phones. This is 
(a) an observational study. 
(b) an uncontrolled experiment. 
(c) a randomized comparative experiment.

9.17 What electrical changes occur in muscles as they get tired? Student subjects hold their arms above their shoulders until they have to drop them. Meanwhile, the electrical activity in their arm muscles is measured. This is 
(a) an observational study. 
(b) an uncontrolled experiment. 
(c) a randomized comparative experiment.

9.18 Can changing diet reduce high blood pressure? Vegetarian diets and low-salt diets are both promising. Men with high blood pressure are assigned at random to four diets: (1) normal diet with unrestricted salt; (2) vegetarian with unrestricted salt; (3) normal with restricted salt; and (4) vegetarian with restricted salt. This experiment has 
(a) one factor, the choice of diet. 
(b) two factors, normal/vegetarian diet and unrestricted/restricted salt. 
(c) four factors, the four diets being compared.

9.19 In the experiment of the previous exercise, the 240 subjects are labeled 001 to 240. Software assigns an SRS of 60 subjects to Diet 1, an SRS of 60 of the remaining 180 to Diet 2, and an SRS of 60 of the remaining 120 to Diet 3. The 60 who are left get Diet 4. This is a 
(a) completely randomized design. 
(b) block design, with four blocks. 
(c) matched pairs design.

9.20 An important response variable in the experiment described in Exercise 9.18 must be 
(a) the amount of salt in the subject's diet. 
(b) which of the four diets a subject is assigned to. 
(c) change in blood pressure after 8 weeks on the assigned diet.

9.21 A medical experiment compares an antidepression medicine with a placebo for relief of chronic headaches. There are 36 headache patients available to serve as subjects. To choose 18 patients to receive the medicine, you would 
(a) assign labels 01 to 36 and use Table B to choose 18. 
(b) assign labels 01 to 18, because only 18 need be chosen. 
(c) assign the first 18 who signed up to get the medicine.

9.22 The Community Intervention Trial for Smoking Cessation asked whether a community-wide advertising campaign would reduce smoking. The researchers located 11 pairs of communities, each pair similar in location, size, economic
status, and so on. One community in each pair participated in the advertising campaign and the other did not. This is
(a) an observational study.
(b) a matched pairs experiment.
(c) a completely randomized experiment.

9.23 To decide which community in each pair in the previous exercise should get the advertising campaign, it is best to
(a) toss a coin.
(b) choose the community that will help pay for the campaign.
(c) choose the community with a mayor who will participate.

9.24 A marketing class designs two videos advertising an expensive Mercedes sports car. They test the videos by asking fellow students to view both (in random order) and say which makes them more likely to buy the car. Mercedes should be reluctant to agree that the video favored in this study will sell more cars because
(a) the study used a matched pairs design instead of a completely randomized design.
(b) results from students may not generalize to the older and richer customers who might buy a Mercedes.
(c) this is an observational study, not an experiment.

CHAPTER 9 EXERCISES

In all exercises that require randomization, you may use Table B, the Simple Random Sample applet, or other software. See Exercise 9.6 for directions on using the applet for more than two treatment groups.

9.25 Wine, beer, or spirits? Example 8.2 (page 191) describes a study that compared three groups of people: the first group drinks mostly wine, the second drinks mostly beer, and the third drinks mostly spirits. This study is comparative, but it is not an experiment. Why not?

9.26 Treating breast cancer. The most common treatment for breast cancer discovered in its early stages was once removal of the breast. It is now usual to remove only the tumor and nearby lymph nodes, followed by radiation. To study whether these treatments differ in their effectiveness, a medical team examines the records of 25 large hospitals and compares the survival times after surgery of all women who have had either treatment.
(a) What are the explanatory and response variables?
(b) Explain carefully why this study is not an experiment.
(c) Explain why confounding will prevent this study from discovering which treatment is more effective. (The current treatment was in fact recommended after several large randomized comparative experiments.)

9.27 Wine, beer, or spirits? You have recruited 300 adults aged 45 to 65 who are willing to follow your orders about alcohol consumption over the next five years. You want to compare the effects on heart disease of moderate drinking of just wine, just beer, or just spirits. Outline the design of a completely randomized
9.28 Marijuana and work. How does smoking marijuana affect willingness to work? Canadian researchers persuaded young adult men who used marijuana to live for 98 days in a “planned environment.” The men earned money by weaving belts. They used their earnings to pay for meals and other consumption and could keep any money left over. One group smoked two potent marijuana cigarettes every evening. The other group smoked two weak marijuana cigarettes. All subjects could buy more cigarettes but were given strong or weak cigarettes depending on their group. Did the weak and strong groups differ in work output and earnings?9

(a) Outline the design of this experiment.
(b) Here are the names of the 20 subjects. Use software or Table B at line 131 to carry out the randomization your design requires.

Abate Dubois Gutierrez Lucero Rosen
Afifi Engel Huang McNeill Thompson
Brown Fluharty Iselin Morse Travers
Cheng Gerson Kaplan Quinones Ullmann

9.29 The benefits of red wine. Does red wine protect moderate drinkers from heart disease better than other alcoholic beverages? Red wine contains substances called polyphenols that may change blood chemistry in a desirable way. This calls for a randomized comparative experiment. The subjects were healthy men aged 35 to 65. They were randomly assigned to drink red wine (9 subjects), drink white wine (9 subjects), drink white wine and also take polyphenols from red wine (6 subjects), take polyphenols alone (9 subjects), or drink vodka and lemonade (6 subjects).10 Outline the design of the experiment and randomly assign the 39 subjects to the 5 groups. If you use Table B, start at line 107.

9.30 Response to TV ads. You decide to use a completely randomized design in the two-factor experiment on response to advertising described in Example 9.2 (page 214). The 36 students named below will serve as subjects. (Ignore the asterisks.) Outline the design and randomly assign the subjects to the 6 treatments. If you use Table B, start at line 130.

Alomar Denman Han Liang Padilla* Valasco
Asihiro* Durr* Howard* Maldonado Plochman Vaughn
Bennett Edwards* Hruska Marsden Rosen* Wei
Bikalis Farouk Imrani Montoya* Solomon Wilder*
Chao* Fleming James O’Brien Trujillo Willis
Clemente George Kaplan* Ogle* Tullock Zhang*

9.31 Improving adolescents’ habits. Twenty-four public middle schools agree to participate in the experiment described in Exercise 9.3 (page 215). Use a diagram to outline a completely randomized design for this experiment. Do the randomization required to assign schools to treatments. If you use the Simple Random Sample applet or other software, choose all four treatment groups. If you use Table B, start at line 105 and choose only the first two groups.

9.32 Relieving headaches. Doctors identify “chronic tension–type headaches” as headaches that occur almost daily for at least six months. Can antidepressant
medications or stress management training reduce the number and severity of these headaches? Are both together more effective than either alone?

(a) Use a diagram like Figure 9.1 to display the treatments in a design with two factors: "medication, yes or no" and "stress management, yes or no." Then outline the design of a completely randomized experiment to compare these treatments.

(b) The headache sufferers named below have agreed to participate in the study. Randomly assign the subjects to the treatments. If you use the Simple Random Sample applet or other software, assign all the subjects. If you use Table B, start at line 130 and assign subjects to only the first treatment group.

Abbott  Decker  Herrera  Lucero  Richter
Abdalla  Devlin  Hersch  Masters  Riley
Alawi  Engel  Hurwitz  Morgan  Samuels
Brodan  Fuentes  Irwin  Nelson  Smith
Chai  Garrett  Jiang  Nho  Suarez
Chuang  Gill  Kelley  Ortiz  Upasani
Cordoba  Glover  Kim  Ramdas  Wilson
Custer  Hammond  Landers  Reed  Xiang

9.33 Fabric finishing. A maker of fabric for clothing is setting up a new line to “finish” the raw fabric. The line will use either metal rollers or natural-bristle rollers to raise the surface of the fabric; a dyeing cycle time of either 30 minutes or 40 minutes; and a temperature of either 150°C or 175°C. An experiment will compare all combinations of these choices. Three specimens of fabric will be subjected to each treatment and scored for quality.

(a) What are the factors and the treatments? How many individuals (fabric specimens) does the experiment require?

(b) Outline a completely randomized design for this experiment. (You need not actually do the randomization.)

9.34 Frappuccino light? Here’s the opening of a press release from June 2004: “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 Mocha Frappuccino coffee.

(a) Describe a matched pairs design to answer this question. Be sure to include proper blinding of your subjects.

(b) You have 20 regular Starbucks customers on hand. Use the Simple Random Sample applet or Table B at line 141 to do the randomization that your design requires.

9.35 Growing trees faster. The concentration of carbon dioxide (CO₂) in the atmosphere is increasing rapidly due to our use of fossil fuels. Because green plants use CO₂ to fuel photosynthesis, more CO₂ may cause trees to grow faster. An elaborate apparatus allows researchers to pipe extra CO₂ to a 30-meter circle of forest. We want to compare the growth in base area of trees in treated and untreated areas to see if extra CO₂ does in fact increase growth. We can afford to treat three circular areas.
(a) Describe the design of a completely randomized experiment using six well-separated 30-meter circular areas in a pine forest. Sketch the circles and carry out the randomization your design calls for.

(b) Areas within the forest may differ in soil fertility. Describe a matched pairs design using three pairs of circles that will reduce the extra variation due to different fertility. Sketch the circles and carry out the randomization your design calls for.

9.36 **Athletes taking oxygen.** We often see players on the sidelines of a football game inhaling oxygen. Their coaches think this will speed their recovery. We might measure recovery from intense exertion as follows: Have a football player run 100 yards three times in quick succession. Then allow three minutes to rest before running 100 yards again. Time the final run. Because players vary greatly in speed, you plan a matched pairs experiment using 25 football players as subjects. Discuss the design of such an experiment to investigate the effect of inhaling oxygen during the rest period.

9.37 **Protecting ultramarathon runners.** An ultramarathon, as you might guess, is a footrace longer than the 26.2 miles of a marathon. Runners commonly develop respiratory infections after an ultramarathon. Will taking 600 milligrams of vitamin C daily reduce these infections? Researchers randomly assigned ultramarathon runners to receive either vitamin C or a placebo. Separately, they also randomly assigned these treatments to a group of nonrunners the same age as the runners. All subjects were watched for 14 days after the big race to see if infections developed.12

(a) What is the name for this experimental design?

(b) Use a diagram to outline the design.

9.38 **Reducing spine fractures.** Fractures of the spine are common and serious among women with advanced osteoporosis (low mineral density in the bones). Can taking strontium renate help? A large medical experiment assigned 1649 women to take either strontium renate or a placebo each day. All of the subjects had osteoporosis and had suffered at least one fracture. All were taking calcium supplements and receiving standard medical care. The response variables were measurements of bone density and counts of new fractures over three years. The subjects were treated at 10 medical centers in 10 different countries.13 Outline a block design for this experiment, with the medical centers as blocks. Explain why this is the proper design.

9.39 **Wine, beer, or spirits?** Women as a group develop heart disease much later than men. We can improve the completely randomized design of Exercise 9.27 by using women and men as blocks. Your 300 subjects include 120 women and 180 men. Outline a block design for comparing wine, beer, and spirits. Be sure to say how many subjects you will put in each group in your design.

9.40 **Response to TV ads, continued.** We can improve on the completely randomized design you outlined in Exercise 9.30. The 36 subjects include 24 women and 12 men. Men and women often react differently to advertising. You therefore decide to use a block design with the two genders as blocks. You must assign the 6 treatments at random within each block separately.

(a) Outline the design with a diagram.
(b) The 12 men are marked with asterisks in the list in Exercise 9.30. Use Table B, beginning at line 140, to do the randomization. Report your result in a table that lists the 24 women and 12 men and the treatment you assigned to each.

9.41 Prayer and meditation. You read in a magazine that “nonphysical treatments such as meditation and prayer have been shown to be effective in controlled scientific studies for such ailments as high blood pressure, insomnia, ulcers, and asthma.” Explain in simple language what the article means by “controlled scientific studies.” Why can such studies in principle provide good evidence that, for example, meditation is an effective treatment for high blood pressure?

9.42 College students. Give an example of a question about college students, their behavior, or their opinions that would best be answered by
(a) a sample survey.
(b) an experiment.

9.43 Quick randomizing. Here’s a quick and easy way to randomize. You have 100 subjects, 50 women and 50 men. Toss a coin. If it’s heads, assign the men to the treatment group and the women to the control group. If the coin comes up tails, assign the women to treatment and the men to control. This gives every individual subject a 50-50 chance of being assigned to treatment or control. Why isn’t this a good way to randomly assign subjects to treatment groups?

9.44 Daytime running lights. Canada requires that cars be equipped with “daytime running lights,” headlights that automatically come on at a low level when the car is started. Many 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 9.7 (page 223) discusses center brake lights. What cautions do you draw from that example that apply to an experiment on the effects of running lights?

9.45 Do antioxidants prevent cancer? People who eat lots of fruits and vegetables have lower rates of colon cancer than those who eat little of these foods. Fruits and vegetables are rich in “antioxidants” such as vitamins A, C, and E. Will taking antioxidants help prevent colon cancer? A medical experiment studied this question with 864 people who were at risk of colon cancer. The subjects were divided into four groups: daily beta-carotene, daily vitamins C and E, all three vitamins every day, or daily placebo. After four years, the researchers were surprised to find no significant difference in colon cancer among the groups.¹⁴
(a) What are the explanatory and response variables in this experiment?
(b) Outline the design of the experiment. Use your judgment in choosing the group sizes.
(c) The study was double-blind. What does this mean?
(d) What does “no significant difference” mean in describing the outcome of the study?
(e) Suggest some lurking variables that could explain why people who eat lots of fruits and vegetables have lower rates of colon cancer. The experiment suggests
that these variables, rather than the antioxidants, may be responsible for the observed benefits of fruits and vegetables.

9.46 An herb for depression? Does the herb Saint-John’s-wort relieve major depression? Here are some excerpts from the report of a study of this issue. The study concluded that the herb is no more effective than a placebo.

(a) “Design: Randomized, double-blind, placebo-controlled clinical trial....” A clinical trial is a medical experiment using actual patients as subjects. Explain the meaning of each of the other terms in this description.

(b) “Participants... were randomly assigned to receive either Saint-John’s-wort extract (n = 98) or placebo (n = 102).... The primary outcome measure was the rate of change in the Hamilton Rating Scale for Depression over the treatment period.” Based on this information, use a diagram to outline the design of this clinical trial.

9.47 Explaining medical research. Observational studies had suggested that vitamin E reduces the risk of heart disease. Careful experiments, however, showed that vitamin E has no effect, at least for women. According to a commentary in the *Journal of the American Medical Association*:

Thus, vitamin E enters the category of therapies that were promising in epidemiologic and observational studies but failed to deliver in adequately powered randomized controlled trials. As in other studies, the “healthy user” bias must be considered, ie, the healthy lifestyle behaviors that characterize individuals who care enough about their health to take various supplements are actually responsible for the better health, but this is minimized with the rigorous trial design.

A friend who knows no statistics asks you to explain this.

(a) What is the difference between observational studies and experiments?

(b) What is a “randomized controlled trial”? (We’ll discuss “adequately powered” in Chapter 16.)

(c) How does “healthy user bias” explain how people who take vitamin E supplements have better health in observational studies but not in controlled experiments?

9.48 Randomization avoids bias. Suppose that the 25 even-numbered students among the 50 students available for the comparison of on-campus and online instruction (Example 9.4) are older, employed students. We hope that randomization will distribute these students roughly equally between the on-campus and online groups. Use the Simple Random Sample applet to take 20 samples of size 25 from the 50 students. (Be sure to click “Reset” after each sample.) Record the counts of even-numbered students in each of your 20 samples. You see that there is considerable chance variation but no systematic bias in favor of one or the other group in assigning the older students. Larger samples from a larger population will on the average do an even better job of creating two similar groups.
Data Ethics*

The production and use of data, like all human endeavors, raise 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. It will end your career when uncovered. 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 1** The whole truth?

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.”¹ The situation is no doubt worse in fields that screen published work less carefully.

---

*This short essay concerns a very important topic, but the material is not needed to read the rest of the book.*
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 individuals who are subjects in a study must give their informed consent before data are collected.

All individual 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. 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 its 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.
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 2  Who can 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.

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 individuals 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.

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 individual set of data protects people’s privacy. The government, for example, maintains a vast amount of information about citizens in many separate data bases—census responses, tax returns, Social Security information, data from surveys such as the Current Population Survey, and so on. Many of these data bases can be searched by computers for statistical studies. A clever computer search of several data bases 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 Uncle Sam knows**

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 you paid the right amount of tax and how large a Social Security benefit you are owed when you retire. Some people feel that individuals 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.

**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 4  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.

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.

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.

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 5 Psychologists in the men’s room

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.
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. What should we protect subjects from when physical harm is unlikely? Possible 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 looking for. 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 6  Reducing domestic violence**

How should police respond to domestic violence calls? In the past, the usual practice was to remove the offender and order him 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 him overnight, or warn the suspect and release him. 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 arrest the suspect or warn and release him, 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.\(^7\) 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.”

DISCUSSION EXERCISES

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

1. 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.

2. Who reviews? 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. Getting 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.

4. No consent needed? In which of the circumstances below 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) The social psychologist pretends to be converted to membership in a religious group and attends private meetings to study the behavior patterns of members.
5. **Studying your blood.** Long ago, doctors drew a blood specimen from you as part of treating minor 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 OK 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 nonetheless acceptable, given complete confidentiality and the fact that using the sample can’t physically harm the patient?

6. **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.

7. **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?

8. **Political polls.** The presidential election campaign is in full swing, and the candidates have hired polling organizations to take sample surveys to find out what the voters think about the issues. What information should the pollsters be required to give out?

(a) What does the standard of informed consent require the pollsters to tell potential respondents?

(b) The standards accepted by polling organizations also require giving respondents the name and address of the organization that carries out the poll. Why do you think this is required?

(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?

9. **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 they are paying for the poll in order to gain information for their 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?

10. **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?

11. **Unequal benefits.** Researchers on aging proposed to investigate the effect of supplemental health services on the quality of life of older people. Eligible patients on the rolls 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?

12. **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?

13. **AIDS trials in Africa.** Effective drugs for treating AIDS are very expensive, so some 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 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?

14. **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—it is the future benefits that would go elsewhere. What do you think?

15. **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.

16. **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 ethically OK?

17. **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.