海外优秀数学类教材系列丛书

PEARSON Addison Wesley

影印版

Introductory Statistics

(Sixth Edition)

统计学导论

(第6版)

NEIL A. WEISS

高等教育出版社
Higher Education Press

海外优秀数学类教材系列丛书

影印版

Introductory Statistics

(Sixth Edition)

统计学导论

(第6版)

Neil A. Weiss

Arizona State University

高等教育出版社
Higher Education Press

图字:01-2004-0389 号

图书在版编目 (CIP) 数据

统计学导论 = Introductory Statistics: 第 6 版 / (美) 韦斯 (Weiss, N.A.). —影印本. —北京: 高等教育出版社, 2004.10
(海外优秀数学类教材系列丛书)
ISBN 7-04-015559-1

Ⅰ.统... Ⅱ.韦... Ⅲ.统计学-高等学校-教材-英文 Ⅳ.C8

中国版本图书馆 CIP 数据核字 (2004) 第 105783 号

出版发行	高等教育出版社	购书热线	010-64054588
社 址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100011	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
经 销	新华书店北京发行所		
印 刷	北京外文印刷厂		
开 本	880×1230 1/16	版 次	2004 年 10 月第 1 版
印 张	60	印 次	2004 年 10 月第 1 次印刷
字 数	700 000	定 价	69.00 元(含光盘)

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

物料号:15559-00

出版者的话

在我国已经加入WTO、经济全球化的今天，为适应当前我国高校各类创新人才培养的需要，大力推进教育部倡导的双语教学，配合教育部实施的“高等学校教学质量与教学改革工程”和“精品课程”建设的需要，高等教育出版社有计划、大规模地开展了海外优秀数学类系列教材的引进工作。

高等教育出版社和 Pearson Education, John Wiley & Sons, McGraw-Hill, Thomson Learning 等国外出版公司进行了广泛接触，经国外出版公司的推荐并在国内专家的协助下，提交引进版权总数 100 余种。收到样书后，我们聘请了国内高校一线教师、专家、学者参与这些原版教材的评介工作，并参考国内相关专业的课程设置和教学实际情况，从中遴选出了这套优秀教材组织出版。

这批教材普遍具有以下特点：(1) 基本上是近 3 年出版的，在国际上被广泛使用，在同类教材中具有相当的权威性；(2) 高版次，历经多年教学实践检验，内容翔实准确、反映时代要求；(3) 各种教学资源配套整齐，为师生提供了极大的便利；(4) 插图精美、丰富，图文并茂，与正文相辅相成；(5) 语言简练、流畅、可读性强，比较适合非英语国家的学生阅读。

本系列丛书中，有 Finney、Weir 等编的《托马斯微积分》(第 10 版，Pearson)，其特色可用“呈传统特色、富革新精神”概括，本书自 20 世纪 50 年代第 1 版以来，平均每四五年就有一个新版面世，长达 50 余年始终盛行于西方教坛，作者既有相当高的学术水平，又热爱教学，长期工作在教学第一线，其中，年近 90 的 G.B.Thomas 教授长年在 MIT 工作，具有丰富的教学经验；Finney 教授也在 MIT 工作达 10 年；Weir 是美国数学建模竞赛委员会主任。Stewart 编的立体化教材《微积分》(第 5 版，Thomson Learning) 配备了丰富的教学资源，是国际上最畅销的微积分原版教材，2003 年全球销量约 40 余万册，在美国，占据了约 50%～60% 的微积分教材市场，其用户包括耶鲁等名牌院校及众多一般院校 600 余所。本系列丛书还包括 Anton 编的经典教材《线性代数及其应用》(第 8 版，Wiley)；Jay L. Devore 编的优秀教材《概率论与数理统计》(第 5 版，Thomson Learning) 等。在努力降低引进教材售价方面，高等教育出版社做了大量和细致的工作，这套引进的教材体现了一定的权威性、系统性、先进性和经济性等特点。

通过影印、翻译、编译这批优秀教材，我们一方面要不断地分析、学习、消化吸收国外优秀教材的长处，吸取国外出版公司的制作经验，提升我们自编教材的立体化配套标准，使我国高校教材建设水平上一个新的台阶；与此同时，我们还将尝试组织海外作者和国内作者合编外文版基础课数学教材，并约请国内专家改编部分国外优秀教材，以适应我国实际教学环境。

这套教材出版后，我们将结合各高校的双语教学计划，开展大规模的宣传、培训工作，及时地将本套丛书推荐给高校使用。在使用过程中，我们衷心希望广大高校教师和同学提出宝贵的意见和建议，如有好的教材值得引进，请与高等教育出版社高等理科分社联系。

联系电话：010-58581384，E-mail：xuke@hep.com.cn。

高等教育出版社
2004 年 4 月 20 日

part i

Introduction

CHAPTER 1 | The Nature of Statistics

chapter 1

The Nature of Statistics

CHAPTER OUTLINE

1.1 Two Kinds of Statistics

1.2 The Technology Center

1.3 Simple Random Sampling

1.4 Other Sampling Designs

1.5 Experimental Designs

GENERAL OBJECTIVES What does the word *statistics* bring to mind? Most people immediately think of numerical facts or data, such as unemployment figures, farm prices, or the number of marriages and divorces. *Webster's New World Dictionary* gives two definitions of the word *statistics:*

1. facts or data of a numerical kind, assembled, classified, and tabulated so as to present significant information about a given subject.
2. [construed as sing.], the science of assembling, classifying, and tabulating such facts or data.

But statistics encompasses much more than these definitions convey. Not only do statisticians assemble, classify, and tabulate data, but they also analyze data for the purpose of making generalizations and decisions. For example, a political analyst can use data from a portion of the voting population to predict the political preferences of the entire voting population. And a city council can decide where to build a new airport runway based on environmental impact statements and demographic reports that include a variety of statistical data.

In this chapter, we introduce some basic terminology so that the various meanings of the word *statistics* will become clear to you. We also examine two primary ways of producing data, namely, through sampling and experimentation. We discuss sampling designs in Sections 1.3 and 1.4, and experimental designs in Section 1.5.

case study

TOP FILMS OF ALL TIME

The American Film Institute (AFI) conducted a survey as part of a celebration of the 100th anniversary of cinema. AFI polled 1500 filmmakers, actors, critics, politicians, and film historians, asking them to pick their 100 favorite films from a list of 400. The films on the list were made between 1896 and 1996.

After tallying the responses, AFI compiled a list representing the top 100 films. *Citizen Kane,* made in 1941, finished in first place, followed by *Casablanca,* which was made in 1942. Following are the top 40 finishers in the poll.

Rank	Film	Year	Rank	Film	Year
1	Citizen Kane	1941	21	The Grapes of Wrath	1940
2	Casablanca	1942	22	2001: A Space Odyssey	1968
3	The Godfather	1972	23	The Maltese Falcon	1941
4	Gone With the Wind	1939	24	Raging Bull	1980
5	Lawrence of Arabia	1962	25	E.T. The Extra-Terrestrial	1982
6	The Wizard of Oz	1939	26	Dr. Strangelove	1964
7	The Graduate	1967	27	Bonnie & Clyde	1967
8	On the Waterfront	1954	28	Apocalypse Now	1979
9	Schindler's List	1993	29	Mr. Smith Goes to Washington	1939
10	Singin' in the Rain	1952	30	The Treasure of the Sierra Madre	1948
11	It's a Wonderful Life	1946	31	Annie Hall	1977
12	Sunset Blvd.	1950	32	The Godfather, Part II	1974
13	The Bridge on the River Kwai	1957	33	High Noon	1952
14	Some Like It Hot	1959	34	To Kill a Mockingbird	1962
15	Star Wars	1977	35	It Happened One Night	1934
16	All About Eve	1950	36	Midnight Cowboy	1969
17	The African Queen	1951	37	The Best Years of Our Lives	1946
18	Psycho	1960	38	Double Indemnity	1944
19	Chinatown	1974	39	Doctor Zhivago	1965
20	One Flew Over the Cuckoo's Nest	1975	40	North by Northwest	1959

Armed with the knowledge gained in this chapter, you will be asked to analyze further the AFI poll at the end of the chapter.

1.1 TWO KINDS OF STATISTICS

You probably already know something about statistics. If you read newspapers, surf the Web, watch the news on television, or follow sports, you see and hear the word *statistics* frequently. In this section, we use familiar examples such as baseball statistics and voter polls to introduce the two major types of statistics: **descriptive statistics** and **inferential statistics.** We also examine how to classify studies as either descriptive or inferential.

DESCRIPTIVE STATISTICS

Each spring in the late 1940s President Harry Truman officially opened the major league baseball season by throwing out the "first ball" at the opening game of the Washington Senators. Both President Truman and the Washington Senators had reason to be interested in statistics in 1948. We use the 1948 baseball season to illustrate the first major type of statistics, descriptive statistics, in Example 1.1.

Example 1.1 ***Descriptive Statistics***

The 1948 Baseball Season In 1948, the Washington Senators played 153 games, winning 56 and losing 97. They finished seventh in the American League and were led in hitting by Bud Stewart, whose batting average was .279. These and many other statistics were compiled by baseball statisticians who took the complete records for each game of the season and organized that large mass of information effectively and efficiently.

Although baseball fans take baseball statistics for granted, a great deal of time and effort is required to gather and organize them. Moreover, without such statistics, baseball would be much harder to understand. For instance, picture yourself trying to select the best hitter in the American League with only the official score sheets for each game. (More than 600 games were played in 1948; the best hitter was Ted Williams, who led the league with a batting average of .369.) ◆

The work of baseball statisticians provides an excellent illustration of descriptive statistics. A formal definition of the term *descriptive statistics* follows.

DEFINITION 1.1 **Descriptive Statistics**

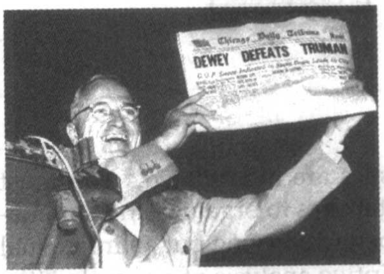
Descriptive statistics consists of methods for organizing and summarizing information.

Descriptive statistics includes the construction of graphs, charts, and tables and the calculation of various descriptive measures such as averages, measures of variation, and percentiles. We discuss descriptive statistics in detail in Chapters 2 and 3.

INFERENTIAL STATISTICS

We use the 1948 presidential election to introduce the other major type of statistics, inferential statistics, in Example 1.2.

Example 1.2 *Inferential Statistics*



The 1948 Presidential Election In the fall of 1948, President Truman was also concerned about statistics. The *Gallup Poll* taken just prior to the election predicted that he would win only 44.5% of the vote and be defeated by the Republican nominee, Thomas E. Dewey. But this time the statisticians had predicted incorrectly. Truman won more than 49% of the vote and, with it, the presidency. The Gallup Organization modified some of its procedures and has correctly predicted the winner ever since. ◆

Political polling provides an example of inferential statistics. Interviewing everyone of voting age in the United States on their voting preferences would be expensive and unrealistic. Statisticians who want to gauge the sentiment of the entire **population** of U.S. voters can afford to interview only a carefully chosen group of a few thousand voters. This group is called a **sample** of the population. Statisticians analyze the information obtained from a sample of the voting population to make inferences (draw conclusions) about the preferences of the entire voting population. Inferential statistics provides methods for making such inferences.

The terminology just introduced in the context of political polling is used in general in statistics. Specifically, the terms *population* and *sample* are defined as follows.

DEFINITION 1.2 **Population and Sample**

Population: The collection of all individuals or items under consideration in a statistical study.

Sample: That part of the population from which information is obtained.

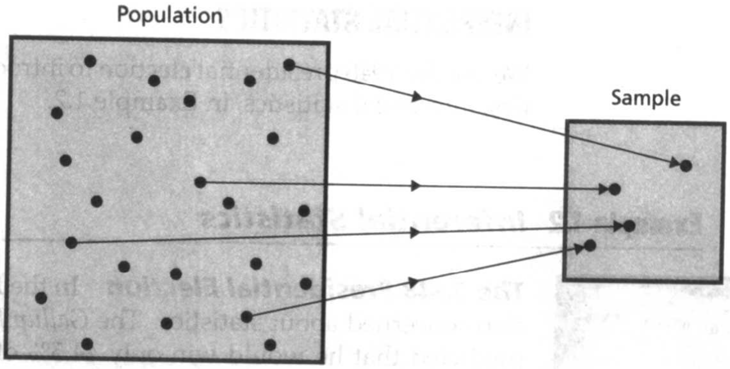
Figure 1.1 depicts the relationship between a population and a sample from the population.

Now that we have discussed the terms *population* and *sample,* we can define *inferential statistics.*

DEFINITION 1.3 **Inferential Statistics**

Inferential statistics consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population.

FIGURE 1.1
Relationship between population and sample



Descriptive statistics and inferential statistics are interrelated. You must almost always use techniques of descriptive statistics to organize and summarize the information obtained from a sample before carrying out an inferential analysis. Furthermore, the preliminary descriptive analysis of a sample often reveals features that lead you to the choice of (or to a reconsideration of the choice of) the appropriate inferential method.

CLASSIFYING STATISTICAL STUDIES

As you proceed through this book, you will obtain a thorough understanding of the principles of descriptive and inferential statistics. At this point, you should be able to classify statistical studies as either descriptive or inferential. In doing so, you should consider the intent of the study.

On the one hand, if the intent of the study is to examine and explore the information obtained for its own intrinsic interest only, the study is descriptive. On the other hand, if the information is obtained from a sample of a population and the intent of the study is to use that information to draw conclusions about the population, the study is inferential.

Thus a descriptive study may be performed on a sample as well as on a population. Only when an inference is made about the population, based on information obtained from the sample, does the study become inferential.

Examples 1.3 and 1.4 further illustrate the distinction between descriptive and inferential studies. In each example, we present the result of a statistical study and classify the study as either descriptive or inferential. Try to classify each study yourself before reading our explanation.

Example 1.3 *Classifying Statistical Studies*

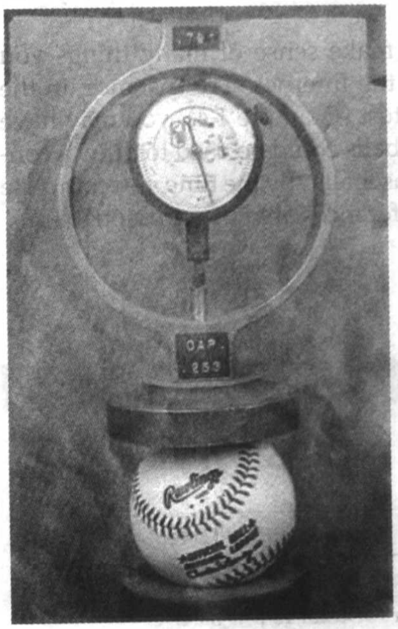
The 1948 Presidential Election *The study* Table 1.1 displays the voting results for the 1948 presidential election.

Classification This study is descriptive. It is a summary of the votes cast by U.S. voters in the 1948 presidential election. No inferences are made. ◆

TABLE 1.1
Final results of the 1948 presidential election

Ticket	Votes	Percentage
Truman–Barkley (Democratic)	24,179,345	49.7
Dewey–Warren (Republican)	21,991,291	45.2
Thurmond–Wright (States Rights)	1,176,125	2.4
Wallace–Taylor (Progressive)	1,157,326	2.4
Thomas–Smith (Socialist)	139,572	0.3

Example 1.4 *Classifying Statistical Studies*



Testing Baseballs *The study* For the 101 years preceding 1977, baseballs used by the major leagues were purchased from the Spalding Company. In 1977, that company stopped manufacturing major league baseballs, and the major leagues arranged to buy their baseballs from the Rawlings Company.

Early in the 1977 season, pitchers began to complain that the Rawlings ball was "livelier" than the Spalding ball. They claimed it was harder, bounced farther and faster, and gave hitters an unfair advantage. There was some evidence for this claim: In the first 616 games of 1977, 1033 home runs were hit, compared to only 762 home runs hit in the first 616 games of 1976.

Sports Illustrated magazine sponsored a careful study of the liveliness question, and the results appeared in the article "They're Knocking the Stuffing Out of It" by Larry Keith (*Sports Illustrated*, June 13, 1977, pp. 23–27). In this study, an independent testing company randomly selected 85 baseballs from the current (1977) supplies of various major league teams. The bounce, weight, and hardness of the baseballs chosen were carefully measured. Those measurements were then compared with measurements obtained from similar tests on baseballs used in 1952, 1953, 1961, 1963, 1970, and 1973.

The conclusion, presented on page 24 of the *Sports Illustrated* issue, was that "... the 1977 Rawlings ball is livelier than the 1976 Spalding, but not as lively as it could be under big league rules, or as the ball has been in the past."

Classification This study is inferential. The independent testing company used a sample of 85 baseballs from the 1977 supplies of major league teams to make an inference about the population of all such baseballs. (An estimated 360,000 baseballs were used by the major leagues in 1977.) ◆

The *Sports Illustrated* study also shows that it is often not feasible to obtain information for the entire population. Indeed, after the bounce and hardness tests, all of the baseballs sampled were taken to a butcher in Plainfield, New Jersey, to be sliced in half so that researchers could look inside them. Clearly, testing every baseball in this way would not have been practical.

THE DEVELOPMENT OF STATISTICS

According to the *Dictionary of Scientific Biography*, "The word *Statistik*, first printed in 1672, meant *Staatswissenschaft*, or, rather, a science concerning the states. It was cultivated at the German universities, where it consisted of more or less systematically collecting 'state curiosities' rather than quantitative material."

The modern science of statistics is much broader than just collecting "state curiosities." Historically, descriptive statistics appeared first. Censuses were taken as long ago as Roman times. Over the centuries, records of such things as births, deaths, marriages, and taxes led naturally to the development of descriptive statistics.

Inferential statistics is a newer arrival. Major developments began to occur with the research of Karl Pearson (1857–1936) and Ronald Fisher (1890–1962), who published their findings in the early years of the twentieth century. Since the work of Pearson and Fisher, inferential statistics has evolved rapidly and is now applied in a myriad of fields.

What Does it Mean?

An understanding of statistical reasoning and of the basic concepts of descriptive and inferential statistics has become mandatory for virtually everyone, in both their private and professional lives.

Familiarity with statistics will help you make sense of many things you read in newspapers and magazines and on the Internet. For instance, in the description of the *Sports Illustrated* baseball test (Example 1.4), you may have questioned whether a sample of only 85 baseballs could be used to draw a conclusion about a population of some 360,000 baseballs. By the time you complete Chapter 9, you will understand why such inferences are not unreasonable.

Exercises 1.1

Statistical Concepts and Skills

1.1 Define the following terms.
a. Population **b.** Sample

1.2 What are the two major types of statistics? Describe them in detail.

1.3 Identify some of the methods used in descriptive statistics.

1.4 Explain two ways in which descriptive statistics and inferential statistics are interrelated.

In Exercises 1.5–1.10, classify each of the studies as either descriptive or inferential.

1.5 TV Viewing Times. The A. C. Nielsen Company collects and publishes information on the television viewing habits of Americans. Data from a sample of Americans yielded the following estimates of average TV viewing time per week for all Americans. The times are in hours and minutes. [SOURCE: Nielsen Media Research, *Nielsen Report on Television*.]

Group (by age)		Time
Average all persons		*30:14*
Women	Total 18+	34:47
	18–24	28:54
	25–54	31:05
	55+	44:11
Men	Total 18+	30.41
	18–24	23:31
	25–54	28:44
	55+	38:47
Teens	12–17	21:50
Children	2–11	23:01

1.6 Professional Athlete Salaries. In the *Statistical Abstract of the United States, 1999*, average professional athletes' salaries in baseball, basketball, and football were compiled and compared for the years 1987 and 1997.

Sport	Average Salary ($1000) 1987	1997
Baseball (MLB)	412	1,337
Basketball (NBA)	440	2,200
Football (NFL)	203	725

1.7 Causes of Death. The U.S. National Center for Health Statistics published the following rate estimates for the leading causes of death in 1997 in *Vital Statistics of the United States*. The estimates are based on a 10% sampling of all 1997 U.S. death certificates. Rates are per 100,000 population.

Cause	Rate
Major cardiovascular diseases	352.2
Malignancies (cancers)	200.8
Chronic obstructive pulmonary diseases	41.3
Accidents	34.4
Pneumonia and influenza	33.0

1.8 Drug Use. The U.S. Substance Abuse and Mental Health Services Administration collects and publishes data on drug use, by type of drug and age group, in *National Household Survey on Drug Abuse*. The following table provides information for the years 1990 and 1997. The percentages shown are estimates obtained from national samples.

Type of drug	Percentage, 12 years old and over: *Ever used* 1990	*Ever used* 1997	*Current user* 1990	*Current user* 1997
Marijuana	30.5	30.9	5.4	5.1
Cocaine	11.2	10.5	0.9	0.7
Inhalants	5.7	5.7	0.4	0.4
Hallucinogens	7.9	9.6	0.4	0.8
Heroin	0.8	0.9	—	0.2
Stimulants[1]	5.5	4.5	0.6	0.3
Sedatives[1]	2.8	1.9	0.2	0.1
Tranquilizers[1]	4.0	3.2	0.6	0.4
Analgesics[1]	6.3	4.9	0.9	0.7
Alcohol	82.2	81.9	52.6	51.4

1 = Nonmedical use.

1.9 Dow Jones Industrial Averages. The following table provides the closing values of the Dow Jones Industrial Averages as of the end of December for the years 1994–1999. [SOURCE: Global Financial Data.]

Year	Closing value
1994	3,834.4
1995	5,117.1
1996	6,448.3
1997	7,908.3
1998	9,181.4
1999	11,497.1

1.10 The Music People Buy. Results of monthly telephone surveys yielded the following percentage estimates of all music expenditures. These statistics were published in *1998 Consumer Profile*. [SOURCE: Recording Industry Association of America, Inc.]

Music Type	Expenditure (%)
Rock	25.7
Country	14.1
R&B	12.8
Pop	10.0
Rap	9.7
Oldies Gospel	6.3
Classical	3.3
Jazz	1.9
Soundtracks	1.7
New Age	0.6
Children's	0.4
Other	13.5

Extending the Concepts and Skills

1.11 Organically Grown Produce. A *Newsweek* poll of a sample of Americans revealed that "84% of those surveyed would choose organically grown produce over produce grown using chemical fertilizers, pesticides, and herbicides."

a. Is the statement in quotes an inferential or a descriptive statement?

b. Based on the same information, what if the statement had been "84% of Americans would choose organically grown produce over produce grown using chemical fertilizers, pesticides, and herbicides"?

1.12 Nuclear Weapons Test Ban. In a press release dated September 24, 1997, dateline Washington, D.C., the Mellman Group for the Coalition to Reduce Nuclear Dangers reported that "a new nationwide poll shows that 70.3% of Americans think the U.S. Senate should approve a treaty with 140 countries that would prohibit underground nuclear weapons explosions worldwide."

a. Do you think that the statement in the press release is inferential or descriptive? Can you be sure?
b. Actually, the Mellman Group conducted an opinion survey of 800 adults and determined that 70.3% of them thought that the U.S. Senate should approve a treaty with 140 countries to prohibit underground nuclear weapons explosions worldwide. How would you rephrase the statement in the press release to make clear that it is a descriptive statement?

1.2 THE TECHNOLOGY CENTER

Today, programs for conducting statistical and data analyses are available in dedicated statistical software packages, general-use spreadsheet software, and graphing calculators. In this book, we discuss three of the most popular technologies for doing statistics: Minitab, Excel, and the TI-83 Plus.

For the most part, for Excel we use Data Desk/XL (DDXL) from Data Description, Inc. This statistics add-in complements Excel's standard statistics capabilities; it is included on the WeissStats CD, which comes with your book. Further details of Minitab, Excel, the TI-83 Plus, and other statistical technologies are provided in supplements written specifically to accompany this book.

At the end of appropriate sections of the book, in subsections titled The Technology Center, we present and interpret output from the three technologies mentioned. The output from each technology addresses problems that were solved by hand earlier in the section. For this aspect of The Technology Center, you need neither a computer nor a graphing calculator, nor do you need any working knowledge of the technologies under discussion.

For those who want to learn how to obtain the output for one or more of the three technologies, step-by-step instructions are presented in subsections titled Obtaining the Output (Optional). When studying this material, you will get the best results by using your computer or graphing calculator to perform the steps described.

Each statistical technology has a slightly different method for data input and output. At this point, you should spend some time learning the method for each technology you want to study. The documentation or online help for your technology gives you the necessary details. For Minitab, Excel, and the TI-83 Plus, you can also find this information in the appropriate technology supplement to this book.

1.3 SIMPLE RANDOM SAMPLING

Throughout this book, we present examples of organizations or people conducting studies: A consumer group wants information about the gas mileage of a particular make of car, so it performs mileage tests on a sample of such cars and statistically analyzes the resulting data; or a teacher wants to know about the comparative merits of two teaching methods, so she tests those methods on two groups of students. This approach reflects a healthy attitude: To obtain information about a subject of interest, plan and conduct a study.

However, the possibility always exists that a study being considered has already been done. Repeating it would be a waste of time, energy, and money.

What Does it Mean?

You can often avoid the effort and expense of a study if someone else has already done that study and published the results.

Therefore, before a study is planned and conducted, a literature search should be made. Doing so does not require going through all the books in the library or making an extensive Internet search. Many information collection agencies specialize in finding studies on specific topics in specific areas.

CENSUS, SAMPLING, AND EXPERIMENTATION

If information required is not already available from a previous study, you can plan a new study to obtain the information. One method for acquiring information is to conduct a **census,** that is, obtain information on the entire population of interest. However, conducting a census is generally time consuming and costly, frequently impractical, and sometimes impossible.

Two methods other than a census for obtaining information are **sampling** and **experimentation.** In much of this book, we concentrate on sampling. However, we introduce experimentation in Section 1.5, discuss it sporadically throughout the text, and examine it in detail in the chapter *Design of Experiments and Analysis of Variance* on the WeissStats CD accompanying this book or on the Weiss Web site, `www.aw.com/weiss`.

If sampling is deemed appropriate, you must then decide how to select the sample; that is, you must choose the method for obtaining a sample from the population. In making that choice, keep in mind that the sample will be used to draw conclusions about the entire population. Consequently, the sample should be a **representative sample,** that is, it should reflect as closely as possible the relevant characteristics of the population under consideration.

For instance, it would not make sense to use the average weight of a sample of professional football players to make an inference about the average weight of all adult males. Nor would it be reasonable to estimate the median income of California residents by sampling the incomes of Beverly Hills residents.

To see what can happen when a sample is not representative, consider the presidential election of 1936. Before the election, the *Literary Digest* magazine conducted an opinion poll of the voting population. Its survey team asked a sample of the voting population whether they would vote for Franklin D. Roosevelt, the Democratic candidate, or for Alfred Landon, the Republican candidate.

Based on the results of the survey, the magazine predicted an easy win for Landon. But when the actual election results were in, Roosevelt won by the greatest landslide in the history of presidential elections! What happened? Here are two reasons given for the failure of the poll.

- The sample was obtained from among people who owned a car or had a telephone. In 1936, that group included only the more well-to-do people, and historically such people tend to vote Republican.
- The response rate was low (less than 25% of those polled responded), and there was a nonresponse bias (a disproportionate number of those who responded to the poll were Landon supporters).

Whatever the reason for the poll's failure, the sample obtained by the *Literary Digest* obviously was not representative.

Most modern sampling procedures involve the use of **probability sampling.** In probability sampling, a random device, such as tossing a coin or consulting a table of random numbers, is used to decide which members of the population will constitute the sample instead of leaving such decisions to human judgment.

The use of probability sampling may still yield a nonrepresentative sample. However, probability sampling eliminates unintentional selection bias and permits the researcher to control the chance of obtaining a nonrepresentative sample. Furthermore, the use of probability sampling guarantees that the techniques of inferential statistics can be applied. In this section and the next, we will examine the most important probability-sampling methods.

SIMPLE RANDOM SAMPLING

The inferential techniques considered in this book are intended for use with only one particular sampling procedure: **simple random sampling,** or just **random sampling,** which yields a **simple random sample.** Simple random sampling is the basic type of probability sampling and is also the foundation for the more complex types of probability sampling.

DEFINITION 1.4 **Simple Random Sampling; Simple Random Sample**

Simple random sampling: A sampling procedure for which each possible sample of a given size is equally likely to be the one obtained.

Simple random sample: A sample obtained by simple random sampling.

There are two types of simple random sampling. One is **simple random sampling with replacement,** whereby a member of the population can be selected more than once; the other is **simple random sampling without replacement,** whereby a member of the population can be selected at most once. *Unless we specify otherwise, assume that simple random sampling is done without replacement.*

In Example 1.5, we chose a very small population—the five top Oklahoma state officials—to illustrate simple random sampling. In practice, we would not sample from such a small population but would instead take a census. Using a small population here makes understanding the concept of simple random sampling easier.

Example 1.5 *Simple Random Samples*

TABLE 1.2
Five top Oklahoma state officials

Governor (G)
Lieutenant Governor (L)
Secretary of State (S)
Attorney General (A)
Treasurer (T)

Sampling Oklahoma State Officials As reported by *The World Almanac,* the top five state officials of Oklahoma are as shown in Table 1.2. Consider these five officials a population of interest.

a. List the possible samples (without replacement) of two officials from this population of five officials.

b. Describe a method for obtaining a simple random sample of two officials from this population of five officials.

c. For the sampling method described in part (b), what are the chances that any particular sample of two officials will be the one selected?
d. Repeat parts (a)–(c) for samples of size 4.

Solution For convenience, we use the letters in parentheses after the officials in Table 1.2 to represent them.

TABLE 1.3
The 10 possible samples of two officials

G, L	G, S	G, A	G, T
L, S	L, A	L, T	S, A
S, T	A, T		

a. There are 10 possible samples of two officials from the population of five officials, as listed in Table 1.3.
b. To obtain a simple random sample of size 2 we could first write the letters that correspond to the five officials, G, L, S, A, and T, on separate pieces of paper. Next, we could place the five slips of paper in a box and shake it. Then, while blindfolded, we could pick two slips of paper.
c. The sampling procedure described in part (b) ensures that we are taking a simple random sample. Consequently, each of the possible samples of two officials is equally likely to be the one selected. There are 10 possible samples, so the chances are $\frac{1}{10}$ (1 in 10) that any particular sample will be the one selected.

TABLE 1.4
The five possible samples of four officials

G, L, S, A	G, L, S, T
G, L, A, T	G, S, A, T
L, S, A, T	

d. There are five possible samples of four officials from the population of five officials, as listed in Table 1.4. In this case, a simple random sampling procedure, such as picking four slips of paper out of a box, gives each of the five possible samples in Table 1.4 a 1 in 5 chance of being the one selected. ◆

RANDOM-NUMBER TABLES

Obtaining a simple random sample by picking slips of paper out of a box is usually not practical, especially when the population to be sampled is large. But there are several practical procedures to get simple random samples. One common method is to use a **table of random numbers**—a table of randomly chosen digits. In Example 1.6, we explain how a table of random numbers can be used to obtain a simple random sample.

Example 1.6 *Random-Number Tables*

Sampling Student Opinions Student questionnaires, known as "teacher evaluations," gained widespread use in the late 1960s and early 1970s. Generally, student evaluations of teachers are not done at final exam time. More commonly, professors hand out evaluation forms a week or so before the final.

That practice, however, poses several problems. On some days, less than 60% of the students registered for a class may actually attend. Moreover, because many of those who are present have preparations to make for other classes, they often complete their teacher evaluation forms in a hurry so that they can leave class early. A better method, therefore, might be to select a sample of students from the class and interview them individually. In this kind of situation, a simple random sample is appropriate.