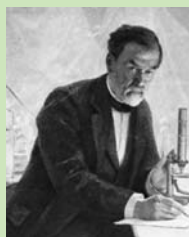# Summer assignment for Honors Probability and Statistics

<u>Please read the following pages from Chapter 1</u>:
- ➢ Lesson 1.1 – What is Statistics? (pages 2 – 10)
- ➢ Lesson 1.2 – Random Samples (pages 12 – 17)
- ➢ Lesson 1.3 – Introduction to Experimental Design (pages 20 – 26)

After reading the above pages from Chapter 1, then do problems #1 – 33 all (the last 10 pages of this document).

1

Chance favors the prepared mind.

—Louis Pasteur

Statistical techniques are tools of thought . . . not substitutes for thought.

—Abrahm Kaplan

For on-line student resources, visit the Brase/Brase, *Understanding Basic Statistics,* 5th edition web site at **www.cengage.com/statistics/Brase/UBS5e.**

Louis Pasteur (1822–1895) is the founder of modern bacteriology. When studying cholera, he accidentally left some bacillus culture unattended over the summer. In the fall, he injected laboratory animals with this bacilli. To his surprise, the animals did not die— in fact, they had become vaccinated against cholera!

As the first quote at the left reminds us, our chances of success are greatly improved if we have a "prepared mind." The statistical methods you will learn in this book will help you achieve a prepared mind for the study of many different fields. The second quote reminds us that statistics is an important tool, but it is not a replacement for an in-depth knowledge of the field to which it is being applied.

The authors of this book want you to understand and enjoy statistics. The reading material will *tell you* about the subject. The examples will *show you* how it works. To understand, however, you must *get involved*. Guided exercises, calculator and computer applications, section and chapter problems, and writing exercises are all designed to get you involved in the subject. As you grow in your understanding of statistics, we believe you will enjoy learning a subject that has a world full of interesting applications.
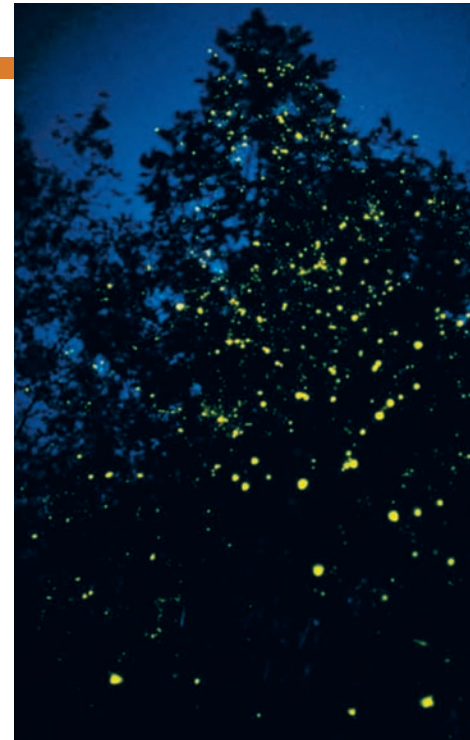
Getting Started

PREVIEW QUESTIONS

Why is statistics important? (SECTION 1.1)

What is the nature of data? (SECTION 1.1)

How can you draw a random sample? (SECTION 1.2)

What are other sampling techniques? (SECTION 1.2)

How can you design ways to collect data? (SECTION 1.3)

FOCUS PROBLEM

Where Have All the Fireflies Gone?

A feature article in *The Wall Street Journal* discusses the disappearance of fireflies. In the article, Professor Sara Lewis of Tufts University and other scholars express concern about the decline in the worldwide population of fireflies.

There are a number of possible explanations for the decline, including habitat reduction of woodlands, wetlands, and open fields; pesticides; and pollution. Artificial nighttime lighting might interfere with the Morse-code-like mating ritual of the fireflies. Some chemical companies pay a bounty for fireflies because the insects contain two rare chemicals used in medical research and electronic detection systems in spacecraft.

What does any of this have to do with statistics?

The truth, at this time, is that no one really knows (a) how much the world firefly population has declined or (b) how to explain the decline. The population of all fireflies is simply too large to study in its entirety.

In any study of fireflies, we must rely on incomplete information from samples. Furthermore, from these samples we must draw realistic conclusions that have statistical integrity. This is the kind of work that makes use of statistical methods to determine ways to collect, analyze, and investigate data.

Adapted from Ohio State University Firefly Files logo

Suppose you are conducting a study to compare firefly populations exposed to normal daylight/darkness conditions with firefly populations exposed to continuous light (24 hours a day). You set up two firefly colonies in a laboratory environment. The two colonies are identical

except that one colony is exposed to normal daylight/darkness conditions and the other is exposed to continuous light. Each colony is populated with the same number of mature fireflies. After 72 hours, you count the number of living fireflies in each colony.

After completing this chapter, you will be able to answer the following questions.

(a) Is this an experiment or an observation study? Explain.

(b) Is there a control group? Is there a treatment group?

(c) What is the variable in this study?

(d) What is the level of measurement (nominal, interval, ordinal, or ratio) of the variable?

(See Problem 9 of the Chapter 1 Review Problems.)

SECTION 1.1

What Is Statistics?

FOCUS POINTS
- Identify variables in a statistical study.
- Distinguish between quantitative and qualitative variables.
- Identify populations and samples.
- Distinguish between parameters and statistics.
- Determine the level of measurement.
- Compare descriptive and inferential statistics.

Introduction

Decision making is an important aspect of our lives. We make decisions based on the information we have, our attitudes, and our values. Statistical methods help us examine information. Moreover, statistics can be used for making decisions when we are faced with uncertainties. For instance, if we wish to estimate the proportion of people who will have a severe reaction to a flu shot without giving the shot to everyone who wants it, statistics provides appropriate methods. Statistical methods enable us to look at information from a small collection of people or items and make inferences about a larger collection of people or items.

Procedures for analyzing data, together with rules of inference, are central topics in the study of statistics.

Statistics

> **Statistics** is the study of how to collect, organize, analyze, and interpret numerical information from data.

The statistical procedures you will learn in this book should supplement your built-in system of inference—that is, the results of statistical procedures and good sense should dovetail. Of course, statistical methods themselves have no power to work miracles. These methods can help us make some decisions, but not all conceivable decisions. Remember, a properly applied statistical procedure is no more accurate than the data, or facts, on which it is based. Finally, statistical results should be interpreted by one who understands not only the methods, but also the subject matter to which they have been applied.

The general prerequisite for statistical decision making is the gathering of data. First, we need to identify the individuals or objects to be included in the study and the characteristics or features of the individuals that are of interest.

Individuals
Variable

> **Individuals** are the people or objects included in the study.
> A **variable** is a characteristic of the individual to be measured or observed.

For instance, if we want to conduct a study about the people who have climbed Mt. Everest, then the individuals in the study are all people who have actually made it to the summit. One variable might be the height of such individuals. Other variables might be age, weight, gender, nationality, income, and so on. Regardless of the variables we use, we would not include measurements or observations from people who have not climbed the mountain.

The variables in a study may be *quantitative* or *qualitative* in nature.

Quantitative variable
Qualitative variable

> A **quantitative variable** has a value or numerical measurement for which operations such as addition or averaging make sense. A **qualitative variable** describes an individual by placing the individual into a category or group, such as male or female.

For the Mt. Everest climbers, variables such as height, weight, age, or income are *quantitative* variables. *Qualitative variables* involve nonnumerical observations such as gender or nationality. Sometimes qualitative variables are referred to as *categorical variables*.

Another important issue regarding data is their source. Do the data comprise information from *all* individuals of interest, or from just *some* of the individuals?

Population data
Sample data

> In **population data,** the data are from *every* individual of interest.
> In **sample data,** the data are from *only some* of the individuals of interest.

It is important to know whether the data are population data or sample data. Data from a specific population are fixed and complete. Data from a sample may vary from sample to sample and are *not* complete.

Parameter
Statistic

> A **parameter** is a numerical measure that describes an aspect of a population.
> A **statistic** is a numerical measure that describes an aspect of a sample.

For instance, if we have data from *all* the individuals who have climbed Mt. Everest, then we have population data. The proportion of males in the *population* of all climbers who have conquered Mt. Everest is an example of a *parameter.*

On the other hand, if our data come from just some of the climbers, we have sample data. The proportion of male climbers in the *sample* is an example of a *statistic.* Note that different samples may have different values for the proportion of male climbers. One of the important features of sample statistics is that they can vary from sample to sample, whereas population parameters are fixed for a given population.

EXAMPLE 1   USING BASIC TERMINOLOGY

The Hawaii Department of Tropical Agriculture is conducting a study of ready-to-harvest pineapples in an experimental field.

(a) The pineapples are the *objects* (individuals) of the study. If the researchers are interested in the individual weights of pineapples in the field, then the *variable* consists of weights. At this point, it is important to specify units of measurement and degree of accuracy of measurement. The weights could be

measured to the nearest ounce or gram. Weight is a *quantitative* variable because it is a numerical measure. If the weights of *all* the ready-to-harvest pineapples in the field are included in the data, then we have a *population*. The average weight of all ready-to-harvest pineapples in the field is a *parameter*.

(b) Suppose the researchers also want data on taste. A panel of tasters rates the pineapples according to the categories "poor," "acceptable," and "good." Only some of the pineapples are included in the taste test. In this case, the *variable* is taste. This is a *qualitative* or *categorical* variable. Because only some of the pineapples in the field are included in the study, we have a *sample*. The proportion of pineapples in the sample with a taste rating of "good" is a *statistic*.

Throughout this text, you will encounter *guided exercises* embedded in the reading material. These exercises are included to give you an opportunity to work immediately with new ideas. The questions guide you through appropriate analysis. Cover the answers on the right side (an index card will fit this purpose). After you have thought about or written down *your own response,* check the answers. If there are several parts to an exercise, check each part before you continue. You should be able to answer most of these exercise questions, but don't skip them—they are important.

GUIDED EXERCISE 1 | **Using basic terminology**

Television station QUE wants to know the proportion of TV owners in Virginia who watch the station's new program at least once a week. The station asked a group of 1000 TV owners in Virginia if they watch the program at least once a week.

(a) Identify the individuals of the study and the variable.

⟹ The individuals are the 1000 TV owners surveyed. The variable is the response does, or does not, watch the new program at least once a week.

(b) Do the data comprise a sample? If so, what is the underlying population?

⟹ The data comprise a sample of the population of responses from all TV owners in Virginia.

(c) Is the variable qualitative or quantitative?

⟹ Qualitative—the categories are the two possible responses, does or does not watch the program.

(d) Identify a quantitative variable that might be of interest.

⟹ Age or income might be of interest.

(e) Is the proportion of viewers in the sample who watch the new program at least once a week a statistic or a parameter?

⟹ Statistic—the proportion is computed from sample data.

Levels of Measurement: Nominal, Ordinal, Interval, Ratio

We have categorized data as either qualitative or quantitative. Another way to classify data is according to one of the four *levels of measurement*. These levels indicate the type of arithmetic that is appropriate for the data, such as ordering, taking differences, or taking ratios.

Levels of measurement

Nominal level

The **nominal level of measurement** applies to data that consist of names, labels, or categories. There are no implied criteria by which the data can be ordered from smallest to largest.

Ordinal level

The **ordinal level of measurement** applies to data that can be arranged in order. However, differences between data values either cannot be determined or are meaningless.

Interval level

The **interval level of measurement** applies to data that can be arranged in order. In addition, differences between data values are meaningful.

Ratio level

The **ratio level of measurement** applies to data that can be arranged in order. In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.

EXAMPLE 2   LEVELS OF MEASUREMENT

Identify the type of data.

(a) Taos, Acoma, Zuni, and Cochiti are the names of four Native American pueblos from the population of names of all Native American pueblos in Arizona and New Mexico.

SOLUTION: These data are at the *nominal* level. Notice that these data values are simply names. By looking at the name alone, we cannot determine if one name is "greater than or less than" another. Any ordering of the names would be numerically meaningless.

(b) In a high school graduating class of 319 students, Jim ranked 25th, June ranked 19th, Walter ranked 10th, and Julia ranked 4th, where 1 is the highest rank.

SOLUTION: These data are at the *ordinal* level. Ordering the data clearly makes sense. Walter ranked higher than June. Jim had the lowest rank, and Julia the highest. However, numerical differences in ranks do not have meaning. The difference between June's and Jim's rank is 6, and this is the same difference that exists between Walter's and Julia's rank. However, this difference doesn't really mean anything significant. For instance, if you looked at grade point average, Walter and Julia may have had a large gap between their grade point averages, whereas June and Jim may have had closer grade point averages. In any ranking system, it is only the relative standing that matters. Differences between ranks are meaningless.

(c) Body temperatures (in degrees Celsius) of trout in the Yellowstone River.

SOLUTION: These data are at the *interval* level. We can certainly order the data, and we can compute meaningful differences. However, for Celsius-scale temperatures, there is not an inherent starting point. The value 0°C may seem to be a starting point, but this value does not indicate the state of "no heat." Furthermore, it is not correct to say that 20°C is twice as hot as 10°C.

(d) Length of trout swimming in the Yellowstone River.

SOLUTION: These data are at the *ratio* level. An 18-inch trout is three times as long as a 6-inch trout. Observe that we can divide 6 into 18 to determine a meaningful *ratio* of trout lengths.

In summary, there are four levels of measurement. The nominal level is considered the lowest, and in ascending order we have the ordinal, interval, and ratio levels. In general, calculations based on a particular level of measurement may not be appropriate for a lower level.

PROCEDURE

HOW TO DETERMINE THE LEVEL OF MEASUREMENT

The levels of measurement, listed from lowest to highest, are nominal, ordinal, interval, and ratio. To determine the level of measurement of a set of data, state the *highest level* that can be justified for the entire collection of data. Consider which calculations are suitable for the data.

Level of Measurement	Suitable Calculation
Nominal	We can put the data into categories.
Ordinal	We can order the data from smallest to largest or "worst" to "best." Each data value can be *compared* with another data value.
Interval	We can order the data and also take the differences between data values. At this level, it makes sense to compare the differences between data values. For instance, we can say that one data value is 5 more than or 12 less than another data value.
Ratio	We can order the data, take differences, and also find the ratio between data values. For instance, it makes sense to say that one data value is twice as large as another.

GUIDED EXERCISE 2 | **Levels of measurement**

The following describe different data associated with a state senator. For each data entry, indicate the corresponding *level of measurement.*

(a) The senator's name is Sam Wilson.

⟹ Nominal level

(b) The senator is 58 years old.

⟹ Ratio level. Notice that age has a meaningful zero. It makes sense to give age ratios. For instance, Sam is twice as old as someone who is 29.

(c) The years in which the senator was elected to the Senate are 1992, 1998, and 2004.

⟹ Interval level. Dates can be ordered, and the difference between dates has meaning. For instance, 2004 is six years later than 1998. However, ratios do not make sense. The year 2000 is not twice as large as the year 1000. In addition, the year 0 does not mean "no time."

(d) The senator's total taxable income last year was $878,314.

⟹ Ratio level. It makes sense to say that the senator's income is 10 times that of someone earning $87,831.40.

Continued

GUIDED EXERCISE 2 *continued*

(e) The senator surveyed his constituents regarding his proposed water protection bill. The choices for response were strong support, support, neutral, against, or strongly against.

⟹ Ordinal level. The choices can be ordered, but there is no meaningful numerical difference between two choices.

(f) The senator's marital status is "married."

⟹ Nominal level

(g) A leading news magazine claims the senator is ranked seventh for his voting record on bills regarding public education.

⟹ Ordinal level. Ranks can be ordered, but differences between ranks may vary in meaning.

CRITICAL THINKING

"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay." Sherlock Holmes said these words in *The Adventure of the Copper Beeches* by Sir Arthur Conan Doyle.

Reliable statistical conclusions require reliable data. This section has provided some of the vocabulary used in discussing data. As you read a statistical study or conduct one, pay attention to the nature of the data and the ways they were collected.

When you select a variable to measure, be sure to specify the process and requirements for measurement. For example, if the variable is the weight of ready-to-harvest pineapples, specify the unit of weight, the accuracy of measurement, and maybe even the particular scale to be used. If some weights are in ounces and others in grams, the data are fairly useless.

Another concern is whether or not your measurement instrument truly measures the variable. Just asking people if they know the geographic location of the island nation of Fiji may not provide accurate results. The answers may reflect the fact that the respondents want you to think they are knowledgeable. Asking people to locate Fiji on a map may give more reliable results.

The level of measurement is also an issue. You can put numbers into a calculator or computer and do all kinds of arithmetic. However, you need to judge whether the operations are meaningful. For ordinal data such as restaurant rankings, you can't conclude that a 4-star restaurant is "twice as good" as a 2-star restaurant, even though the number 4 is twice 2.

Are the data from a sample, or do they comprise the entire population? Sample data can vary from one sample to another! This means that if you are studying the same statistic from two different samples of the same size, the data values may be different. In fact, the ways in which sample statistics vary among different samples of the same size will be the focus of our study from Chapter 7 on.

Looking Ahead

The purpose of collecting and analyzing data is to obtain information. Statistical methods provide us tools with which we can obtain information from data. These methods break into two branches.

Descriptive statistics

Descriptive statistics involves methods of organizing, picturing, and summarizing information from samples or populations.

Inferential statistics

Inferential statistics involves methods of using information from a sample to draw conclusions regarding the population.

We will look at methods of descriptive statistics in Chapters 2, 3, and 4. These methods may be applied to data from samples or populations.

Sometimes we do not have access to an entire population. At other times, the difficulties or expense of working with the entire population is prohibitive. In such cases, we will use inferential statistics together with probability. These are the topics of Chapters 5 through 11.

VIEWPOINT | The First Measured Century

The twentieth century saw measurements of aspects of American life that never had been systematically studied before. Social conditions involving crime, sex, food, fun, religion, and work were numerically investigated. The measurements and survey responses taken over the entire century reveal unsuspected statistical trends. The First Measured Century *is a book by Caplow, Hicks, and Wattenberg. It is also a PBS documentary available on video. For more information, visit the Online Study Center at* **www.cengage.com/statistics/Brase/UBS5e** *and find the link to the PBS* First Measured Century *documentary.*

SECTION 1.1 PROBLEMS

1. *Statistical Literacy* What is the difference between an individual and a variable?

2. *Statistical Literacy* Are data at the nominal level of measurement quantitative or qualitative?

3. *Statistical Literacy* What is the difference between a parameter and a statistic?

4. *Statistical Literacy* For a set population, does a parameter ever change? If there are three different samples of the same size from a set population, is it possible to get three different values for the same statistic?

5. *Marketing: Fast Food* A nationwide survey of adults asks, "How many times per week do you eat in a fast-food restaurant?" Possible answers: 0, 1–3, 4 or more.
   (a) Identify the variable.
   (b) Is the variable quantitative or qualitative?
   (c) What is the implied population?

6. *Advertising: Auto Mileage* What is the average miles per gallon (mpg) for all new cars? Using *Consumer Reports*, a random sample of 35 new cars gave an average of 21.1 mpg.
   (a) Identify the variable.
   (b) Is the variable quantitative or qualitative?
   (c) What is the implied population?

7. *Ecology: Wetlands* Government agencies carefully monitor water quality and its effect on wetlands (Reference: *Environmental Protection Agency Wetland Report* EPA 832-R-93-005). Of particular concern is the concentration of nitrogen in water draining from fertilized lands. Too much nitrogen can kill fish and wildlife. Twenty-eight samples of water were taken at random from a lake. The nitrogen concentration (milligrams of nitrogen per liter of water) was determined for each sample.
   (a) Identify the variable.
   (b) Is the variable quantitative or qualitative?
   (c) What is the implied population?

8. *Archaeology: Ireland* The archaeological site of Tara is more than 4000 years old. Tradition states that Tara was the seat of the high kings of Ireland. Because of its archaeological importance, Tara has received extensive study (Reference: *Tara: An Archaeological Survey* by Conor Newman, Royal Irish Academy, Dublin). Suppose an archaeologist wants to estimate the density of ferromagnetic artifacts in the Tara region. For this purpose, a random sample of 55 plots, each of size 100 square meters, is used. The number of ferromagnetic artifacts for each plot is determined.
   (a) Identify the variable.
   (b) Is the variable quantitative or qualitative?
   (c) What is the implied population?

9. *Student Life: Levels of Measurement* Categorize these measurements associated with student life according to level: nominal, ordinal, interval, or ratio.
   (a) Length of time to complete an exam
   (b) Time of first class
   (c) Major field of study
   (d) Course evaluation scale: poor, acceptable, good
   (e) Score on last exam (based on 100 possible points)
   (f) Age of student

10. *Business: Levels of Measurement* Categorize these measurements associated with a robotics company according to level: nominal, ordinal, interval, or ratio.
    (a) Salesperson's performance: below average, average, above average
    (b) Price of company's stock
    (c) Names of new products
    (d) Temperature (°F) in CEO's private office
    (e) Gross income for each of the past 5 years
    (f) Color of product packaging

11. *Fishing: Levels of Measurement* Categorize these measurements associated with fishing according to level: nominal, ordinal, interval, or ratio.
    (a) Species of fish caught: perch, bass, pike, trout
    (b) Cost of rod and reel
    (c) Time of return home
    (d) Guidebook rating of fishing area: poor, fair, good
    (e) Number of fish caught
    (f) Temperature of water in degrees Fahrenheit

12. *Education: Teacher Evaluation* If you were going to apply *statistical methods* to analyze teacher evaluations, which question form, A or B, would be better?
    Form A: In your own words, tell how this teacher compares with other teachers you have had.
    Form B: Use the following scale to rank your teacher as compared with other teachers you have had.

1	2	3	4	5
worst	**below average**	**average**	**above average**	**best**

13. *Critical Thinking* You are interested in the weights of backpacks students carry to class and decide to conduct a study using the backpacks carried by 30 students.
    (a) Give some instructions for weighing the backpacks. Include unit of measure, accuracy of measure, and type of scale.
    (b) Do you think each student asked will allow you to weigh his or her backpack?
    (c) Do you think telling students ahead of time that you are going to weigh their backpacks will make a difference in the weights?

SECTION 1.2

Random Samples

FOCUS POINTS
- Explain the importance of random samples.
- Construct a simple random sample using random numbers.
- Simulate a random process.
- Describe stratified sampling, cluster sampling, systematic sampling, multistage sampling, and convenience sampling.

Simple Random Samples

Eat lamb—20,000 coyotes can't be wrong!

This slogan is sometimes found on bumper stickers in the western United States. The slogan indicates the trouble that ranchers have experienced in protecting their flocks from predators. Based on their experience with this sample of the coyote population, the ranchers concluded that *all* coyotes are dangerous to their flocks and should be eliminated! The ranchers used a special poison bait to get rid of the coyotes. Not only was this poison distributed on ranch land, but with government cooperation it also was distributed widely on public lands.

The ranchers found that the results of the widespread poisoning were not very beneficial. The sheep-eating coyotes continued to thrive while the general population of coyotes and other predators declined. What was the problem? The sheep-eating coyotes the ranchers had observed were not a representative sample of all coyotes. Modern methods of predator control target the sheep-eating coyotes. To a certain extent, the new methods have come about through a closer examination of the sampling techniques used.

In this section, we will examine several widely used sampling techniques. One of the most important sampling techniques is a *simple random sample*.

Simple random sample

> A **simple random sample** of *n* measurements from a population is a subset of the population selected in a manner such that every sample of size *n* from the population has an equal chance of being selected.

In a simple random sample, not only does every sample of the specified size have an equal chance of being selected, but also every individual of the population has an equal chance of being selected. However, the fact that each individual has an equal chance of being selected does not necessarily imply a simple random sample. Remember, for a simple random sample, every sample of the given size must also have an equal chance of being selected.

GUIDED EXERCISE 3 | *Simple random sample*

Is open space around metropolitan areas important? Players of the Colorado Lottery might think so, since some of the proceeds of the game are used to fund open space and outdoor recreational space. To play the game, you pay 1 dollar and choose any six different numbers from the group of numbers 1 through 42. If your group of six numbers matches the winning group of six numbers selected by simple random sampling, then you are a winner of a grand prize of at least 1.5 million dollars.

(a) Is the number 25 as likely to be selected in the winning group of six numbers as the number 5?

⟹   Yes. Because the winning numbers constitute a simple random sample, each number from 1 through 42 has an equal chance of being selected.

Continued

(b) Could all the winning numbers be even?   ➡   Yes, since six even numbers is one of the possible groups of six numbers.

(c) Your friend always plays the numbers   ➡   Yes. In a simple random sample, the listed group of six numbers is *as likely as any* of the 5,245,786 possible groups of six numbers to be selected as the winner. (See Section 5.3 to learn how to compute the number of possible groups of six numbers that can be selected from 42 numbers.)

    1 2 3 4 5 6

    Could she ever win?

How do we get random samples? Suppose you need to know if the emission systems of the latest shipment of Toyotas satisfy pollution-control standards. You want to pick a random sample of 30 cars from this shipment of 500 cars and test them. One way to pick a random sample is to number the cars 1 through 500, write these numbers on cards, mix up the cards, and then draw 30 numbers. The sample will consist of the cars with the chosen numbers. If you mix the cards sufficiently, this procedure produces a random sample.

Random-number table

An easier way to select the numbers is to use a *random-number table*. You can make one yourself by writing the digits 0 through 9 on separate cards and mixing up these cards in a hat. Then draw a card, record the digit, return the card, and mix up the cards again. Draw another card, record the digit, and so on. Table 1 in the Appendix is a ready-made random-number table (adapted from Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates*). Let's see how to pick our random sample of 30 Toyotas by using this random-number table.

EXAMPLE 3   RANDOM-NUMBER TABLE

Use a random-number table to pick a random sample of 30 cars from a population of 500 cars.

SOLUTION: Again, we assign each car a different number between 1 and 500, inclusive. Then we use the random-number table to choose the sample. Table 1 in the Appendix has 35 rows and 10 blocks of five digits each; it can be thought of as a solid mass of digits that has been broken up into rows and blocks for user convenience.

You read the digits by beginning anywhere in the table. We dropped a pin on the table, and the head of the pin landed in row 15, block 5. We'll begin there and list all the digits in that row. If we need more digits, we'll move on to row 16, and so on. The digits we begin with are

   99281   59640   15221   96079   09961   05371

Since the highest number assigned to a car is 500, and this number has three digits, we regroup our digits into blocks of 3:

   992   815   964   015   221   960   790   996   105   371

To construct our random sample, we use the first 30 car numbers we encounter in the random-number table when we start at row 15, block 5. We skip the first three groups—992, 815, and 964—because these numbers are all too large. The next group of three digits is 015, which corresponds to 15. Car number 15 is the first car included in our sample, and the next is car number 221. We skip the next three groups and then include car numbers 105 and 371. To get the rest of the cars in the sample, we continue to the next line and use the random-number table in the same fashion. If we encounter a number we've used before, we skip it.

COMMENT   When we use the term *(simple) random sample,* we have very specific criteria in mind for selecting the sample. One proper method for selecting a simple random sample is to use a computer- or calculator-based random-number generator or a table of random numbers as we have done in the example. The term *random* should not be confused with *haphazard!*

PROCEDURE

HOW TO DRAW A RANDOM SAMPLE

1. Number all members of the population sequentially.
2. Use a table, calculator, or computer to select random numbers from the numbers assigned to the population members.
3. Create the sample by using population members with numbers corresponding to those randomly selected.

Simulation

   Another important use of random-number tables is in *simulation.* We use the word *simulation* to refer to the process of providing numerical imitations of "real" phenomena. Simulation methods have been productive in studying a diverse array of subjects such as nuclear reactors, cloud formation, cardiology (and medical science in general), highway design, production control, shipbuilding, airplane design, war games, economics, and electronics. A complete list would probably include something from every aspect of modern life. In Guided Exercise 4 we'll perform a brief simulation.

> A **simulation** is a numerical facsimile or representation of a real-world phenomenon.

GUIDED EXERCISE 4 | **Simulation**

Use a random-number table to simulate the outcomes of tossing a balanced (that is, fair) penny 10 times.

(a) How many outcomes are possible when you toss a coin once?

⟹   Two—heads or tails

(b) There are several ways to assign numbers to the two outcomes. Because we assume a fair coin, we can assign an even digit to the outcome "heads" and an odd digit to the outcome "tails." Then, starting at block 3 of row 2 of Table 1 in the Appendix, list the first 10 single digits.

⟹   7 1 5 4 9 4 4 8 4 3

(c) What are the outcomes associated with the 10 digits?

⟹   T T T H T H H H H T

(d) If you start in a different block and row of Table 1 in the Appendix, will you get the same sequence of outcomes?

⟹   It is possible, but not very likely. (In Section 5.3 you will learn how to determine that there are 1024 possible sequences of outcomes for 10 tosses of a coin.)

Sampling with replacement

Most statistical software packages, spreadsheet programs, and statistical calculators generate random numbers. In general, these devices sample with replacement. *Sampling with replacement* means that although a number has been selected for the sample, it is *not removed* from the population. Therefore, the same number may be selected for the sample more than once. If you need to sample without replacement, generate more items than you need for the sample. Then sort the sample and remove duplicate values. Specific procedures for generating random samples using the TI-84Plus/TI-83Plus calculator, Excel, Minitab, and SPSS are shown in Using Technology at the end of this chapter. More details are given in the separate Technology Guides for each of these technologies.

Other Sampling Techniques

Although we will assume throughout this text that (simple) random samples are used, other methods of sampling also are widely used. Appropriate statistical techniques exist for these sampling methods, but they are beyond the scope of this text.

Stratified sampling

One of these alternative sampling methods is called *stratified sampling*. Groups or classes inside a population that share a common characteristic are called *strata* (plural of *stratum*). For example, in the population of all undergraduate college students, some strata might be freshmen, sophomores, juniors, or seniors. Other strata might be men or women, in-state students or out-of-state students, and so on. In the method of stratified sampling, the population is divided into at least two distinct strata. Then a (simple) random sample of a certain size is drawn from each stratum, and the information obtained is carefully adjusted or weighted in all resulting calculations.

The groups or strata are often sampled in proportion to their actual percentages of occurrence in the overall population. However, other (more sophisticated) ways of determining the optimal sample size in each stratum may give the best results. In general, statistical analysis and tests based on data obtained from stratified samples are somewhat different from techniques discussed in an introductory course in statistics. Such methods for stratified sampling will not be discussed in this text.

Systematic sampling

Another popular method of sampling is called *systematic sampling*. In this method, it is assumed that the elements of the population are arranged in some natural sequential order. We select a (random) starting point and then select every *k*th element for our sample. For example, people lining up to buy rock concert tickets are "in order." To generate a systematic sample of these people (and ask questions regarding topics such as age, smoking habits, income level, etc.), we could include every fifth person in line. The "starting" person is selected at random from the first five.

The advantage of a systematic sample is that it is easy to select. However, there are dangers in using systematic sampling. When the population is repetitive or cyclic in nature, systematic sampling should not be used. For example, consider a fabric mill that produces dress material. Suppose the loom that produces the material makes a mistake every 17th yard, but we check only every 16th yard with an automated electronic scanner. In this case, a random starting point may or may not result in detection of fabric flaws before a large amount of fabric is produced.

Cluster sampling

Cluster sampling is a method used extensively by government agencies and certain private research organizations. In cluster sampling, we begin by dividing the demographic area into sections. Then we randomly select sections or clusters. Every member of the cluster is included in the sample. For example, in conducting a survey of school children in a large city, we could first randomly select five schools and then include all the children from each selected school.

Multistage samples

Often a population is very large or geographically spread out. In such cases, samples are constructed through a *multistage sample design* of several stages, with the final stage consisting of clusters. For instance, the government Current

Population Survey interviews about 60,000 households across the United States each month by means of a multistage sample design.

For the Current Population Survey, the first stage consists of selecting samples of large geographic areas that do not cross state lines. These areas are further broken down into smaller blocks, which are stratified according to ethnic and other factors. Stratified samples of the blocks are then taken. Finally, housing units in each chosen block are broken into clusters of nearby housing units. A random sample of these clusters of housing units is selected, and each household in the final cluster is interviewed.

Convenience sampling

Convenience sampling simply uses results or data that are conveniently and readily obtained. In some cases, this data may be all that is available, and in many cases, it is better than no information at all. However, convenience sampling does run the risk of being severely biased. For instance, consider a newsperson who wishes to get the "opinions of the people" about a proposed seat tax to be imposed on tickets to all sporting events. The revenues from the seat tax will then be used to support the local symphony. The newsperson stands in front of a classical music store at noon and surveys the first five people coming out of the store who will cooperate. This method of choosing a sample will produce some opinions, and perhaps some human interest stories, but it certainly has bias. It is hoped that the city council will not use these opinions as the sole basis for a decision about the proposed tax. It is good advice to be very cautious indeed when the data come from the method of convenience sampling.

Sampling techniques

Random sampling: Use a simple random sample from the entire population.

Stratified sampling: Divide the entire population into distinct subgroups called strata. The strata are based on a specific characteristic such as age, income, education level, and so on. All members of a stratum share the specific characteristic. Draw random samples from each stratum.

Systematic sampling: Number all members of the population sequentially. Then, from a starting point selected at random, include every kth member of the population in the sample.

Cluster sampling: Divide the entire population into pre-existing segments or clusters. The clusters are often geographic. Make a random selection of clusters. Include every member of each selected cluster in the sample.

Multistage sampling: Use a variety of sampling methods to create successively smaller groups at each stage. The final sample consists of clusters.

Convenience sampling: Create a sample by using data from population members that are readily available.

CRITICAL THINKING

Sampling frame

We call the list of individuals from which a sample is actually selected the *sampling frame*. Ideally, the sampling frame is the entire population. However, from a practical perspective, not all members of a population may be accessible. For instance, a telephone directory used as the sample frame for residential telephone contacts would not include unlisted numbers.

Undercoverage

When the sample frame does not match the population, we have what is called *undercoverage*. In demographic studies, undercoverage could result if the homeless, fugitives from the law, and so forth were not included in the study.

> A **sampling frame** is a list of individuals from which a sample is actually selected.
>
> **Undercoverage** results when population members are omitted from the sample frame.

In general, even when the sampling frame and the population match, a sample is not a perfect representation of a population. Therefore, information drawn from a sample may not exactly match corresponding information from the population. To the extent that sample information does not match the corresponding population information, we have an error, called a *sampling error*.

Sampling error

> A **sampling error** is the difference between measurements from a sample and corresponding measurements from the respective population. It is caused by the fact that the sample does not perfectly represent the population.
>
> A **nonsampling error** is the result of poor sample design, sloppy data collection, faulty measuring instruments, bias in questionnaires, and so on.

Sampling errors do not represent mistakes! They are simply the consequences of using samples instead of populations. However, be alert to nonsampling errors, which may sometimes occur inadvertently.

VIEWPOINT | Extraterrestrial Life?

Do you believe intelligent life exists on other planets? Using methods of random sampling, a Fox News opinion poll found that about 54% of all U.S. men do believe in intelligent life on other planets, whereas only 47% of U.S. women believe there is such life. How could you conduct a random survey of students on your campus regarding belief in extraterrestrial life?

SECTION 1.2 PROBLEMS

1. *Statistical Literacy* Explain the difference between a stratified sample and a cluster sample.

2. *Statistical Literacy* Explain the difference between a simple random sample and a systematic sample.

3. *Statistical Literacy* Marcie conducted a study of the cost of breakfast cereal. She recorded the costs of several boxes of cereal. However, she neglected to take into account the number of servings in each box. Someone told her not to worry because she just had some sampling error. Comment on that advice.

4. *Critical Thinking* Consider the students in your statistics class as the population and suppose they are seated in four rows of 10 students each. To select a sample, you toss a coin. If it comes up heads, you use the 20 students sitting in the first two rows as your sample. If it comes up tails, you use the 20 students sitting in the last two rows as your sample.
   (a) Does every student have an equal chance of being selected for the sample? Explain.
   (b) Is it possible to include students sitting in row 3 with students sitting in row 2 in your sample? Is your sample a simple random sample? Explain.
   (c) Describe a process you could use to get a simple random sample of size 20 from a class of size 40.

5. *Critical Thinking* Suppose you are assigned the number 1, and the other students in your statistics class call out consecutive numbers until each person in the class has his or her own number. Explain how you could get a random sample of four students from your statistics class.
   (a) Explain why the first four students walking into the classroom would not necessarily form a random sample.
   (b) Explain why four students coming in late would not necessarily form a random sample.
   (c) Explain why four students sitting in the back row would not necessarily form a random sample.
   (d) Explain why the four tallest students would not necessarily form a random sample.

6. *Critical Thinking* In each of the following situations, the sampling frame does not match the population, resulting in undercoverage. Give examples of population members that might have been omitted.
   (a) The population consists of all 250 students in your large statistics class. You plan to obtain a simple random sample of 30 students by using the sampling frame of students present next Monday.
   (b) The population consists of all 15-year-olds living in the attendance district of a local high school. You plan to obtain a simple random sample of 200 such residents by using the student roster of the high school as the sampling frame.

7. *Sampling: Random* Use a random-number table to generate a list of 10 random numbers between 1 and 99. Explain your work.

8. *Sampling: Random* Use a random-number table to generate a list of eight random numbers from 1 to 976. Explain your work.

9. *Sampling: Random* Use a random-number table to generate a list of six random numbers from 1 to 8615. Explain your work.

10. *Simulation: Coin Toss* Use a random-number table to simulate the outcomes of tossing a quarter 25 times. Assume that the quarter is balanced (i.e., fair).

11. *Computer Simulation: Roll of a Die* A die is a cube with dots on each face. The faces have 1, 2, 3, 4, 5, or 6 dots. The table below is a computer simulation (from the software package Minitab) of the results of rolling a fair die 20 times.

DATA DISPLAY

ROW	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1	5	2	2	2	5	3	2	3	1	4
2	3	2	4	5	4	5	3	5	3	4

   (a) Assume that each number in the table corresponds to the number of dots on the upward face of the die. Is it appropriate that the same number appears more than once? Why? What is the outcome of the fourth roll?
   (b) If we simulate more rolls of the die, do you expect to get the same sequence of outcomes? Why or why not?

12. *Simulation: Birthday Problem* Suppose there are 30 people at a party. Do you think any two share the same birthday? Let's use the random-number table to simulate the birthdays of the 30 people at the party. Ignoring leap year, let's assume that the year has 365 days. Number the days, with 1 representing January 1, 2 representing January 2, and so forth, with 365 representing December 31. Draw a random sample of 30 days (with replacement). These days represent the birthdays of the people at the party. Were any two of the birthdays the same? Compare your results with those obtained by other students in the class. Would you expect the results to be the same or different?

13. *Education: Test Construction* Professor Gill is designing a multiple-choice test. There are to be 10 questions. Each question is to have five choices for answers. The choices are to be designated by the letters *a, b, c, d,* and *e*. Professor Gill wishes to use a random-number table to determine which letter choice should correspond to the correct answer for a question. Using the number correspondence 1 for *a,* 2 for *b,* 3 for *c,* 4 for *d,* and 5 for *e,* use a random-number table to determine the letter choice for the correct answer for each of the 10 questions.

14. *Education: Test Construction* Professor Gill uses true–false questions. She wishes to place 20 such questions on the next test. To decide whether to place a true statement or a false statement in each of the 20 questions, she uses a random-number table. She selects 20 digits from the table. An even digit tells her to use a true statement. An odd digit tells her to use a false statement. Use a random-number table to pick a sequence of 20 digits, and describe the corresponding sequence of 20 true–false questions. What would the test key for your sequence look like?

15. *Sampling Methods: Benefits Package* An important part of employee compensation is a benefits package, which might include health insurance, life insurance, child care, vacation days, retirement plan, parental leave, bonuses, etc. Suppose you want to conduct a survey of benefits packages available in private businesses in Hawaii. You want a sample size of 100. Some sampling techniques are described below. Categorize each technique as *simple random sample, stratified sample, systematic sample, cluster sample,* or *convenience sample.*
    (a) Assign each business in the Island Business Directory a number, and then use a random-number table to select the businesses to be included in the sample.
    (b) Use postal ZIP Codes to divide the state into regions. Pick a random sample of 10 ZIP Code areas and then include all the businesses in each selected ZIP Code area.
    (c) Send a team of five research assistants to Bishop Street in downtown Honolulu. Let each assistant select a block or building and interview an employee from each business found. Each researcher can have the rest of the day off after getting responses from 20 different businesses.
    (d) Use the Island Business Directory. Number all the businesses. Select a starting place at random, and then use every 50th business listed until you have 100 businesses.
    (e) Group the businesses according to type: medical, shipping, retail, manufacturing, financial, construction, restaurant, hotel, tourism, other. Then select a random sample of 10 businesses from each business type.

16. *Sampling Methods: Health Care* Modern Managed Hospitals (MMH) is a national for-profit chain of hospitals. Management wants to survey patients discharged this past year to obtain patient satisfaction profiles. They wish to use a sample of such patients. Several sampling techniques are described below. Categorize each technique as *simple random sample, stratified sample, systematic sample, cluster sample,* or *convenience sample.*
    (a) Obtain a list of patients discharged from all MMH facilities. Divide the patients according to length of hospital stay (2 days or less, 3–7 days, 8–14 days, more than 14 days). Draw simple random samples from each group.
    (b) Obtain lists of patients discharged from all MMH facilities. Number these patients, and then use a random-number table to obtain the sample.
    (c) Randomly select some MMH facilities from each of five geographic regions, and then include all the patients on the discharge lists of the selected hospitals.
    (d) At the beginning of the year, instruct each MMH facility to survey every 500th patient discharged.
    (e) Instruct each MMH facility to survey 10 discharged patients this week and send in the results.

SECTION 1.3

Introduction to Experimental Design

FOCUS POINTS

- Discuss what it means to take a census.
- Describe simulations, observational studies, and experiments.
- Identify control groups, placebo effects, completely randomized experiments, and randomized block experiments.
- Discuss potential pitfalls that might make your data unreliable.

Planning a Statistical Study

Planning a statistical study and gathering data are essential components of obtaining reliable information. Depending on the nature of the statistical study, a great deal of expertise and resources may be required during the planning stage. In this section, we look at some of the basics of planning a statistical study.

PROCEDURE

> Basic guidelines for planning a statistical study
>
> 1. First, identify the individuals or objects of interest.
> 2. Specify the variables as well as protocols for taking measurements or making observations.
> 3. Determine if you will use an entire population or a representative sample. If using a sample, decide on a viable sampling method.
> 4. In your data collection plan, address issues of ethics, subject confidentiality, and privacy. If you are collecting data at a business, store, college, or other institution, be sure to be courteous and to obtain permission as necessary.
> 5. Collect the data.
> 6. Use appropriate descriptive statistics methods (Chapters 2, 3, and 4) and make decisions using appropriate inferential statistics methods (Chapters 8–11).
> 7. Finally, note any concerns you might have about your data collection methods and list any recommendations for future studies.

Census

One issue to consider is whether to use the entire population in a study or a representative sample. If we use data from the entire population, we have a *census*.

> In a **census,** measurements or observations from the *entire* population are used.

When the population is small and easily accessible, a census is very useful because it gives complete information about the population. However, obtaining a census can be both expensive and difficult. Every 10 years, the U.S. Department of Commerce Census Bureau is required to conduct a census of the United States. However, contacting some members of the population—such as the homeless—is almost impossible. Sometimes members of the population will not respond. In such cases, statistical estimates for the missing responses are often supplied.

Sample

If we use data from only part of the population of interest, we have a *sample*.

> In a **sample,** measurements or observations from *part* of the population are used.

In the previous section, we examined several sampling strategies: simple random, stratified, cluster, systematic, multistage, and convenience. In this text, we will study methods of inferential statistics based on simple random samples.

Simulation

As discussed in Section 1.2, a *simulation* is a numerical facsimile of a real-world phenomenon. Sometimes simulation is called a "dry lab" approach, in the sense that it is a mathematical imitation of a real situation. Advantages of simulation are that numerical and statistical simulations can fit real-world problems extremely well. The researcher can explore procedures through simulation that might be very dangerous in real life.

Experiments and Observation

When gathering data for a statistical study, we want to distinguish between observational studies and experiments.

> In an **observational study,** observations and measurements of individuals are conducted in a way that doesn't change the response or the variable being measured.
>
> In an **experiment,** a *treatment* is deliberately imposed on the individuals in order to observe a possible change in the response or variable being measured.

EXAMPLE 4   Experiment

In 1778, Captain James Cook landed in what we now call the Hawaiian Islands. He gave the islanders a present of several goats, and over the years these animals multiplied into wild herds totaling several thousand. They eat almost anything, including the famous silver sword plant, which was once unique to Hawaii. At one time, the silver sword grew abundantly on the island of Maui (in Haleakala, a national park on that island, the silver sword can still be found), but each year there seemed to be fewer and fewer plants. Biologists suspected that the goats were partially responsible for the decline in the number of plants and conducted a statistical study that verified their theory.

(a) To test the theory, park biologists set up stations in remote areas of Haleakala. At each station, two plots of land similar in soil conditions, climate, and plant count were selected. One plot was fenced to keep out the goats, while the other was not. At regular intervals a plant count was made in each plot. This study involves an *experiment* because a *treatment* (the fence) was imposed on one plot.

(b) The experiment involved two plots at each station. The plot that was not fenced represents the *control* plot. This is the plot on which a treatment was specifically not imposed, although the plot was similar to the fenced plot in every other way.

*Silver sword plant,
Haleakala National Park*

Placebo effect

Statistical experiments are commonly used to determine the effect of a treatment. However, the design of the experiment needs to *control* for other possible causes of the effect. For instance, in medical experiments, the *placebo effect* is the improvement or change that is the result of patients just believing in the treatment, whether or not the treatment itself is effective.

> The **placebo effect** occurs when a subject receives no treatment but (incorrectly) believes he or she is, in fact, receiving treatment and responds favorably.

To account for the placebo effect, patients are divided into two groups. One group receives the prescribed treatment. The other group, called the *control group,* receives a dummy or placebo treatment that is disguised to look like the real treatment. Finally, after the treatment cycle, the medical condition of the patients in the *treatment group* is compared to that of the patients in the control group.

A common way to assign patients to treatment and control groups is by using a random process. This is the essence of a *completely randomized experiment.*
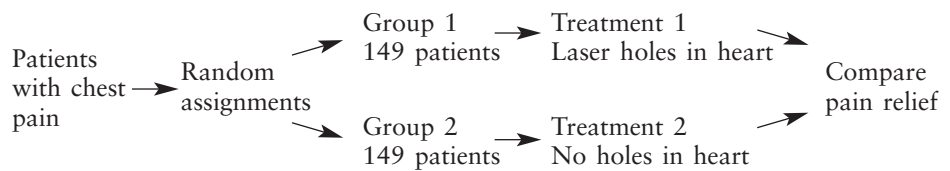
Completely randomized experiment

> A **completely randomized experiment** is one in which a random process is used to assign each individual to one of the treatments.

EXAMPLE 5   COMPLETELY RANDOMIZED EXPERIMENT

Can chest pain be relieved by drilling holes in the heart? For more than a decade, surgeons have been using a laser procedure to drill holes in the heart. Many patients report a lasting and dramatic decrease in angina (chest pain) symptoms. Is the relief due to the procedure, or is it a placebo effect? A recent research project at Lenox Hill Hospital in New York City provided some information about this issue by using a completely randomized experiment. The laser treatment was applied through a less invasive (catheter laser) process. A group of 298 volunteers with severe, untreatable chest pain were randomly assigned to get the laser treatment or not. The patients were sedated but awake. They could hear the doctors discuss the laser process. Each patient thought he or she was receiving the treatment.

The experimental design can be pictured as

Patients with chest pain → Random assignments →
Group 1 149 patients → Treatment 1 Laser holes in heart →
Group 2 149 patients → Treatment 2 No holes in heart →
Compare pain relief

The laser patients did well. But shockingly, the placebo group showed more improvement in pain relief. The medical impacts of this study are still being investigated.
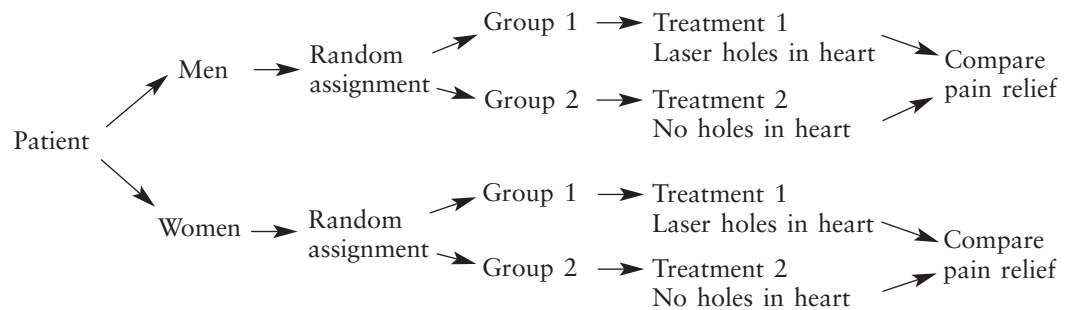
It is difficult to control all the variables that might influence the response to a treatment. One way to control some of the variables is through *blocking.*

> A **block** is a group of individuals sharing some common features that might affect the treatment.

Randomized block design

> In a **randomized block experiment,** individuals are first sorted into blocks, and then a random process is used to assign each individual in the block to one of the treatments.

A randomized block design utilizing gender for blocks in the experiment involving laser holes in the heart would be

```
                              Group 1  →  Treatment 1
                    Random  ↗             Laser holes in heart ↘
          Men  →    assignment                                    Compare
                            ↘  Group 2  →  Treatment 2          ↗  pain relief
                                           No holes in heart
Patient ↗
        ↘
                              Group 1  →  Treatment 1
                    Random  ↗             Laser holes in heart ↘
          Women →   assignment                                    Compare
                            ↘  Group 2  →  Treatment 2          ↗  pain relief
                                           No holes in heart
```

The study cited in Example 5 has many features of good experimental design.

There is a **control group.** This group received a dummy treatment, enabling the researchers to control for the placebo effect. In general, a control group is used to account for the influence of other known or unknown variables that might be an underlying cause of a change in response in the experimental group. Such variables are called **lurking** or **confounding variables.**

Randomization is used to assign individuals to the two treatment groups. This helps prevent bias in selecting members for each group.

Replication of the experiment on many patients reduces the possibility that the differences in pain relief for the two groups occurred by chance alone.

Double-blind experiment

Many experiments are also *double-blind.* This means that neither the individuals in the study nor the observers know which subjects are receiving the treatment. Double-blind experiments help control for subtle biases that a doctor might pass on to a patient.

GUIDED EXERCISE 5 | **Collecting data**

Which technique for gathering data (sampling, experiment, simulation, or census) do you think might be the most appropriate for the following studies?

(a) Study of the effect of stopping the cooling process of a nuclear reactor.

⇨ Simulation, since you probably do not want to risk a nuclear meltdown.

(b) Study of the amount of time college students taking a full course load spend watching television.

⇨ Sampling and using an observational study would work well. Notice that obtaining the information from a student probably will not change the amount of time the student spends watching television.

(c) Study of the effect on bone mass of a calcium supplement given to young girls.

⇨ Experimentation. A study by Tom Lloyd reported in the *Journal of the American Medical Association* utilized 94 young girls. Half were randomly selected and given a placebo. The other half were given calcium supplements to bring their daily calcium intake up to about 1400 milligrams per day. The group getting the experimental treatment of calcium gained 1.3% more bone mass in a year than the girls getting the placebo.

(d) Study of the credit hours load of *each* student enrolled at your college at the end of the drop/add period this semester.

⇨ Census. The registrar can obtain records for *every* student.

Surveys

Once you decide whether you are going to use sampling, census, observation, or experiments, a common means to gather data about people is to ask them questions. This process is the essence of *surveying*. Sometimes the possible responses are simply yes or no. Other times the respondents choose a number on a scale that represents their feelings from, say, strongly disagree to strongly agree. Such a scale is called a *Likert scale*. In the case of an open-ended, discussion-type response, the researcher must determine a way to convert the response to a category or number.

A number of issues can arise when using a survey.

Some potential pitfalls of a survey

Nonresponse: Individuals either cannot be contacted or refuse to participate. Nonresponse can result in significant undercoverage of a population.

Truthfulness of response: Respondents may lie intentionally or inadvertently.

Faulty recall: Respondents may not accurately remember when or whether an event took place.

Hidden bias: The question may be worded in such a way as to elicit a specific response. The order of questions might lead to biased responses. Also, the number of responses on a Likert scale may force responses that do not reflect the respondent's feelings or experience.

Vague wording: Words such as "often," "seldom," and "occasionally" mean different things to different people.

Interviewer influence: Factors such as tone of voice, body language, dress, gender, authority, and ethnicity of the interviewer might influence responses.

Voluntary response: Individuals with strong feelings about a subject are more likely than others to respond. Such a study is interesting but not reflective of the population.

Lurking and confounding variables

Sometimes our goal is to understand the cause-and-effect relationships between two or more variables. Such studies can be complicated by *lurking variables* or *confounding variables*.

A **lurking variable** is one for which no data have been collected but that nevertheless has influence on other variables in the study.

Two variables are **confounded** when the effects of one cannot be distinguished from the effects of the other. Confounding variables may be part of the study, or they may be outside lurking variables.

For instance, consider a study involving just two variables, amount of gasoline used to commute to work and time to commute to work. Level of traffic congestion is a likely lurking variable that increases both of the study variables. In a study involving several variables such as grade point average, difficulty of courses, IQ, and available study time, some of the variables might be confounded. For instance, students with less study time might opt for easier courses.

Generalizing results

Some researchers want to generalize their findings to a situation of wider scope than that of the actual data setting. The true scope of a new discovery must be determined by repeated studies in various real-world settings. Statistical experiments showing that a drug had a certain effect on a collection of laboratory rats do not guarantee that the drug will have a similar effect on a herd of wild horses in Montana.

Study sponsor

The sponsorship of a study is another area of concern. Subtle bias may be introduced. For instance, if a pharmaceutical company is paying freelance researchers to work on a study, the researchers may dismiss rare negative findings about a drug or treatment.

GUIDED EXERCISE 6 | *Cautions about data*

Comment on the usefulness of the data collected as described.

(a) A uniformed police officer interviews a group of 20 college freshmen. She asks each one his or her name and then if he or she has used an illegal drug in the last month.

⇒ Respondents may not answer truthfully. Some may refuse to participate.

(b) Jessica saw some data that show that cities with more low-income housing have more homeless people. Does building low-income housing cause homelessness?

⇒ There may be some confounding or lurking variables, such as the size of the city. Larger cities may have more low-income housing and more homeless.

(c) A survey about quality of food in the student cafeteria was conducted by having forms available for customers to pick up at the cash register. A drop box for completed forms was available outside the cafeteria.

⇒ The voluntary response likely will produce more negative comments.

(d) Extensive studies on coronary problems were conducted using men over age 50 as the subjects.

⇒ Conclusions for men over age 50 may or may not generalize to other age and gender groups. These results may be useful for women or younger people, but studies specifically involving these groups may need to be performed.

Choosing Data Collection Techniques

We've briefly discussed three common techniques for gathering data: observational studies, experiments, and surveys. Which technique is best? The answer depends on the number of variables of interest and the level of confidence needed regarding statements of relationships among the variables.

- Surveys may be the best choice for gathering information across a wide range of many variables. Many questions can be included in a survey. However, great care must be taken in the construction of the survey instrument and in the administration of the survey. Nonresponse and other issues discussed earlier can introduce bias.

- Observational studies are the next most convenient technique for gathering information on many variables. Protocols for taking measurements or recording observations need to be specified carefully.

- Experiments are the most stringent and restrictive data-gathering technique. They can be time-consuming, expensive, and difficult to administer. In experiments, the goal is often to study the effects of changing only one variable at a time. Because of the requirements, the number of variables may be more limited. Experiments must be designed carefully to ensure that the resulting data are relevant to the research questions.

COMMENT   An experiment is the best technique for reaching valid conclusions. By carefully controlling for other variables, the effect of changing one variable on a treatment group and comparing it to a control group yields results carrying high confidence.

The next most effective technique for obtaining results that have high confidence is the use of observational studies. Care must be taken that the act of observation does not change the behavior being measured or observed.

The least effective technique for drawing conclusions is the survey. Surveys have many pitfalls and by their nature cannot give exceedingly precise results. A medical survey asking patients if they feel better after taking a specific drug gives some information, but not precise information about the drug's effects. However, surveys are widely used to gauge attitudes, gather demographic information, study social and political trends, and so on.

VIEWPOINT | Is the Placebo Effect a Myth?

Henry Beecher, former Chief of Anesthesiology at Massachusetts General Hospital, published a paper in the Journal of the American Medical Association *(1955) in which he claimed that the placebo effect is so powerful that about 35% of patients would improve simply if they believed a dummy treatment (placebo) was real. However, two Danish medical researchers refuted this widely accepted claim in the* New England Journal of Medicine. *They said the placebo effect is nothing more than a "regression effect," referring to a well-known statistical observation that patients who feel especially bad one day will almost always feel better the next day, no matter what is done for them. However, other respected statisticians question the findings of the Danish researchers. Regardless of the new controversy surrounding the placebo effect, medical researchers agree that placebos are still needed in clinical research. Double-blind research using placebos prevents the researchers from inadvertently biasing results.*

SECTION 1.3 PROBLEMS

1. *Statistical Literacy* A study involves three variables: income level, hours spent watching TV per week, and hours spent at home on the Internet per week. List some ways the variables might be confounded.

2. *Statistical Literacy* Consider a completely randomized experiment in which a control group is given a placebo for congestion relief and a treatment group is given a new drug for congestion relief. Describe a double-blind procedure for this experiment and discuss some benefits of such a procedure.

3. *Ecology: Gathering Data* Which technique for gathering data (observational study or experiment) do you think was used in the following studies?
   (a) The Colorado Division of Wildlife netted and released 774 fish at Quincy Reservoir. There were 219 perch, 315 blue gill, 83 pike, and 157 rainbow trout.
   (b) The Colorado Division of Wildlife caught 41 bighorn sheep on Mt. Evans and gave each one an injection to prevent heartworm. A year later, 38 of these sheep did not have heartworm, while the other three did.
   (c) The Colorado Division of Wildlife imposed special fishing regulations on the Deckers section of the South Platte River. All trout under 15 inches had to be released. A study of trout before and after the regulation went into effect showed that the average length of a trout increased by 4.2 inches after the new regulation.
   (d) An ecology class used binoculars to watch 23 turtles at Lowell Ponds. It was found that 18 were box turtles and 5 were snapping turtles.

4. *General: Gathering Data* Which technique for gathering data (sampling, experiment, simulation, or census) do you think was used in the following studies?
   (a) An analysis of a sample of 31,000 patients from New York hospitals suggests that the poor and the elderly sue for malpractice at one-fifth the rate of wealthier patients (*Journal of the American Medical Association*).
   (b) The effects of wind shear on airplanes during both landing and takeoff were studied by using complex computer programs that mimic actual flight.
   (c) A study of all league football scores attained through touchdowns and field goals was conducted by the National Football League to determine whether field goals account for more scoring events than touchdowns (*USA Today*).
   (d) An Australian study included 588 men and women who already had some precancerous skin lesions. Half got a skin cream containing a sunscreen with a sun protection factor of 17; half got an inactive cream. After 7 months, those using the sunscreen with the sun protection had fewer new precancerous skin lesions (*New England Journal of Medicine*).

5. *General: Completely Randomized Experiment* How would you use a completely randomized experiment in each of the following settings? Is a placebo being used or not? Be specific and give details.
   (a) A veterinarian wants to test a strain of antibiotic on calves to determine their resistance to common infection. In a pasture are 22 newborn calves. There is enough vaccine for 10 calves. However, blood tests to determine resistance to infection can be done on all calves.
   (b) The Denver Police Department wants to improve its image with teenagers. A uniformed officer is sent to a school one day a week for 10 weeks. Each day the officer visits with students, eats lunch with students, attends pep rallies, and so on. There are 18 schools, but the police department can visit only half of these schools this semester. A survey regarding how teenagers view police is sent to all 18 schools at the end of the semester.
   (c) A skin patch contains a new drug to help people quit smoking. A group of 75 cigarette smokers have volunteered as subjects to test the new skin patch. For one month, 40 of the volunteers receive skin patches with the new drug. The other volunteers receive skin patches with no drugs. At the end of two months, each subject is surveyed regarding his or her current smoking habits.

6. *Surveys: Manipulation* The *New York Times* did a special report on polling that was carried in papers across the nation. The article pointed out how readily the results of a survey can be manipulated. Some features that can influence the results of a poll include the following: the number of possible responses, the phrasing of the question, the sampling techniques used (voluntary response or sample designed to be representative), the fact that words may mean different things to different people, the questions that precede the question of interest, and finally, the fact that respondents can offer opinions on issues they know nothing about.
   (a) Consider the expression "over the last few years." Do you think that this expression means the same time span to everyone? What would be a more precise phrase?
   (b) Consider this question: "Do you think fines for running stop signs should be doubled?" Do you think the response would be different if the question "Have you ever run a stop sign?" preceded the question about fines?
   (c) Consider this question: "Do you watch too much television?" What do you think the responses would be if the only responses possible were yes or no? What do you think the responses would be if the possible responses were rarely, sometimes, or frequently?

7. *Critical Thinking* An agricultural study is conducted to compare the harvest volumes of two types of barley. The site for the experiment is bordered by a river. The field is divided into eight plots of approximately the same size. The experiment calls for the plots to be blocked into four plots per block. Then, two plots of each block will be randomly assigned to one of the two barley types.

Two blocking schemes are shown below, with one block indicated by the white region and the other by the grey region. Which blocking scheme, A or B, would be better? Explain.
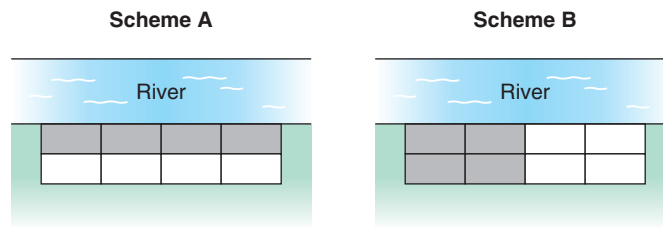
Chapter Review

SUMMARY

In this chapter, you've seen that statistics is the study of how to collect, organize, analyze, and interpret numerical information from populations or samples. This chapter discussed some of the features of data and ways to collect data. In particular, the chapter discussed

- Individuals or subjects of a study and the variables associated with those individuals
- Data classification as qualitative or quantitative, and levels of measurement of data
- Sample and population data. Summary measurements from sample data are called statistics, and those from population data are called parameters.

- Sampling strategies, including simple random, stratified, systematic, multistage, and convenience. Inferential techniques presented in this text are based on simple random samples.
- Methods of obtaining data: Use of a census, simulation, observational studies, experiments, and surveys
- Concerns: Undercoverage of a population, nonresponse, bias in data from surveys and other factors, effects of confounding or lurking variables on other variables, generalization of study results beyond the population of the study, and study sponsorship

IMPORTANT WORDS & SYMBOLS

Section 1.1*
Statistics
Individual
Variable
Quantitative variable
Qualitative variable
Population data
Sample data
Parameter
Statistic
Levels of measurement
　Nominal
　Ordinal
　Interval
　Ratio
Descriptive statistics
Inferential statistics

Section 1.2
Simple random sample
Random-number table

Simulation
Sampling with replacement
Stratified sample
Systematic sample
Cluster sample
Multistage sample
Convenience sample
Sampling frame
Undercoverage
Sample error
Nonsample error

Section 1.3
Census
Observational study
Experiment
Placebo effect
Completely randomized experiment
Block
Randomized block experiment
Double-blind experiment

*Indicates section of first appearance.

Control group	Replication
Treatment group	Survey
Confounding variable	Nonresponse
Lurking variable	Voluntary response
Randomization	Hidden bias

VIEWPOINT | Is Chocolate Good for Your Heart?

A study of 7841 Harvard alumni showed that the death rate was 30% lower in those who ate candy compared with those who abstained. It turns out that candy, especially chocolate, contains antioxidants that help slow the aging process. Also, chocolate, like aspirin, reduces the activity of blood platelets that contribute to plaque and blood clotting. Furthermore, chocolate seems to raise levels of high-density lipoprotein (HDL), the good cholesterol. However, these results are all preliminary. The investigation is far from complete. A wealth of information on this topic was published in the August 2000 issue of the Journal of Nutrition. *Statistical studies and reliable experimental design are indispensable in this type of research.*

CHAPTER REVIEW PROBLEMS

1. *Statistical Literacy* You are conducting a study of students doing work-study jobs on your campus. Among the questions on the survey instrument are:

   A. How many hours are you scheduled to work each week? Answer to the nearest hour.

   B. How applicable is this work experience to your future employment goals?

   Respond using the following scale: 1 = not at all, 2 = somewhat, 3 = very

   (a) Suppose you take random samples from the following groups: freshmen, sophomores, juniors, and seniors. What kind of sampling technique are you using (simple random, stratified, systematic, cluster, multistage, convenience)?

   (b) Describe the individuals of this study.

   (c) What is the variable for question A? Classify the variable as qualitative or quantitative. What is the level of the measurement?

   (d) What is the variable for question B? Classify the variable as qualitative or quantitative. What is the level of the measurement?

   (e) Is the proportion of responses "3 = very" to question B a statistic or a parameter?

   (f) Suppose only 40% of the students you selected for the sample respond. What is the nonresponse rate? Do you think the nonresponse rate might introduce bias into the study? Explain.

   (g) Would it be appropriate to generalize the results of your study to all work-study students in the nation? Explain.

2. *Radio Talk Show: Sample Bias* A radio talk show host asked listeners to respond either yes or no to the question, Is the candidate who spends the most on a campaign the most likely to win? Fifteen people called in and nine said yes. What is the implied population? What is the variable? Can you detect any bias in the selection of the sample?

3. *Simulation: TV Habits* One cable station knows that approximately 30% of its viewers have TIVO and can easily skip over advertising breaks. You are to design a simulation of how a random sample of seven station viewers would respond to the question, "Do you have TIVO?" How would you assign the random digits 0 through 9 to the responses "Yes" and "No" to the TIVO question? Use your random digit assignment and the random-number table to generate the responses from a random sample of seven station viewers.

4. *General: Type of Sampling* Categorize the type of sampling (simple random, stratified, systematic, cluster, or convenience) used in each of the following situations.
   (a) To conduct a preelection opinion poll on a proposed amendment to the state constitution, a random sample of 10 telephone prefixes (first three digits of the phone number) was selected, and all households from the phone prefixes selected were called.
   (b) To conduct a study on depression among the elderly, a sample of 30 patients in one nursing home was used.
   (c) To maintain quality control in a brewery, every 20th bottle of beer coming off the production line was opened and tested.
   (d) Subscribers to the magazine *Sound Alive* were assigned numbers. Then a sample of 30 subscribers was selected by using a random-number table. The subscribers in the sample were invited to rate new compact disc players for a "What the Subscribers Think" column.
   (e) To judge the appeal of a proposed television sitcom, a random sample of 10 people from each of three different age categories was selected and those chosen were asked to rate a pilot show.

5. *General: Gathering Data* Which technique for gathering data (observational study or experiment) do you think was used in the following studies? Explain.
   (a) The U.S. Census Bureau tracks population age. In 1900, the percentage of the population that was 19 years old or younger was 44.4%. In 1930, the percentage was 38.8%; in 1970, the percentage was 37.9%; and in 2000, the percentage in the age group was down to 28.5% (*The First Measured Century*, T. Caplow, L. Hicks, B. J. Wattenberg).
   (b) After receiving the same lessons, a class of 100 students was randomly divided into two groups of 50 each. One group was given a multiple-choice exam covering the material in the lessons. The other group was given an essay exam. The average test scores for the two groups were then compared.

6. *General: Experiment* How would you use a completely randomized experiment in each of the following settings? Is a placebo being used or not? Be specific and give details.
   (a) A charitable nonprofit organization wants to test two methods of fundraising. From a list of 1000 past donors, half will be sent literature about the successful activities of the charity and asked to make another donation. The other 500 donors will be contacted by phone and asked to make another donation. The percentage of people from each group who make a new donation will be compared.
   (b) A tooth-whitening gel is to be tested for effectiveness. A group of 85 adults have volunteered to participate in the study. Of these, 43 are to be given a gel that contains the tooth-whitening chemicals. The remaining 42 are to be given a similar-looking package of gel that does not contain the tooth-whitening chemicals. A standard method will be used to evaluate the whiteness of teeth for all participants. Then the results for the two groups will be compared. How could this experiment be designed to be double-blind?
   (c) Consider the experiment described in part (a). Describe how you would use a randomized block experiment with blocks based on age. Use three blocks: donors under 30 years old, donors 30 to 59 years old, donors 60 and over.

7. *Student Life: Data Collection Project* Make a statistical profile of your own statistics class. Items of interest might be
   (a) Height, age, gender, pulse, number of siblings, marital status
   (b) Number of college credit hours completed (as of beginning of term); grade point average
   (c) Major; number of credit hours enrolled in this term
   (d) Number of scheduled hours working per week
   (e) Distance from residence to first class; time it takes to travel from residence to first class
   (f) Year, make, and color of car usually driven

What directions would you give to people answering these questions? For instance, how accurate should the measurements be? Should age be recorded as of last birthday?

8. *Census: Web Site* *Census and You,* a publication of the Census Bureau, indicates that "Wherever your Web journey ends up, it should start at the Census Bureau's site." Visit the Online Study Center at **www.cengage.com/statistics/Brase/UBS5e** and find a link to the Census Bureau's site, as well as to Fedstats, another extensive site offering links to federal data. The Census Bureau site touts itself as the source of "official statistics." But it is willing to share the spotlight. The web site now has links to other "official" sources: other federal agencies, foreign statistical agencies, and state data centers. If you have access to the Internet, check out the Census Bureau's site.

9. *Focus Problem: Fireflies* Suppose you are conducting a study to compare firefly populations exposed to normal daylight/darkness conditions with firefly populations exposed to continuous light (24 hours a day). You set up two firefly colonies in a laboratory environment. The two colonies are identical except that one colony is exposed to normal daylight/darkness conditions and the other is exposed to continuous light. Each colony is populated with the same number of mature fireflies. After 72 hours, you count the number of living fireflies in each colony.
   (a) Is this an experiment or an observation study? Explain.
   (b) Is there a control group? Is there a treatment group?
   (c) What is the variable TI in this study?
   (d) What is the level of measurement (nominal, interval, ordinal, or ratio) of the variable?

DATA HIGHLIGHTS: GROUP PROJECTS

1. Use a random-number table or random-number generator to simulate tossing a fair coin 10 times. Generate 20 such simulations of 10 coin tosses. Compare the simulations. Are there any strings of 10 heads? of 4 heads? Does it seem that in most of the simulations half the outcomes are heads? half are tails? In Chapter 6, we will study the probabilities of getting from 0 to 10 heads in such a simulation.

2. Use a random-number table or random-number generator to generate a random sample of 30 distinct values from the set of integers from 1 to 100. Instructions for doing this using the TI-84Plus/TI-83Plus, Excel, Minitab, or SPSS are given in Using Technology at the end of this chapter. Generate five such samples. How many of the samples include the number 1? the number 100? Comment about the differences among the samples. How well do the samples seem to represent the numbers between 1 and 100?

LINKING CONCEPTS: WRITING PROJECTS

Discuss each of the following topics in class or review the topics on your own. Then write a brief but complete essay in which you summarize the main points. Please include formulas and graphs as appropriate.

1. What does it mean when we say that we are going to use a sample to draw an inference about a population? Why is a random sample so important to this process? If we wanted a random sample of students in the cafeteria, why couldn't we just take the students who order Diet Pepsi with their lunch? Comment on the statement, "A random sample is like a miniature population, whereas samples that are not random are likely to be biased." Why would the students who order Diet Pepsi with lunch not be a random sample of students in the cafeteria?

2. In your own words, explain the differences among the following sampling techniques: simple random sample, stratified sample, systematic sample, cluster sample, multistage sample, and convenience sample. Describe situations in which each type might be useful.

USING TECHNOLOGY

General spreadsheet programs such as Microsoft Excel, specific statistical software packages such as Minitab or SPSS, and graphing calculators such as the TI-84Plus and TI-83Plus all offer computing support for statistical methods. Applications in this section may be completed using software or calculators with statistical functions. Select keystroke or menu choices are shown for the TI-84Plus and TI-83Plus calculators, Minitab, Excel, and SPSS in the Technology Hints portion of this section. More details can be found in the software-specific Technology Guide that accompanies this text.

Applications

Most software packages sample *with replacement*. That is, the same number may be used more than once in the sample. If your applications require sampling without replacement, draw more items than you need. Then use sort commands in the software to put the data in order, and delete repeated data.

1. Simulate the results of tossing a fair die 18 times. Repeat the simulation. Are the results the same? Did you expect them to be the same? Why or why not? Do there appear to be equal numbers of outcomes 1 through 6 in each simulation? In Chapter 5, we will encounter the law of large numbers, which tells us that we would expect equal numbers of particular outcomes only when the simulation is very large.

2. A college has 5000 students, and the registrar wishes to use a random sample of 50 students to examine credit hour enrollment for this semester. Write a brief description of how a random sample can be drawn. Draw a random sample of 50 students. Are you sampling with or without replacement?

Technology Hints: Random Numbers

TI-84Plus/TI-83Plus

To select a random set of integers between two specified values, press the **MATH** key and highlight **PRB** with **5:randInt** (low value, high value, sample size). Press Enter and fill in the low value, high value, and sample size. To store the sample in list L1, press the **STO➡** key and then L1. The screen display shows two random samples of size 5 drawn from the integers between 1 and 100.

```
randInt(1,100,5)
{63 89 13 46 47}
randInt(1,100,5)
{29 82 99 50 41}
```

Excel

To select a random number between two specified values, type the command **=Randbetween(low value, high value)** in the formula bar. Alternatively, access a dialogue box for the command by clicking on the **paste function** f_x key on the menu bar. Then choose **All** in the left drop-down menu and **RANDBETWEEN** in the right menu. Fill in the dialogue box.

A5	▼	=	=RANDBETWEEN(1,100)	
	A	B	C	D	E
1	99				
2	58				
3	69				
4	86				
5	13				

Minitab

To generate random integers between specified values, use the menu selection **Calc ➤ Random Data ➤ Integer.** Fill in the dialogue box to get five random numbers between 1 and 100.

	C1
↓	
1	8
2	35
3	33
4	9
5	15

Worksheet 2 ***

SPSS

SPSS is a research statistical package for the social sciences. Data are entered in the data editor, which has a spreadsheet format. In the data editor window, you have a choice of data view (default) or variable view. In the variable view, you name variables, declare type (numeric for measurements, string for category), determine format, and declare measurement type. The choices for measurement type are scale (for ratio or interval data), ordinal, or nominal. Once you have entered data, you can use the menu bar at the top of the screen to select activities, graphs, or analysis appropriate to the data.

SPSS supports several random sample activities. In particular, you can select a random sample from an existing data set or from a variety of probability distributions.

Selecting a random integer between two specified values involves several steps. First, in the data editor, enter the sample numbers in the first column. For instance, to generate five random numbers, list the values 1 through 5 in the first column. Notice that the label for the first column is now var00001. SPSS does not have a direct function for selecting a random sample of integers. However, there is a function for sampling values from the uniform distribution of all real numbers between two specified values. We will use that function and then truncate the values to obtain a random sample of integers between two specified values.

Use the menu options **Transform ➤ Compute.** In the dialog box, type in var00002 as the target variable. Then, in the function box, select the function **RV.UNIFORM(min, max).** Use 1 as the minimum and 101 as the maximum. The maximum is 101 because numbers between 100 and 101 truncate to 100.
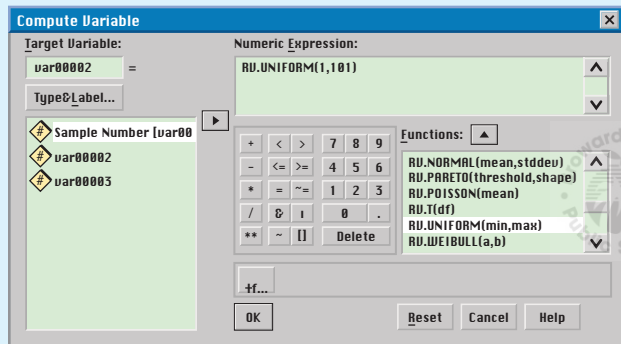
The random numbers from the uniform distribution now appear in the second column under var00002. You can visually truncate the values to obtain random integers. However, if you want SPSS to truncate the values for you, you can again use the menu choices **Transform ➤ Compute.** In the dialogue box, enter var00003 for the target variable. From the functions box, select **TRUNC(numexpr).** Use var00002 in place of numexpr. The random integers between 1 and 100 appear in the third column under var00003.

	var00001	var00002	var00003
1	1.00	14.96	14
2	2.00	44.13	44
3	3.00	62.22	62
4	4.00	30.08	30
5	5.00	16.57	16

33

Honors Probabality and Statistics Summer Assignment

Multiple Choice
Identify the choice that best completes the statement or answers the question.

____   1. Data may be classified by one of the four levels of measurement. What is the name of the lowest
level?
A)   interval
B)   nominal
C)   ordinal
D)   ratio
E)   simple

____   2. Identify the variable in the information below.
What is the average miles per gallon (mpg) for all new cars? Using Consumer Reports, a
random sample of 35 new cars gave an average of 21.1 mpg.
A)   number of new cars surveyed
B)   cars
C)   miles
D)   miles per gallon
E)   gallons

____   3. Identify the variable in the information below.
The archeological site of Tara is more than 4000 years
old. Tradition states that Tara was the seat of the high kings of Ireland. Because of its archeological
importance, Tara has received extensive study (Reference:
Tara: An Archeological survey by Conor Newman, Royal Irish
Academy, Dublin). Suppose an archeologist wants to
estimate the density of ferromagnetic artifacts in the
Tara region. For this purpose, a random sample of 55
plots, each of size 100 square meters, is used. The number
of ferromagnetic artifacts for each plot is determined.
A)   number of plots
B)   size of the plots
C)   none of these choices
D)   number of ferromagnetic artifacts per 100 square meters
E)   number of plots as well as size of the plots

____   4. Identify the variable in the information below.
        Government agencies carefully monitor water quality and its effect on wetlands (Reference:
        Environment Protection Agency Wetland Report EPA 832-R-93-005). Of particular concern is the
        concentration of nitrogen in water draining from fertilized lands. Too much nitrogen can kill fish
        and wildlife. Twenty-eight samples of water were taken at random from a lake. The nitrogen
        concentration (milligrams of nitrogen per liter of water) was determined for each sample.
        A)   quality of water
        B)   lakes
        C)   none of these choices
        D)   nitrogen concentration
        E)   quantity of water

____   5. Identify the variable in the information below.
        The archeological site of Tara is more than 4000 years
        old. Tradition states that Tara was the seat of the high kings of Ireland. Because of its archeological
        importance, Tara has received extensive study (Reference:
        Tara: An Archeological survey by Conor Newman, Royal Irish
        Academy, Dublin). Suppose an archeologist wants to
        estimate the density of ferromagnetic artifacts in the
        Tara region. For this purpose, a random sample of 55
        plots, each of size 100 square meters, is used. The number
        of ferromagnetic artifacts for each plot is determined.
        A)   number of plots
        B)   size of the plots
        C)   none of these choices
        D)   number of ferromagnetic artifacts per 100 square meters
        E)   number of plots as well as size of the plots

____   6. Identify the variable in the information below.
        USA Today reported that 44.9% of those surveyed (1261 adults) ate in a fast-food restaurant
        from one to three times each week.
        A)   fast-food restaurant as well as response regarding frequency of eating at fast-food
             restaurants
        B)   adults surveyed
        C)   fast-food restaurants
        D)   response regarding frequency of eating at fast-food restaurants
        E)   none of these choices

____ 7. Identify the implied population in the information below.
Government agencies carefully monitor water quality and
its effect on wetlands (Reference: Environment
Protection Agency Wetland Report EPA 832-R-93-005). Of
particular concern is the concentration of nitrogen in water
draining from fertilized lands. Too much nitrogen can
kill fish and wildlife. Twenty-eight samples of water
were taken at random from a lake. The nitrogen
concentration (milligrams of nitrogen per liter of water) was
determined for each sample. The variable in this
information is nitrogen concentration (mg nitrogen/l water).
A) number of fish
B) samples of water taken at random
C) none of these choices
D) nitrogen concentration (mg nitrogen/l water) in the entire lake
E) samples of water taken at random

____ 8. Identify the implied population in the information below.
The archeological site of Tara is more than 4000 years old.
Tradition states that Tara was the seat of the high kings of Ireland.
Because of its archeological importance, Tara has received
extensive study (Reference: Tara: An Archeological survey by Conor
Newman, Royal Irish Academy, Dublin). Suppose an archeologist wants
to estimate the density of ferromagnetic artifacts in the Tara
region. For this purpose, a random sample of 55 plots, each of size
100 square meters, is used. The number of ferromagnetic artifacts
for each plot is determined. The variable in this information is
number of ferromagnetic artifacts per 100 square meters.
A) number of ferromagnetic artifacts per 100-square-meters
B) density of ferromagnetic artifacts in some of the plots
C) none of these choices
D) number of ferromagnetic artifacts per each distinct 100-square-meters plot in the
Tara region
E) number of ferromagnetic artifacts per 100-square-meters in some plots
in the Tara region

____   9. Identify the implied population in the information below.
            USA Today reported that 44.9% of those surveyed (1261 adults) ate in a fast-food restaurant
            from one to three times each week. The variable in this information is response regarding
            frequency of eating at fast-food restaurants.
            A) all fast-food restaurants in the US
            B) response regarding frequency of eating at fast food restaurants
            C) none of these choices
            D) responses for all adults in the US
            E) responses for some of the adults in the US

____   10. Identify the implied population in the information below.
            What is the average miles per gallon (mpg) for all new cars? Using Consumer
            Reports, a random sample of 35 new cars gave an average of 21.1 mpg. The variable in
            this information is miles per gallon.
            A) miles per gallon
            B) all new cars
            C) neither qualitative nor quantitative
            D) gasoline mileage for all new cars
            E) gasoline mileage for some of the new cars

____   11. Identify whether the variable in the information below is qualitative or quantitative.
            The archeological site of Tara is more than 4000 years
            old. Tradition states that Tara was the seat of the high kings of Ireland. Because of its archeological
            importance, Tara has received extensive study (Reference:
            Tara: An Archeological survey by Conor Newman, Royal Irish
            Academy, Dublin). Suppose an archeologist wants to
            estimate the density of ferromagnetic artifacts in the
            Tara region. For this purpose, a random sample of 55
            plots, each of size 100 square meters, is used. The number
            of ferromagnetic artifacts for each plot is determined.
            A) quantitative
            B) qualitative
            C) neither qualitative nor quantitative
            D) qualitative as well as quantitative
            E) Information does not have any variable.

____   12. Identify whether the variable in the information below is qualitative or quantitative.
            Average of height of all students in a class of 50 is 5 feet.
            A) quantitative
            B) qualitative
            C) qualitative as well as quantitative
            D) neither qualitative nor quantitative
            E) Information does not have any variable.

____ 13. Identify whether the variable in the information below is qualitative or quantitative.
Government agencies carefully monitor water quality and its effect on wetlands (Reference:
Environment Protection Agency Wetland Report EPA 832-R-93-005). Of particular concern is the
concentration of nitrogen in water draining from fertilized lands. Too much nitrogen can kill fish
and wildlife. Twenty-eight samples of water were taken at random from a lake. The nitrogen
concentration (milligrams of nitrogen per liter of water) was determined for each sample.
A) quantitative
B) qualitative
C) qualitative as well as quantitative
D) neither qualitative nor quantitative
E) Information does not have any variable.

____ 14. Identify whether the variable in the information below is qualitative or quantitative.
USA Today reported that 44.9% of those surveyed (1261 adults) ate in a fast-food restaurant
from one to three times each week.
A) qualitative
B) quantitative
C) neither qualitative nor quantitative
D) qualitative as well as quantitative
E) Information does not have any variable.

____ 15. Identify the level of measurement corresponding to the data "Gross income for
each of the past 5 years" associated with robotics.
A) interval
B) ordinal
C) ratio
D) none of these choices
E) nominal

____ 16. Identify the level of measurement corresponding to the data "Time of first class" associated with
student life.
A) none of these choices
B) ordinal
C) interval
D) nominal
E) ratio

____ 17. Identify the level of measurement corresponding to the data "Temperature of water" associated
with fishing.
A) none of these choices
B) nominal
C) interval
D) ordinal
E) ratio

____ 18. Identify the level of measurement corresponding to the data "Temperature (°F) in CEO's private office" associated with robotics.
A) ratio
B) nominal
C) interval
D) ordinal
E) none of these choices

____ 19. Identify the level of measurement corresponding to the data "Number of fish caught" associated with fishing.
A) ordinal
B) interval
C) ratio
D) none of these choices
E) nominal

____ 20. Identify the level of measurement corresponding to the data "Guidebook rating area: poor, fair, good" associated with fishing.
A) nominal
B) ratio
C) ordinal
D) interval
E) none of these choices

____ 21. Identify the level of measurement corresponding to the data "Time of return home" associated with fishing.
A) ratio
B) nominal
C) interval
D) none of these choices
E) ordinal

____ 22. Identify the level of measurement corresponding to the data "Length of time to complete an exam" associated with student life.
A) interval
B) nominal
C) ratio
D) ordinal
E) none of these choices

_____ 23. Use a random-number table to get a list of 4 random numbers between 100 and 150 beginning at row 5, block 2.

65321	85623	10204	50218	20321	22315
98532	91972	39800	45670	20510	10451
92012	59826	35456	79289	91483	29754
45652	98653	45863	36963	15326	78952
45678	10100	91251	37041	13712	14672

A) 101 100 137 121
B) 101 125 137 112
C) 101 125 137 121
D) 101 100 137 112
E) 101 125 137 137

_____ 24. Use a random-number table to simulate the outcomes of tossing a quarter 20 times beginning at row 5, block 3. Assume that the quarter is balanced (i.e., fair) and an even digit is assigned to the outcome heads (H) and an odd digit to the outcome tails (T).

65321	85623	10204	50218	20321	22315
98532	91972	39800	45670	20510	10451
92012	59826	35456	79289	91483	29754
45652	98653	45863	36963	15326	78952
45678	10100	91251	37041	13712	14672

A) T T H T T T T H H T T T T T H H T H H T H
B) T T H T T T T H T T T T T T T H T H H T H
C) T T H T T T T H H T T T T T T H T H T H H
D) T T H T T T T H H T T T T T T H T H T T H
E) T T H T T T T H H T T T T T T H T H H T H

_____ 25. Use a random-number table to simulate the outcomes of tossing a quarter 10 times beginning at row 3, block 3. Assume that the quarter is balanced (i.e., fair) and an even digit is assigned to the outcome heads (H) and an odd digit to the outcome tails (T).

65321	85623	10204	50218	20321	22315
98532	91972	39800	45670	20510	10451
92012	59826	35456	79289	91483	29754
45652	98653	45863	36963	15326	78952
45678	10100	91251	37041	13712	14672

A) T T H T H T H H H T
B) T T H T H T T H T T
C) T T H T H T T H H H
D) T T H T H T T T H T
E) T T H T H T T H H T

____   26. Identify the sampling technique used in the following information.
An important part of employee compensation is a benefits package that might include health insurance, life insurance, child care, vacation days, retirement plan, parental leave, bonuses, etc. Suppose you want to conduct a survey of benefits packages available in private businesses in Hawaii. You want a sample size of 100. Sampling technique used to get the sample size of 100 is described below.
Use the Island Business Directory. Number all the businesses. Select a starting place at random, and then use every 50th business listed until you have 100 businesses.
A) Convenience sample
B) Simple random sample
C) Systematic sample
D) Cluster sample
E) Stratified sample

____   27. Identify the sampling technique used in the following information.
An important part of employee compensation is a benefits package that might include health insurance, life insurance, child care, vacation days, retirement plan, parental leave, bonuses, etc. Suppose you want to conduct a survey of benefits packages available in private businesses in Hawaii. You want a sample size of 100. Sampling technique used to get the sample size of 100 is described below.
Send a team of five research assistants to Bishop Street in downtown Honolulu. Let each assistant select a block of building and interview an employee from each business found. Each researcher can have the rest of the day off after getting responses from 20 different businesses.
A) Stratified sample
B) Cluster sample
C) Convenience sample
D) Systematic sample
E) Simple random sample

____   28. Find the technique for gathering data in the study below.
An ecology class used binoculars to watch 23 turtles at Lowell Pond. It was found that 18 were box turtles, and 5 were snapping turtles.
A) experiment
B) observational study
C) sampling
D) census
E) none of these choices

_____ 29. Find the technique for gathering data in the study below.
The Colorado Division of Wildlife caught 41 bighorn sheep on Mt. Evans and gave each
one an injection to prevent heartworm. A year later 38 of these sheep did not have
heartworm while the other three did.
A) observational study
B) experiment
C) sampling
D) none of these choices
E) census

_____ 30. Find the technique for gathering data in the study below.
A study of all league football scores attained through touchdowns and field goals was
conduced by the National Football League to determine whether field goals account for more
scoring events than touchdowns (USA Today).
A) sampling
B) census
C) experiment
D) simulation
E) none of these choices

_____ 31. Find the technique for gathering data in the study below.
The effects of wind shear on airplanes during both landing and takeoff are studied by using complex
computer programs that mimic actual flight.
A) experiment
B) simulation
C) sampling
D) census
E) none of these choices

_____ 32. Use a random-number table to get a list of 10 random numbers between 1 and 99 beginning at row
2, block 2.

65321	85623	10204	50218	20321	22315
98532	91972	39800	45670	20510	10451
92012	59826	35456	79289	91483	29754
45652	98653	45863	36963	15326	78952
45678	10100	91251	37041	13712	14672

A) 91 97 23 98 45 67 20 05 10 45
B) 91 97 23 98 45 67 02 05 10 10
C) 91 97 23 98 45 67 02 05 10 45
D) 91 97 32 98 45 67 02 05 10 45
E) 91 97 32 98 45 67 02 50 10 45

____ 33. Use a random-number table to get a list of 8 random numbers between 1 and 50 beginning at row 1, block 3.

65321	85623	10204	50218	20321	22315
98532	91972	39800	45670	20510	10451
92012	59826	35456	79289	91483	29754
45652	98653	45863	36963	15326	78952
45678	10100	91251	37041	13712	14672

A) 10 20 45 02 18 20 32 12
B) 10 20 54 02 18 32 12 23
C) 10 20 45 02 18 32 12 23
D) 10 02 45 02 18 32 12 23
E) 10 02 45 02 18 32 12 31