STATISTICS

For Engineering and the Sciences

FOURTH EDITION

William Mendenhall
University of Florida

Terry Sincich
University of South Florida

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Contents

CHAPTER FOUR

Discrete Random Variables 143

CHAPTER FIVE

Continuous Random Variables 203

CHAPTER SIX

Bivariate Probability Distributions 257

CHAPTER SEVEN

Sampling Distributions 289

CHAPTER EIGHT

Estimation 337

CHAPTER NINE

Tests of Hypotheses 421

CHAPTER TEN

Categorical Data Analysis 495

CHAPTER ELEVEN

Simple Linear Regression 531

CHAPTER TWELVE

Multiple Regression Analysis 599

CHAPTER THIRTEEN

Model Building 699

CHAPTER FOURTEEN

Analysis of Variance for Designed Experiments 789

CHAPTER FIFTEEN

Nonparametric Statistics 919

APPENDIX I

Matrix Algebra 1065

APPENDIX II

Useful Statistical Tables 1085

APPENDIX III

DDT Analyses on Fish Samples, Tennessee River, Alabama 1125

APPENDIX IV

APPENDIX V

APPENDIX VI

APPENDIX VII

Preface

The fourth edition of *Statistics for Engineering and the Sciences* is a text for a two-semester introductory course in statistics for students majoring in engineering or any of the physical sciences. Inevitably, once these students graduate and are employed, they will be involved in the analysis of data and will be required to make inferences from their analyses. Consequently, they need to acquire knowledge of the basic concepts of statistical inference and familiarity with some of the statistical methods that they will be required to use in their employment.

Pedagogy

Chapters 1–7 identify the objectives of statistics, explain how we can describe data sets, and present the basic concepts of probability. Chapters 8 and 9 introduce the two methods for making inferences about population parameters: estimation and testing hypotheses. These notions are extended in the remaining chapters to cover other topics that are useful in analyzing engineering and scientific data, including the analysis of categorical data (Chapter 10), regression analysis and model building (Chapters 11–13), the analysis of variance for designed experiments (Chapter 14), nonparametric statistics (Chapter 15), statistical quality control (Chapter 16), and product and system reliability (Chapter 17).

The assumed mathematical background is a two-semester sequence in calculus—that is, the course could be taught to students of average mathematical talent and with a basic understanding of the principles of differential and integral calculus. Presentation requires the ability to perform one-variable differentiation and integration, but examples involving topics from multivariable calculus are designated as optional. Thus, the theoretical concepts are sketched and presented in a one-variable context, but it is easy for the instructor to delve deeper into the theoretical and mathematical aspects of statistics using the optional topics, examples, and exercises.

Features

Specific features of the text are the following:

1. **Blend of theory and applications.** The basic theoretical concepts of mathematical statistics are integrated with a two-semester presentation of statistical methodology. Thus, the instructor has the opportunity to present a course with either of two characteristics—a course stressing basic concepts and applied statistics or a course that, while still tilted toward application, presents a modest introduction to the theory underlying statistical inference.
2. **Computer applications with instructions on how to use the computer.** The instructor and student have the option of using a computer to perform the statistical calculations. Printouts from two popular statistical software packages available at

most university computing centers, SAS and MINITAB, are fully integrated into the text. Additionally, we provide the SAS and MINITAB commands required to generate the printouts in "Computer Lab" sections at the end of most chapters. These tutorials are designed for the novice user; no prior computer experience is needed. The instructions on how to use SAS and MINITAB for statistical analysis of data apply to both large mainframe computers and personal computers (PCs).

3. **Broad coverage of topics.** To meet the diverse needs of future engineers and scientists, the text provides coverage of a wide range of data analysis topics. The material on exploratory data analysis (Chapter 2), regression analysis and model building (Chapters 11–13), quality control (Chapter 16), and reliability (Chapter 17) sets the text apart from the typical introductory statistics text. The material often refers to theoretical material covered in earlier chapters, but the presentation is oriented toward applications.
4. **Applied exercises extracted from scientific journals.** The text contains a large number of applied exercises designed to motivate a student and suggest future uses for the methodology. Most of these exercises require the student to analyze actual data or interpret experimental results extracted from professional journals in the engineering and physical sciences.
5. **Optional theoretical exercises.** Where appropriate, theoretical exercises are provided to motivate those students who have a stronger desire to understand the mathematical theory that forms an underpinning for the applications. These exercises are labeled "optional" because they require greater mathematical skill for their solution.
6. **Key concepts highlighted.** Definitions, theorems, formulas, steps to follow in performing a statistical procedure, and warnings (indicating a specific situation where a student might misuse a statistical technique) are boxed and highlighted to enable the student to assimilate easily the most important concepts in a chapter.
7. **Real data sets.** Explanations of basic statistical concepts and methodology are based on and motivated by the use of real scientific data sets. Four large data sets are provided in the appendices for use as instructional vehicles:

   Appendix III. Length, weight, and DDT measurements for 144 fish of various species captured from the Tennessee River by the U.S. Army Corps of Engineers.

   Appendix IV. The central processing unit (CPU) times of 1,000 computer jobs run by a small statistical consulting firm.

   Appendix V. Percentage iron content for 390 1.5-kilogram specimens of iron ore selected from a 20,000-ton consignment of Canadian ore.

   Appendix VI. Federal Trade Commission rankings of 372 domestic cigarette brands.

   These data sets are also available (in ASCII format) on floppy diskette. Consequently, they can be loaded into computer storage and analyzed with SAS, MINITAB, or some other statistical software package. For example, the data sets can be used by the instructor to illustrate the concept of a sampling distribution and the theoretical interpretation of a "95% confidence interval."

8. **Short answers to exercises provided.** To aid the student in working the exercise sets, short answers (mostly numerical in nature) to all exercises are provided at the end of the text.

Revisions

Although the scope and coverage remain the same, the fourth edition contains several substantial changes, additions, and enhancements:

1. **More computer printouts.** Throughout the text, we have greatly increased the number of SAS and MINITAB printouts. A printout now accompanies every statistical technique presented, allowing the instructor to emphasize interpretations of the statistical results rather than the calculations required to obtain the results.
2. **Chapter 2: Summary frequency tables.** A discussion of how to construct and use summary frequency tables has been added to the section on describing qualitative data (Section 2.1).
3. **Chapter 4: Bernoulli distribution.** A new section (Section 4.5) describing Bernoulli trials and their importance in binomial experiments is included.
4. **Chapter 5: Descriptive methods for assessing normality.** A new section (Section 5.6) on determining whether a data set is approximately normal has been added to the chapter on continuous probability distributions. In addition to the traditional graphical methods (histogram, stem-and-leaf display), we present the ratio of the interquartile range to the standard deviation as a check on normality. The emphasis on these techniques early in the text makes the student aware of the importance of checking assumptions in later chapters.
5. **Chapter 9: More emphasis on *p*-values.** Throughout the test of hypothesis chapter, we present both the rejection region approach and observed significance level (*p*-value) approach to making decisions. Since a computer printout is provided with nearly each example, it is easy for the instructor to emphasize the *p*-value approach to hypothesis testing.
6. **Chapter 13: Comprehensive example on model building.** The key ideas and techniques of the chapter are applied to a practical problem on detecting collusive bidding in road construction (Section 13.11).
7. **Chapter 14: Principles of experimental design.** Two new sections (Sections 14.4 and 14.5) present an overview of designed experiments and the principles of noise-reducing and volume-increasing designs.
8. **Chapter 14: Regression approach to ANOVA.** Although we present both the traditional ANOVA approach and the regression approach to analyzing data from designed experiments, our emphasis is on the regression approach. For each design, we give the corresponding regression models and show how to conduct the ANOVA F tests using the models.

9. **Chapter 16: Total quality management (TQM).** A new section on total quality management (Section 16.1) has been added to the chapter on statistical process and quality control.
10. **More exercises with real data.** Many new "real-life" scientific exercises have been added throughout the text. All of these are extracted from news articles, magazines, and professional journals.

Numerous, less obvious changes in details have been made throughout the text in response to suggestions by current users and reviewers of the text.

Supplements

The text is also accompanied by the following supplementary material:

1. **Student's solutions manual** (by Nancy S. Boudreau). The manual contains the full solutions for all the odd-numbered exercises contained in the text. ISBN 0-02-312718-X.
2. **Instructor's solutions manual** (by Mark Dummeldinger). The manual contains the full solutions to all the even-numbered exercises contained in the text. ISBN 0-02-380582-X.
3. **Data sets on diskette.** All four large appendix data sets and numerous smaller data sets (containing 20 or more observations) analyzed in exercises are available (in ASCII format) on a 3½" IBM PC diskette. ISBN 0-02-380583-8.
4. **ASP statistical software diskette.** New to this edition, the text includes (inside the back cover) a 3½" micro disk containing the ASP program, *A Statistical Package for Business, Economics, and the Sciences*. ASP, from DMC Software, Inc., is a user-friendly, totally menu-driven program that contains all of the major statistical applications covered in the text, plus many more. ASP runs on any IBM-compatible PC with at least 512K of memory and two disk drives. With ASP, students with no knowledge of computer programming can create and analyze data sets easily and quickly. The appendix contains start-up procedures and a short tutorial on the use of ASP. Full documentation is provided to adopters of the text.
5. **ASP Tutorial and Student Guide** (by George Blackford). Most students have little trouble learning to use ASP without documentation. Some, however, may want to purchase the *ASP Tutorial and Student Guide*. Bookstores can order the tutorial from DMC Software, Inc., 6169 Pebbleshire Drive, Grand Blanc, MI 48439.

Acknowledgments

We wish to acknowledge several individuals who contributed to the revision of this text. Susan Reiland managed the production of the fourth edition with the same high degree of professionalism she exhibited with the first three editions. Faith Sincich served in several capacities, including word-processing specialist, proofreader, moral supporter, and our toughest critic. Finally, we thank the reviewers of this latest edition, who suggested many of the changes listed above. The names of these reviewers follow:

Carl Bodenschatz (United States Air Force Academy)
Christopher Ennis (Normandale Community College)
Nasrollah Etemadi (University of Illinois-Chicago)
Carol Gattis (University of Arkansas)
Jeffery Maxey (University of Central Florida)
Giovanni Parmigiani (Duke University)
David Powers (Clarkson University)
Shiva Saksena (University of North Carolina-Wilmington)

We again thank the reviewers of previous editions:

George C. Derringer (Battelle Columbus, Ohio, Division)
Danny Dyer (University of Texas-Arlington)
Herberg B. Eisenberg (West Virginia College of Graduate Studies)
Linda Gans (California State Polytechnic University)
K. G. Janardan (Eastern Michigan University)
H. Lennon (Coventry Polytechnic, Coventry, England)
Curtis C. McKnight (University of Oklahoma)
Chand Midha (University of Akron)
Balgobin Nandram (Worcester Polytechnic Institute)
Paul Nelson (Kansas State University)
Carol O'Connor (University of Louisville)
Norbert Oppenheim (City College of New York)
David Robinson (St. Cloud State University)
Dennis Wackerly (University of Florida)
Donald L. Woods (Texas A&M University)

CHAPTER ONE

Introduction

Objective

To identify the role of statistics in the analysis of data from engineering and the sciences

Contents

1.1 Statistics: The Science of Data

According to *The Random House College Dictionary*, statistics is "the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data." In short, **statistics** is the **science of data**.

> **Definition 1.1**
>
> **Statistics** is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data.

The science of statistics is commonly applied to two types of problems:

1. Summarizing, describing, and exploring data
2. Using sample data to infer the nature of the data set from which the sample was selected

As an illustration of the descriptive applications of statistics, consider the United States census, which involves the collection of a data set that purports to characterize the socioeconomic characteristics of the approximately 250 million people living in the United States. Managing this enormous mass of data is a problem for the computer scientist, and describing the data utilizes the methods of statistics. Similarly, an engineer uses statistics to describe the data set consisting of the daily emissions of sulfur oxides of an industrial plant recorded for 365 days last year. The branch of statistics devoted to these applications is called **descriptive statistics**.

> **Definition 1.2**
>
> The branch of statistics devoted to the organization, summarization, and description of data sets is called **descriptive statistics**.

Sometimes the phenomenon of interest is characterized by a data set that is either physically unobtainable, or too costly or time-consuming to obtain. In such situations, we sample the data set and use the sample information to infer its nature. To illustrate, suppose the phenomenon of interest is the waiting time for a data-processing job to be completed. You might expect the waiting time to depend on such factors as the size of the job, the computer utilization factor, etc. In fact, if you were to run the same job over and over again on the computer, the waiting times would vary, even for the same computer utilization factor. Thus, the phenomenon "waiting time before job processing" is characterized by a large data set that exists only conceptually (in our minds). To determine the nature of this data set, we *sample* it—i.e., we process the job a number n of times, record the waiting time for each run, and then use this sample of n waiting times to infer the nature of the large conceptual data set of interest. The branch of statistics used to solve this problem is called **inferential statistics**.

In statistical terminology, the data set that we want to describe, the one that characterizes a phenomenon of interest to us, is called a **population**. A **sample** is a subset of data selected from a population. Sometimes the words *population* and *sample* are used to represent the objects upon which the measurements are taken. In a particular situation, the meaning attached to these terms will be clear by the context in which they are used.

Definition 1.3

A **population** is a data set that is the target of our interest.

Definition 1.4

A **sample** is a subset of data selected from a population.

Definition 1.5

The branch of statistics concerned with using sample data to make an inference about a population is called **inferential statistics**.

CASE STUDY 1.1 / Contamination of Fish in the Tennessee River

Chemical and manufacturing plants often discharge toxic waste materials into nearby rivers and streams. These toxicants have a detrimental effect on the plant and animal life inhabiting the river and the river's bank. One type of pollutant, commonly known as DDT, is especially harmful to fish and, indirectly, to people. The Food and Drug Administration sets the limit for DDT content in individual fish at 5 parts per million (ppm). Fish with DDT content exceeding this limit are considered potentially hazardous to people if consumed. A study was undertaken to examine the DDT content of fish inhabiting the Tennessee River (in Alabama) and its tributaries.

The Tennessee River flows in a west–east direction across the northern part of the state of Alabama, through Wheeler Reservoir, a national wildlife refuge. Ecologists fear that contaminated fish migrating from the mouth of the river to the reservoir could endanger other wildlife that prey on the fish. This concern is more than academic. A manufacturing plant was once located along Indian Creek, which enters the Tennessee River 321 miles upstream from the mouth. Although the plant has been inactive for over 10 years, there is evidence that the plant discharged toxic materials into the creek, contaminating all the fish in the immediate area. Have the fish in the Tennessee River and its tributary creeks also been contaminated? And if so, how far upstream have the contaminated fish migrated? To answer these and other questions, members of the U.S. Army Corps of Engineers in the summer of 1980 collected fish specimens at different locations along the Tennessee River and three tributary creeks: Flint Creek (which enters the river 309 miles upstream from the river's mouth), Limestone Creek (310 miles upstream), and Spring Creek (282 miles upstream). Each fish was first weighed (in grams) and measured (length in centimeters), then the fillet of the fish was extracted and the DDT concentration (in parts per million) in the fillet was measured.

Appendix III contains the length, weight, and DDT measurements for a total of 144 fish specimens.* Obviously, not all the fish in the Tennessee River and its tributaries were captured. Consequently, the data are based on a sample collected from the population of all fish inhabiting the Tennessee River. Here, the words *population* and *sample* are used to describe the objects upon which the measurements are taken, i.e., the fish. We could also use the terms to represent data sets. For example, the 144 DDT measurements represent a sample collected from the population consisting of DDT measurements for all fish inhabiting the river.

Notice that the data set also contains information on the location (i.e., where the fish were captured) and species of the fish. Three species of fish were examined: channel catfish, largemouth bass, and smallmouth buffalo. The different symbols for location are interpreted as follows. The first two characters represent the river or creek, and the remaining characters represent the distance (in miles) from the mouth of the river or creek. For example, FCM5 indicates that the fish was captured in Flint Creek (FC), 5 miles upstream from the mouth of the creek (M5). Similarly, TRM380 denotes a fish sample collected from the Tennessee River (TR), 380 miles upstream from the river's mouth (M380). In subsequent chapters, we will use the data in Appendix III to compare the DDT contents of fish at different locations and among the different species, and to determine the relationship (if any) of length and weight to DDT content.

EXERCISES

1.1 Pesticides applied to an extensively grown crop can result in inadvertent ambient air contamination. *Environmental Science & Technology* (Oct. 1993) reported on thion residues of the insecticide chlorpyrifos used on dormant orchards in the San Joaquin Valley, California. Ambient air specimens were collected daily at an orchard site during an intensive period of spraying—a total of 13 days—and the thion level (ng/m^3) was measured each day.

a. Identify the population of interest to the researchers.

b. Identify the sample.

1.2 Research engineers with the University of Kentucky Transportation Research Program have collected data on accidents occurring at intersections in Lexington, Kentucky, over a period of 5 years. One of the goals of the study was to compare the average number of left-turn accidents at locations with and without left-turn-only lanes to develop numerical warrants (or guidelines) for the installation of left-turn lanes.

a. What is the population of interest?

b. What is the sample?

c. How can the sample information be used to attain the researchers' goal?

---

*Source: U.S. Army Corps of Engineers, Mobile District, Alabama.

1.3 Electrical engineers recognize that high neutral current in computer power systems is a potential problem. To determine the extent of the problem, a survey of the computer power system load currents at 146 U.S. sites was taken (*IEEE Transactions on Industry Applications*, July/Aug. 1990). The survey revealed that less than 10% of the sites had high neutral to full-load current ratios.

a. Identify the population of interest.
b. Identify the sample.
c. Use the sample information to make an inference about the population.

1.4 Researchers have developed a new precooling method for preparing Florida vegetables for market. The system employs an air and water mixture designed to yield effective cooling with a much lower water flow than conventional hydrocooling. To compare the effectiveness of the two systems, 20 batches of green tomatoes were divided into two groups; one group was precooled with the new method, and the other with the conventional method. The water flow (in gallons) required to effectively cool each batch was recorded.

a. Identify the population, the samples, and the type of statistical inference to be made for this problem.
b. How could the sample data be used to compare the cooling effectiveness of the two systems?

1.5 Computer tomography (CT) scanners are highly sensitive, visual computer systems designed to aid a physician's diagnosis by generating radiographlike images of inner organs and physiological functions. Suppose you want to estimate the average *scan time*—that is, the average time required for a CT scanner to project an image. Describe how you could collect the sample data necessary to make the desired inference. What is the population of interest?

1.6 Checking all manufactured items coming off an assembly line for defectives would be a costly and time-consuming procedure. One effective and economical method of checking for defectives involves the selection and examination of a portion of the items by a quality control engineer. The percentage of examined items that are defective is computed and then used to estimate the percentage of all items manufactured on the line that are defective. Identify the population, the sample, and a type of statistical inference to be made for this problem.

1.2 Types of Data

Data can be one of two types, quantitative or qualitative. **Quantitative data** are those that represent the quantity or amount of something, measured on a numerical scale. For example, the power frequency (measured in megahertz) of a semiconductor is a quantitative variable, as is the waiting time (measured in seconds) before a computer job begins processing. In contrast, **qualitative** (or **categorical**) **data** possess no quantitative interpretation. They can only be classified. The set of n occupations corresponding to a group of n engineering graduates is a qualitative data set. A list of the manufacturers of n minicomputers owned by n small businesses is a set of qualitative data.*

*A finer breakdown of data types into nominal, ordinal, interval, and ratio data is possible. **Nominal** data are qualitative data with categories that cannot be meaningfully ordered. **Ordinal** data are also qualitative data, but a distinct ranking of the groups from high to low exists. **Interval** and **ratio** data are two different types of quantitative data. For most statistical applications (and all the methods presented in this introductory text), it is sufficient to classify data as either quantitative or qualitative.

Definition 1.6

Quantitative data are those that represent the quantity or amount of something.

Definition 1.7

Qualitative data are those that have no quantitative interpretation, i.e., they can only be classified into categories.

EXAMPLE 1.1

Refer to the data set in Appendix III (see Case Study 1.1). Classify each of the five variables in the data set (location, species, length, weight, and DDT concentration) as quantitative or qualitative.

Solution

Length (in centimeters), weight (in grams), and DDT concentration (in parts per million) are all measured on a numerical scale; thus, they represent quantitative data. In contrast, location and species cannot be measured on a quantitative scale; they can only be classified (e.g., channel catfish, largemouth bass, and smallmouth buffalo for species). Consequently, data on location and species are qualitative.

The proper statistical tool used to describe and analyze data will depend on the type of data. Consequently, it is important to differentiate between quantitative and qualitative data.

EXERCISES

1.7 Refer to the *IEEE Transactions on Industry Applications* (July/Aug. 1990) survey of computer power system load currents in Exercise 1.3. In addition to the ratio of neutral current to full-load current, the researchers also recorded the type of load (line-to-line or line-to-neutral) and the computer system vendor. Identify the type of data for each variable recorded.

1.8 The *Journal of Performance of Constructed Facilities* (Feb. 1990) reported on the performance dimensions of water distribution networks in the Philadelphia area. For one part of the study, the following data were collected for a sample of water pipe sections. Identify the data as quantitative or qualitative.

a. Pipe diameter (inches)
b. Pipe material
c. Age (year of installation)
d. Location
e. Pipe length (feet)

The theory of statistics uses *probability* to measure the uncertainty associated with an inference. It enables us to calculate the probabilities of observing specific samples, under specific assumptions about the population. The statistician then uses these probabilities to evaluate the uncertainties associated with sample inferences.

Thus, the major contribution of statistics is that it enables us to make inferences—estimates of and decisions about population parameters—with a known measure of uncertainty. It enables us to evaluate the *reliability* of inferences based on sample data.

Although we will present some useful methods for exploring and describing data sets (Chapter 2), the major emphasis in this text and in modern statistics is in the area of inferential statistics. The flowchart in Figure 1.1 is provided as an outline of the chapters in this text and as a guide when selecting the statistical method appropriate for your particular analysis.

1.4 Summary

Statistics—the **science of data**—is concerned with two types of problems: (1) summarizing and describing data (**descriptive statistics**), and (2) using **sample** data to make inferences about a large set of data—a **population**—from which the sample has been selected (**inferential statistics**).

The appropriate statistical method for describing and analyzing the data will depend on whether the data are **quantitative** or **qualitative**. These methods allow us to make inferences about a population and also provide a **measure of reliability** for the inference.

Descriptive statistics is the topic of Chapter 2. The remaining chapters are devoted to inferential statistics.

SUPPLEMENTARY EXERCISES

1.11 The reliability of a computer system is measured in terms of the lifelength of a specified hardware component (for example, the disk drive). To estimate the reliability of a particular system, 100 computer components are tested until they fail, and their lifelengths are recorded.

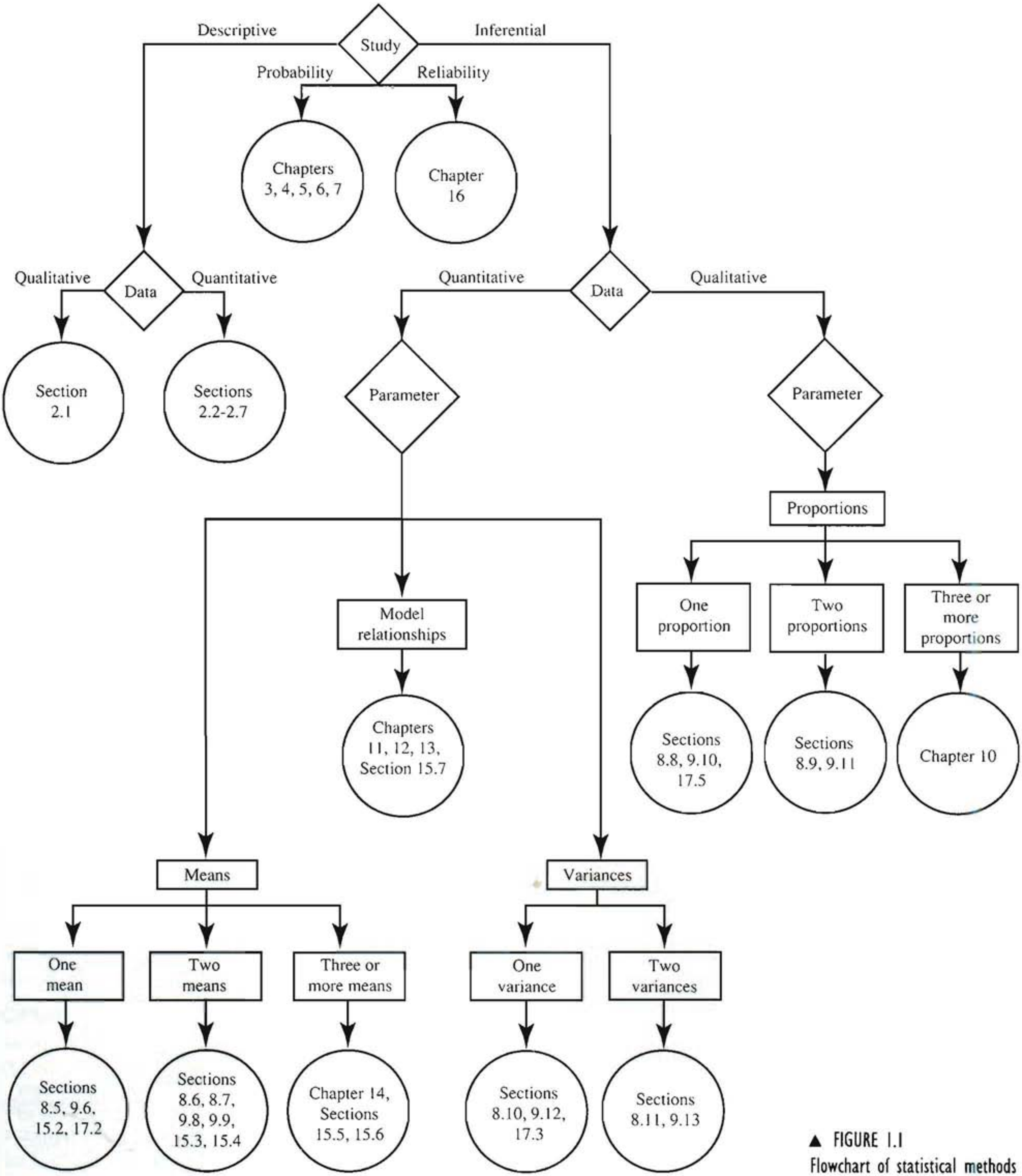
a. What is the population of interest?

b. What is the sample?

c. Are the data quantitative or qualitative?

d. How could the sample information be used to estimate the reliability of the computer system?

1.12 Hundreds of sea turtle hatchlings, instinctively following the bright lights of condominiums, wandered to their deaths across a coastal highway in Florida (*Tampa Tribune*, Sept. 16, 1990). This incident led researchers to begin experimenting with special low-pressure sodium lights. One night, 60 turtle hatchlings were released on a dark beach and their direction of travel noted. The next night, the special lights were installed and the same 60 hatchlings were released. Finally, on the third night, tar paper was placed over the sodium



▲ **FIGURE 1.1**
Flowchart of statistical methods described in the text

lights. Consequently, the direction of travel was recorded for each hatchling under three experimental conditions—darkness, sodium lights, and sodium lights covered with tar paper.

a. Identify the population of interest to the researchers.
b. Identify the sample.
c. What type of data were collected, quantitative or qualitative?

1.13 Every 5 years the Mechanics Division of the American Society of Engineering Education (ASEE) conducts a nationwide survey on undergraduate mechanics education at colleges and universities. In the latest survey, 66 out of the 100 colleges sampled covered fluid statics in their undergraduate engineering program (*Engineering Education,* April 1986).

a. What is the population of interest to the ASEE? The sample?
b. What type of data, quantitative or qualitative, are collected?
c. Use the sample information to make an inference about the population.

1.14 State whether each of the following data sets is quantitative or qualitative.

a. Arrival times of 16 reflected seismic waves
b. Types of computer software used in a database management system
c. Brands of calculator used by 100 engineering students on campus
d. Ash contents in pieces of coal from three different mines
e. Mileages attained by 12 automobiles powered by alcohol
f. Numbers of print characters per line of computer output for 20 line printers
g. Shift supervisors in charge of computer operations at an airline company
h. Accident rates at 46 machine shops

1.15 The data in the accompanying table were obtained from the Environmental Protection Agency (EPA) 1993 *Gas Mileage Guide* for new automobiles. State whether each of the variables measured is quantitative or qualitative.

Model Name	*Manufacturer*	*Transmission*	*Engine Size (liters)*	*Number of Cylinders*	*Estimated City Miles/Gallon*	*Estimated Highway Miles/Gallon*
NSX	Acura	Automatic	3.0	6	18	23
Colt	Dodge	Manual	1.5	4	32	40
318i	BMW	Automatic	1.8	4	22	30
Aerostar	Ford	Automatic	4.0	6	16	22
Camry	Toyota	Manual	2.2	4	22	30

Source: 1993 *Gas Mileage Guide*, EPA Fuel Economy Estimates, Oct. 1992.

COMPUTER LAB: Entering and Listing Data

In the Computer Lab sections of this text, we give the commands necessary to conduct a statistical analysis of data using one of two statistical software packages—SAS or MINITAB. These two packages were selected because of their current popularity, ease of use, and availability at most university computing centers. In addition, both packages have versions available for large mainframe computers and for personal computers (PCs).

Both software packages utilize the following three basic types of instructions:

1. **Data entry commands:** Instructions on how the data will be entered
2. **Input data values:** The values of the variables in the data set
3. **Statistical analysis commands:** Instructions on what type of analysis is to be conducted on the data

In this section we give the **data entry commands** for each package. That is, we give the commands that will enable you to create a data set ready for analysis. (The appropriate statistical analysis commands are provided in the relevant sections of the text.) The data set of interest from Appendix III consists of location, species, length, weight, and DDT levels of a sample of five fish. The data are listed in Table 1.1.

Note: With few exceptions, the commands provided in the following sections are appropriate for the large mainframe and PC versions of both software packages. When a mainframe computer is being used, however, these statements must be preceded by the job control language (JCL) commands required at your institution.

TABLE 1.1 Five Measurements Selected from Appendix III

Observation	Location	Species	Length	Weight	DDT Concentration
1	FCM5	Catfish	42.5	732	10.00
8	LCM3	Catfish	48.0	1,151	7.70
31	TRM280	Buffalo	49.0	1,763	4.50
43	TRM285	Bass	28.5	778	.48
73	TRM300	Buffalo	35.5	1,300	1.30

SAS

Command line

```
 1   DATA FISH;
 2   INPUT LOCATION $ SPECIES $ LENGTH WEIGHT DDT;  } Data entry
 3   LWRATIO = LENGTH/WEIGHT;                       } instructions
 4   CARDS;
 5   FCM5   CATFISH 42.5  732 10.00 }
 6   LCM3   CATFISH 48.0 1151  7.70 }  Input data values
 7   TRM280 BUFFALO 49.0 1763  4.50 }  (1 observation per line)
 8   TRM285 BASS    28.5  778  0.48 }
 9   TRM300 BUFFALO 35.5 1300  1.30 }
10   PROC PRINT;                       Print instruction
```

COMMAND 1 FISH is an arbitrarily chosen name used to identify the data set. (Data set names are restricted to a maximum length of eight characters.)

COMMAND 2 LOCATION, SPECIES, LENGTH, WEIGHT, and DDT are arbitrarily chosen names for the variables in the data set. (Variable names are also restricted to a maximum length of eight characters.) A dollar sign ($) must follow the name of any nonnumeric variable in the data set.

COMMAND 3 LWRATIO (length-to-weight ratio) is calculated by dividing LENGTH by WEIGHT. (The standard arithmetic operations symbols, +, −, *, and /, are used for addition, subtraction, multiplication, and division, respectively.)

COMMAND 4 CARDS signals SAS that the input data values are to follow.

COMMANDS 5–9 Each data line gives the values of the variables in the data set for a single observation (fish) in the order in which the variables are listed in the INPUT command. Input data values must be separated by at least one blank space; commas are not permitted in numeric values.

COMMAND 10 The PRINT procedure (PROC) will produce a listing of the entire data set (see Figure 1.2). In addition to the INPUT variables, the data set will contain any variables created using the standard arithmetic operations (e.g., LWRATIO) in command line 3.

GENERAL All SAS commands must end with a semicolon; the only exceptions to this rule are the input data values.

FIGURE 1.2 ▶ SAS output: Listing of the data in Table 1.1

OBS	LOCATION	SPECIES	LENGTH	WEIGHT	DDT	LWRATIO
1	fcm5	catfish	42.5	732	10.00	0.058060
2	lcm3	catfish	48.0	1151	7.70	0.041703
3	trm280	buffalo	49.0	1763	4.50	0.027794
4	trm285	bass	28.5	778	0.48	0.036632
5	trm300	buffalo	35.5	1300	1.30	0.027308

MINITAB

Command line

```
1   READ C1 C2 C3 C4 C5        Data entry instructions
2   1 1 42.5  732 10.00  )
3   2 1 48.0 1151  7.70  |     Input data values
4   3 2 49.0 1763  4.50  }     (1 observation per line)
5   4 3 28.5  778  0.48  |
6   5 2 35.5 1300  1.30  )
7   DIVIDE C3 BY C4 PUT INTO C6
8   NAME C1 = 'LOCATION' C2 = 'SPECIES'    Data entry instructions
9   NAME C3='LENGTH' C4='WEIGHT' C5='DDT' C6='LWRATIO'
10  PRINT C1-C6                Print instruction
11  STOP
```

COMMAND 1 The five variables to be read onto the MINITAB "worksheet" are identified by the "columns" into which they are placed: C1, C2, C3, C4, and C5. (MINITAB does not, in general, recognize variable names.) Thus, location will be read in column 1, species in column 2, etc.

COMMANDS 2–6 Each data line gives the values of the variables read in the worksheet columns for a single observation (fish). Input data values must be separated by at least one blank space; commas are not permitted. MINITAB also requires all data used in statistical analysis to be numerical. For example, the values of the nonnumeric variable *location* are converted to numbers in C1. (Arbitrarily let 1 represent FCM5, 2 represent LCM3, etc.)

COMMAND 7 MINITAB uses the word commands ADD, SUBTRACT, MULTIPLY, and DIVIDE to perform the usual arithmetic operations on variables. The ratio of length (C3) to weight (C4) is stored in C6.

COMMANDS 8–9 For labeling printed output, the NAME command can be used to give names to the variables stored in the worksheet columns. If the NAME command is omitted, the columns will be labeled C1, C2, etc., on the MINITAB printouts.

COMMAND 10 The PRINT command will produce a listing of the data in the MINITAB worksheet for the specified variables (columns). (See Figure 1.3.)

COMMAND 11 All MINITAB programs terminate with the STOP command.

GENERAL MINITAB permits you to insert extraneous words within each command to help you follow the logic of the program. For example, command line 1 could be entered as follows:

```
READ LOCATION IN C1, SPECIES IN C2, LENGTH IN C3, WEIGHT IN C4, DDT IN C5
```

FIGURE 1.3 ▶
MINITAB output: Listing of data in Table 1.1

ROW	Location	Species	Length	Weight	DDT	LWRatio
1	1	1	42.5	732	10.00	0.0580601
2	2	1	48.0	1151	7.70	0.0417029
3	3	2	49.0	1763	4.50	0.0277935
4	4	3	28.5	778	0.48	0.0366324
5	5	2	35.5	1300	1.30	0.0273077

COMPUTER LAB: Accessing an External Data File (Optional)

Data created by other software and saved in an external file as an ASCII data set can also be accessed and analyzed by SAS and MINITAB. For example, the full data set of Appendix III (DDT measurements and other data for 144 fish specimens) is saved in an ASCII file called FISH.DAT on a 3.5″ micro disk or 5.25″ floppy disk available from the publisher (see the Preface). The program lines shown here give the commands for reading and listing the data on this external file.

SAS

Command line

```
DATA FISH;
INFILE 'FISH.DAT';
INPUT LOCATION $ SPECIES $ LENGTH WEIGHT DDT;
LWRATIO=LENGTH/WEIGHT;
PROC PRINT;
```

MINITAB

Command line

```
READ 'FISH.DAT' C1-C5
DIVIDE C3 C4 C6
NAME C1='LOCATION' C2='SPECIES' C3='LENGTH'
     C4='WEIGHT' C5='DDT' C6='LWRATIO'
PRINT C1-C6
```

References

Careers in Statistics. American Statistical Association and the Institute of Mathematical Statistics, 1974.

Huff, D. *How to Lie with Statistics*. New York: W. W. Norton and Co., 1978.

Ryan, B. F., Joiner, B. L., and Ryan, T. A. *Minitab Handbook*, 2nd ed. Boston: PWS-Kent, 1990.

SAS Procedures Guide for Personal Computers, Version 6 ed. Cary, N.C.: SAS Institute, Inc., 1986.

SAS *User's Guide: Basics*, Version 6 ed. Cary, N.C.: SAS Institute, Inc., 1991.

Tanur, J. M., Mosteller, F., Kruskal, W. H., Link, R. F., Pieters, R. S., and Rising, G. R., eds. *Statistics: A Guide to the Unknown*, 3rd ed. Pacific Grove, California: Wadsworth, Brooks/Cole, 1989.

Index

The accompanying table gives the number of scrams at each of 56 U.S. nuclear reactor units in a recent year. A MINITAB printout showing both a graphical and numerical description of the data is provided.

Number of Scrams

1	0	3	1	4	2	10	6	5	2	0	3	1	5
4	2	7	12	0	3	8	2	0	9	3	3	4	7
2	4	5	3	2	7	13	4	2	3	3	7	0	9
4	3	5	2	7	8	5	2	4	3	4	0	1	7

Minitab printout for Exercise 2.47

```
Stem-and-leaf of SCRAMS       N  = 56
Leaf Unit = 0.10

     6     0 000000
    10     1 0000
    19     2 000000000
   (10)    3 0000000000
    27     4 00000000
    19     5 00000
    14     6 0
    13     7 000000
     7     8 00
     5     9 00
     3    10 0
     2    11
     2    12 0
     1    13 0

                    ---------------
          --------I    +         I-----------------          *    *
                    ---------------
          +---------+---------+---------+---------+---------+------SCRAMS
         0.0       2.5       5.0       7.5      10.0      12.5

                  N       MEAN     MEDIAN     TRMEAN      STDEV     SEMEAN
SCRAMS           56      4.036      3.000      3.820      3.027      0.404

                MIN        MAX         Q1         Q3
SCRAMS        0.000     13.000      2.000      5.750
```

a. Fully interpret the results.
b. Would you expect to observe a nuclear reactor in the future with 11 unplanned scrams? Explain.

2.48 Industrial engineers periodically conduct "work measurement" analyses to determine the time required to produce a single unit of output. At a large processing plant, the number of total worker-hours required per day to perform a certain task was recorded for 50 days. The data are shown here.

128	119	95	97	124	128	142	98	108	120
113	109	124	132	97	138	133	136	120	112
146	128	103	135	114	109	100	111	131	113
124	131	133	131	88	118	116	98	112	138
100	112	111	150	117	122	97	116	92	122

a. Compute the mean, median, and mode of the data set.
b. Find the range, variance, and standard deviation of the data set.
c. Construct the intervals $\bar{y} \pm s$, $\bar{y} \pm 2s$, and $\bar{y} \pm 3s$. Count the number of observations that fall within each interval and find the corresponding proportions. Compare the results to the Empirical Rule. Do you detect any outliers?
d. Construct a box plot for the data. Do you detect any outliers?
e. Find the 70th percentile for the data on total daily worker-hours. Interpret its value.

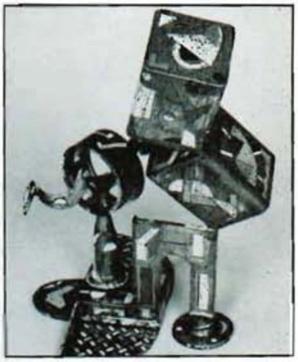
2.49 A marketing research study of consulting engineering services to industrial firms in the Midwest was recently conducted. The main goal of the study was to gather information that will enable consulting engineers to effectively market their services to industrial firms. Of the 70 firms surveyed, 40 indicated that they have no need for outside consulting engineering services. The accompanying table gives the primary reasons cited by the "nonneeders" and corresponding breakdown in percentages for both the large and small industrial firms in the survey.

Reason	*Large Firms*	*Small Firms*
Assistance obtained from corporate headquarters	62%	30%
No wastes, therefore, no need to improve	0	32
No improvements planned	0	24
Assistance obtained from staff engineers	19	6
Unfamiliar with consulting	10	2
Waiting for regulations	9	0
Other reasons	0	6
TOTALS	100%	100%

Source: Carey, R. J., and Brunner, J. A. "A study of marketing of consulting engineering services to industrial firms." *Journal of the Boston Society of Civil Engineers Section*, American Society of Civil Engineers, Vol. 71, Nos. 1 and 2, 1985, p. 152.

a. Construct a pie chart that describes the reasons cited for not needing consulting engineering services at large industrial firms.
b. Repeat part **a** for small industrial firms.
c. Compare the two pie charts in parts **a** and **b**. Do you detect major differences in the reasons cited by large and small firms?

2.50 The nuclear mishap on Three Mile Island near Harrisburg, Pennsylvania, on March 28, 1979, forced many local residents to evacuate their homes—some temporarily, others permanently. To assess the impact of the accident on the area population, a questionnaire was designed and mailed to a sample of 150 households within 2 weeks after the accident occurred. Two questions asked of the sampled residents were: (1) When did you learn about the accident? and (2) How did you learn about the accident? The responses to the two questions are illustrated in the frequency distributions shown at the top of page 66. Based on these graphical descriptions, find each of the following:
a. The percentage of the 150 respondents who learned about the accident on Wednesday afternoon
b. The percentage of the 150 respondents who learned about the accident on Friday
c. The percentage of the 150 respondents who learned about the accident from a radio report
d. The percentage of the 150 respondents who learned about the accident from television



CHAPTER SEVEN

Sampling Distributions

Objective

To present methods for finding the probability distribution (sampling distribution) of a statistic; to identify the sampling distributions for some useful statistics

Contents

7.1 Random Sampling

Recall from Chapter 1 that statistical inference involves sampling from a well-defined population. How a sample is selected from a population is of vital importance because the probability of an observed sample will be used to infer the characteristics of the sampled population.

To illustrate, suppose you deal yourself 4 cards from a deck of 52 cards and all 4 cards are aces. Do you conclude that your deck is an ordinary bridge deck, containing only 4 aces, or do you conclude that the deck is stacked with more than 4 aces? It depends on how the cards were drawn. If the 4 aces are always placed at the top of a standard bridge deck, drawing 4 aces is not unusual—it is certain. On the other hand, if the cards are thoroughly mixed, drawing 4 aces in a sample of 4 cards is highly improbable. The point, of course, is that in order to use the observed sample of 4 cards to draw inferences about the population (the deck of 52 cards), you need to know how the sample was selected from the deck.

One of the simplest and most frequently employed sampling procedures produces what is known as a **random sample**.

> **Definition 7.1**
>
> If n elements are selected from a population in such a way that every set of n elements in the population has an equal probability of being selected, the n elements are said to be a **random sample.***

EXAMPLE 7.1

An experiment was conducted in which each of 10 different antiscalants was added to an aliquot of brine. One of the 10 brine solutions is to be selected, filtered, and the amount of silica determined. How would you select the brine solution so that the choice is random?

Solution

If the choice is to be random, each brine solution must have the same probability of being drawn. That is, each solution should have a probability of $\frac{1}{10}$ of being selected. A method to achieve the objective of equal selection probabilities is to *thoroughly* mix the 10 brine solutions and *blindly* pick one of the solutions. If this procedure were

*Strictly speaking, this is a **simple random sample**. There are many different types of random samples. For example, a **stratified random sample** is obtained by partitioning the population into groups (*strata*) and selecting a random sample of elements from each group; a **cluster sample** involves randomly selecting groups (or *clusters*) of elements from the population and sampling *all* elements in each cluster; and a **systematic sample** is collected by systematically selecting every *k*th element from the population. Since it is the most common, we focus our discussion on the random sample.

repeatedly used, each time replacing the selected solution, a particular solution should be chosen approximately $\frac{1}{10}$ of the time in a long series of draws. This method of sampling is known as **random sampling**.

How can a random sample be generated? If a population is not too large and the elements can be numbered on slips of paper, poker chips, etc., you can physically mix the slips of paper or chips and remove n elements from the total. The numbers that appear on the chips selected would indicate the population elements to be included in the sample. Such a procedure will not guarantee a random sample, because it is often difficult to achieve a thorough mix, but it provides a reasonably good approximation to random sampling.

Another, more formal, technique is to use a **table of random numbers**. Random number tables are generated by computer in such a way that every number of the same length (2-digit, 3-digit, 4-digit, etc.), occurs with equal probability. Further, the occurrence of any one number in the table is independent of any of the other numbers in the table. Consequently, the numbers that are selected from a random number table identify the elements to be included in the random sample.

Although this method of random sampling is easy to implement, it can become time-consuming and tedious if the number of observations in the sample is large. Therefore, large-sample scientific studies rely on computers (with built-in random number generators) to automatically select the random sample.

EXAMPLE 7.2

Suppose you want to randomly sample 5 (we will keep the number in the sample small to simplify our example) from a shipment of 100,000 bolts for quality control testing.

a. Use a random number table to select this random sample.
b. Use the computer to generate the random sample.

Solution

a. Since there are 100,000 bolts in the shipment (target population), we first number the bolts from 1 to 100,000. Then, we turn to a table of random numbers (see Table 6, Appendix II), and select a page, say, the first page. (A partial reproduction of the first page of Table 6 is shown in Table 7.1 on page 292.) Now, randomly select a starting number, say, the random number appearing in the third row, second column. This number is 48360. Proceed down the second column to obtain the remaining four random numbers. (Proceeding down or across is an arbitrary choice.) The five selected random numbers are shaded in Table 7.1. Using the first five digits to represent the bolts from 1 to 99,999 and the number 00000 to represent bolt 100,000, you can see that the bolts numbered

48,360   93,093   39,975   6,907   72,905

should be included in your sample.

TABLE 7.1 Partial Reproduction of Table 6 of Appendix II

Row \ Column	1	2	3	4	5	6
1	10480	15011	01536	02011	81647	91646
2	22368	46573	25595	85393	30995	89198
3	24130	48360	22527	97265	76393	64809
4	42167	93093	06243	61680	07856	16376
5	37570	39975	81837	16656	06121	91782
6	77921	06907	11008	42751	27756	53498
7	99562	72905	56420	69994	98872	31016
8	96301	91977	05463	07972	18876	20922
9	89579	14342	63661	10281	17453	18103
10	85475	36857	53342	53988	53060	59533
11	28918	69578	88231	33276	70997	79936
12	63553	40961	48235	03427	49626	69445
13	09429	93969	52636	92737	88974	33488
14	10365	61129	87529	85689	48237	52267
15	07119	97336	71048	08178	77233	13916

b. Almost all of the commercial statistical software packages available (e.g., SAS and MINITAB) have procedures for generating random samples. The output from a SAS program designed to generate a sample of size 5 from a population of 100,000 elements is displayed in Figure 7.1.* From the printout, you can see that bolts numbered

6,181   35,982   76,110   58,667   59,592

comprise the random sample of size 5.

FIGURE 7.1 ► SAS-generated random sample for Example 7.2

```
OBS    SELECT

 1       6181
 2      35982
 3      76110
 4      58667
 5      59592
```

Although random sampling represents one of the simplest of the multitude of sampling techniques available for research, most of the statistical methods presented

*The commands for generating a random sample in SAS and MINITAB are provided in the Computer Lab at the end of this chapter.

in this text assume that such a sample has been collected. If a researcher knows that a sample is nonrandom, any inferences derived from the analysis may be invalid.

EXERCISES

7.1 Refer to the DDT levels for 144 contaminated fish specimens, Appendix III. Use Table 6 of Appendix II or a computer to generate a random sample of $n = 10$ DDT levels from the data set.

7.2 Refer to the CPU times for 1,000 computer jobs, Appendix IV. Use Table 6 of Appendix II or a computer to generate a random sample of $n = 25$ CPU times from the data set.

7.3 Refer to the percentage iron contents for 390 iron-ore specimens, Appendix V. Use Table 6 of Appendix II or a computer to generate a random sample of $n = 5$ percentage iron measurements from the data set.

7.4 Laboratory tests were conducted to compare the permeability of open-graded asphalt concrete with asphalt contents of 3% and 7% (*Journal of Testing and Evaluation*, July 1981). Eight batches of cement were prepared—four with a 3% asphalt mix and four with a 7% asphalt mix. Use Table 6 of Appendix II to randomly select the four batches that receive the 3% asphalt mix.

7.5 One of the most infamous examples of improper sampling was conducted in 1936 by the *Literary Digest* to determine the winner of the Landon–Roosevelt presidential election. The poll, which predicted Landon to be the winner, was conducted by sending ballots to a random sample of persons selected from among the names listed in the telephone directories of that year. In the actual election, Landon won in Maine and Vermont but lost in the remaining 46 states. The *Literary Digest*'s erroneous forecast is believed to be the major reason for its eventual failure.

What was the cause of the *Digest*'s erroneous forecast? That is, why might the sampling procedure described above yield a sample of people whose opinions might be biased in favor of Landon?

7.6 Every 10 years the United States population census provides essential information about our nation and its people. The basic constitutional purpose of the census is to apportion the membership of the House of Representatives among the states. However, the census has many other important uses. For example, private business uses the census for plant location and marketing.

The 1990 census included questions on age, sex, race, marital status, family relationship, and income; this census was mailed to every household in the United States. In some cities, however, a series of questions was added for a 5% sample of the city's households. That is, each of a random sample of the city's households was mailed a census form that included additional questions. Suppose that a particular city contained 100,000 households and, of these, 5,000 were selected and mailed the longer census form.

a. If you worked for the Bureau of the Census and were assigned the task of selecting a random sample of 5,000 of the city's households, describe how you would proceed.
b. Suppose that one of the additional questions on the long form of the census concerned energy consumption. The city used this sample information to project the average energy consumption for the city's 100,000 households. Explain why it is important that the sample of 5,000 households be random.
c. Using the procedure you described in part **a**, randomly select a sample of 10 households from the 100,000 households in the city.

7.2 Sampling Distributions

Recall (Chapter 6) that the n measurements in a sample can be viewed as observations on n random variables, $y_1, y_2, \ldots, y_n$. Consequently, the sample mean $\bar{y}$, the sample variance s^2, and other statistics are functions of random variables—functions that we will use in the following chapters to make inferences about population parameters. Thus, a primary reason for presenting the theory of probability and probability distributions in the preceding chapters was to enable us to find and evaluate the properties of the probability distribution of a statistic. This probability distribution is often called the **sampling distribution** of the statistic. As is the case for a single random variable, its mean is the expected value of the statistic. Its standard deviation is called the **standard error** of the statistic.

Definition 7.2

The **sampling distribution** of a statistic is its probability distribution.

Definition 7.3

The **standard error** of a statistic is the standard deviation of its sampling distribution.

The mathematical techniques for finding the sampling distribution of a statistic are difficult to apply and, except for very simple examples, are beyond the scope of this text. We will introduce this topic in Section 7.3, where we will develop a procedure for using a computer to generate random samples from theoretical populations of data. We will use this simulated sampling procedure to draw many samples of a specified size, calculate the value of a statistic for each sample, and form a relative frequency histogram of these values. The resulting relative frequency histogram will be an approximation to the sampling distribution of the statistic.

Even if we are unable to find the exact mathematical form of the probability distribution of a statistic and are unable to approximate it using simulation, we can always find its mean and variance using the methods of Chapters 4–6. Then we can obtain an approximate description of the sampling distribution by applying the Empirical Rule.

7.3 Probability Distributions of Functions of Random Variables (Optional)

There are essentially three methods for finding the density function for a function of one or more random variables. Two of these—the moment generating function method and the transformation method—are beyond the scope of this text, but a discussion of them can be found in the references at the end of the chapter. The third method, which we will call the **cumulative distribution function method**, will be demonstrated with examples.

Suppose w is a function of one or more random variables. The cumulative distribution function method finds the density function for w by first finding the probability $P(w \leq w_0)$, which (dropping the subscript 0) is equal to $F(w)$. The density function $f(w)$ is then found by differentiating $F(w)$ with respect to w. We will demonstrate the method in Examples 7.3 and 7.4.

EXAMPLE 7.3

Suppose the random variable y has a density function

$$f(y) = \begin{cases} \dfrac{e^{-y/\beta}}{\beta} & \text{if } 0 \leq y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and let $w(y) = y^2$. Find the density function for the random variable w.

Solution

A graph of $w = y^2$ is shown in Figure 7.2 on page 296. We will denote the cumulative distribution functions of w and y as $G(w)$ and $F(y)$, respectively. We note from the figure that w will be less than w_0 whenever y is less than y_0; it follows that $P(w \leq w_0) = G(w_0) = F(y_0)$. Since $w = y^2$, we have $y_0 = \sqrt{w_0}$ and

$$F(y_0) = F(\sqrt{w_0}) = \int_{-\infty}^{\sqrt{w_0}} f(y)dy = \int_0^{\sqrt{w_0}} \frac{e^{-y/\beta}}{\beta}\,dy = -e^{-y/\beta}\Big]_0^{\sqrt{w_0}} = 1 - e^{-(\sqrt{w_0}/\beta)}$$

Therefore, the cumulative distribution function for w is

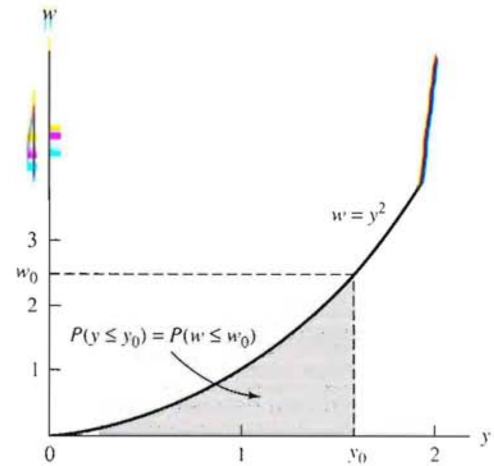
$$G(w) = 1 - e^{-(\sqrt{w}/\beta)}$$

Differentiating, we obtain the density function for w:

$$\frac{dG(w)}{dw} = f(w) = \frac{w^{-1/2}e^{-(\sqrt{w}/\beta)}}{2\beta}$$

FIGURE 7.2 ▶

A graph of $w = y^2$



EXAMPLE 7.4

If the random variables x and y possess a uniform joint density function over the unit square, then $f(x, y) = 1$ for $0 \le x \le 1$ and $0 \le y \le 1$. Find the density function for the sum $w = x + y$.

Solution

Each value of w corresponds to a series of points on the line $w_0 = x + y$ (see Figure 7.3). Written in the slope–intercept form, $y = w_0 - x$, this is the equation of a line with slope equal to -1 and y-intercept equal to w_0. The values of w that are less than or equal to w_0 are those corresponding to points (x, y) below the line $w_0 = x + y$. (This area is shaded in Figure 7.3.) Then, for values of the y-intercept w_0, $0 \le w_0 \le 1$, the probability that w is less than or equal to w_0 is equal to the volume of a solid over the shaded area shown in the figure. We could find this probability by multiple integration, but it is easier to obtain it with the aid of geometry. Each of the two equal sides of the triangle has length w_0. Therefore, the area of the shaded triangular region is $w_0^2/2$, the height of the solid over the region is $f(x, y) = 1$, and the volume is

$$P(w \le w_0) = F(w_0) = w_0^2/2$$

We now drop the subscript to obtain

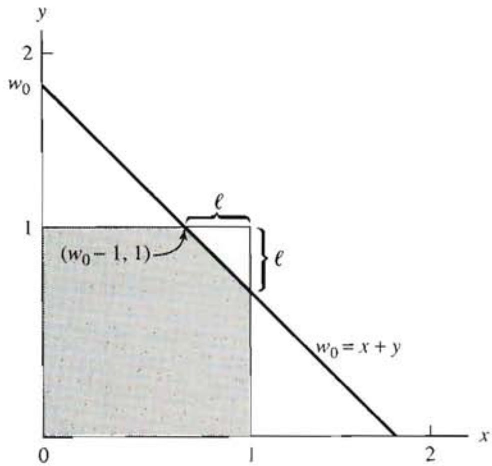
$$F(w) = w^2/2 \quad (0 \le w \le 1)$$

FIGURE 7.3 ▶
A graph showing the region of integration to find $F(w_0)$,

The equation for $F(w)$ is different over the interval $1 \le w \le 2$. The probability $P(w \le w_0) = F(w_0)$ is the integral of $f(x, y) = 1$ over the shaded area shown in Figure 7.4. The integral can be found by subtracting from 1 the volume corresponding to the small triangular (nonshaded) area that lies above the line $w_0 = x + y$. To find the length of one side of this triangle, we need to locate the point where the line $w_0 = x + y$ intersects the line $y = 1$. Substituting $y = 1$ into the equation of the line, we find

$$w_0 = x + 1 \quad \text{or} \quad x = w_0 - 1$$

FIGURE 7.4 ▶
A graph showing the region of integration to find $F(w_0)$, $1 \le w_0 \le 2$



The point $(w_0 - 1, 1)$ is shown in Figure 7.4. The two equal sides of the triangle each have length $\ell = 1 - (w_0 - 1) = 2 - w_0$. The area of the triangle lying above the line $w_0 = x + y$ is then

$$\begin{aligned} \text{Area} &= \frac{1}{2}(\text{Base})(\text{Height}) \\ &= \frac{1}{2}(2 - w_0)(2 - w_0) = \frac{(2 - w_0)^2}{2} \end{aligned}$$

Since the height of the solid constructed over the triangle is $f(x, y) = 1$, the probability that w lies above the line $w_0 = x + y$ is $(2 - w_0)^2/2$. Subtracting this probability from 1, we find the probability that w lies below the line to be

$$F(w_0) = P(w \le w_0) = 1 - \frac{(2 - w_0)^2}{2}$$

We drop the subscript and simplify to obtain

$$F(w) = -1 + 2w - w^2/2 \quad (1 \le w \le 2)$$

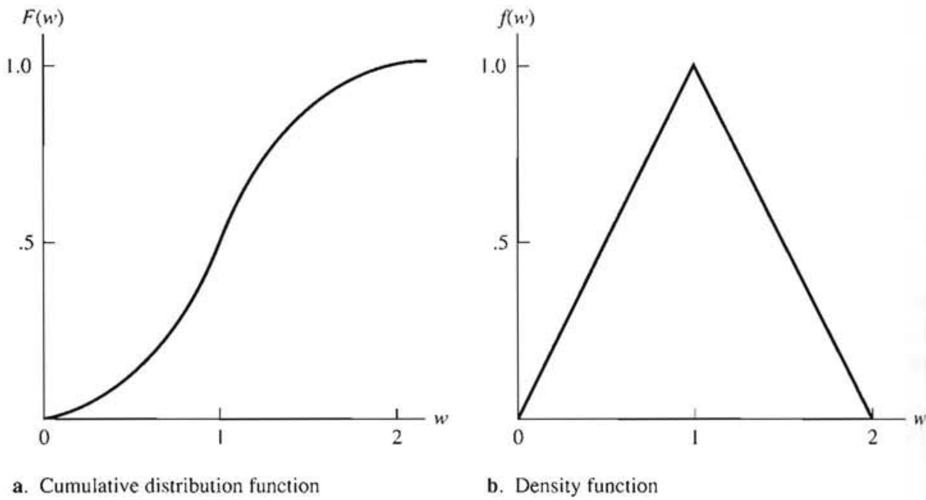
The density function for the sum of the two random variables x and y is now obtained by differentiating $F(w)$:

$$f(w) = \frac{dF(w)}{dw} = \frac{d(w^2/2)}{dw} = w \quad (0 \le w \le 1)$$

$$f(w) = \frac{dF(w)}{dw} = \frac{d(-1 + 2w - w^2/2)}{dw} = 2 - w \quad (1 \le w \le 2)$$

Graphs of the cumulative distribution function and the density function for $w = x + y$ are shown in Figures 7.5a and 7.5b, respectively. Note that the area under the density function over the interval $0 \le w \le 2$ is equal to 1.

FIGURE 7.5 ▶ Graphs of the cumulative distribution function and density function for $w = x + y$



One of the most useful functions of a single continuous random variable is the cumulative distribution function itself. We will show that if y is a continuous random variable with density function $f(y)$ and cumulative distribution function $F(y)$, then $w = F(y)$ has a uniform probability distribution over the interval $0 \le w \le 1$. Using a computer program for generating random numbers, we can generate a random sample of w values. For each value of w, we can solve for the corresponding value of y using the equation $w = F(y)$ and, thereby, obtain a random sample of y values from a population modeled by the density function $f(y)$. We will present this important transformation as a theorem, prove it, and then demonstrate its use with an example.

Theorem 7.1

Let y be a continuous random variable with density function $f(y)$ and cumulative distribution $F(y)$. Then the density function of $w = F(y)$ will be a uniform distribution defined over the interval $0 \le w \le 1$, i.e.,

$$f(w) = 1 \quad (0 \le w \le 1)$$

PROOF OF THEOREM 7.1 Figure 7.6 shows the graph of $w = F(y)$ for a continuous random variable y. You can see from the figure that there is a one-to-one correspondence between y values and w values, and that values of y corresponding to values of w in the interval $0 \le w \le w_0$ will be those in the interval $0 \le y \le y_0$. Therefore,

$$P(w \le w_0) = P(y \le y_0) = F(y_0)$$

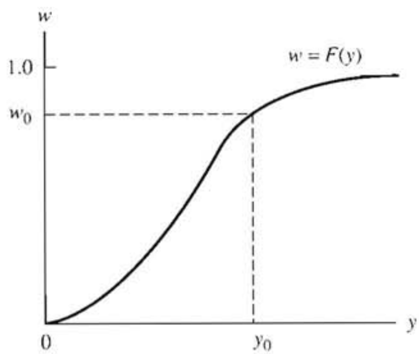
But since $w = F(y)$, we have $F(y_0) = w_0$. Therefore, we can write

$$P(w \le w_0) = F(y_0) = w_0$$

The cumulative distribution function for w is obtained by dropping the subscript:

$$F(w) = w$$

FIGURE 7.6 ▶ Cumulative distribution function $F(y)$



Finally, we differentiate over the range $0 \le w \le 1$ to obtain the density function:

$$f(w) = \frac{dF(w)}{dw} = 1$$

EXAMPLE 7.5

Use Theorem 7.1 to generate a random sample of $n = 3$ observations from an exponential distribution with $\beta = 2$.

Solution

The density function for the exponential distribution with $\beta = 2$ is

$$f(y) = \begin{cases} \frac{e^{-y/2}}{2} & \text{if } 0 \le y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and the cumulative distribution function is

$$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{0}^{y} \frac{e^{-t/2}}{2}\,dt = -e^{-t/2}\Big]_{0}^{y} = 1 - e^{-y/2}$$

If we let $w = F(y) = 1 - e^{-y/2}$, then Theorem 7.1 tells us that w has a uniform density function over the interval $0 \le w \le 1$.

To draw a random number y from the exponential distribution, we first randomly draw a value of w from the uniform distribution. This can be done by drawing a random number from Table 6 of Appendix II or using a computer. Suppose, for example, that we draw the random number 10480. This corresponds to the random selection of the value $w_1 = .10480$ from a uniform distribution over the interval $0 \le w \le 1$. Substituting this value of w_1 into the formula for $w = F(y)$ and solving for y, we obtain

$$\begin{aligned} w_1 &= F(y) = 1 - e^{-y_1/2} \\ .10480 &= 1 - e^{-y_1/2} \\ e^{-y_1/2} &= .8952 \\ \frac{-y_1}{2} &= -.111 \\ y_1 &= .222 \end{aligned}$$

If the next two random numbers selected are 22368 and 24130, then the corresponding values of the uniform random variable are $w_2 = .22368$ and $w_3 = .24130$. By substituting these values into the formula $w = 1 - e^{-y/2}$, you can verify that $y_2 = .506$ and $y_3 = .552$. Thus, $y_1 = .222$, $y_2 = .506$, and $y_3 = .552$ represent three randomly selected observations on an exponential random variable with mean equal to 2.

EXERCISES

7.7 Consider the density function

$$f(y) = \begin{cases} 2y & \text{if } 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the density function of w, where:
a. $w = y^2$ **b.** $w = 2y - 1$ **c.** $w = 1/y$

7.8 Consider the density function

$$f(y) = \begin{cases} e^{-(y-3)} & \text{if } y > 3 \\ 0 & \text{elsewhere} \end{cases}$$

Find the density function of w, where:
a. $w = e^{-y}$ **b.** $w = y - 3$ **c.** $w = y/3$

7.9 The amount y of paper used per day by a line printer at a university computing center has an exponential distribution with mean equal to five boxes (i.e., $\beta = 5$). The daily cost of the paper is proportional to $c = (3y + 2)$. Find the probability density function of the daily cost of paper used by the line printer.

7.10 An environmental engineer has determined that the amount y (in parts per million) of pollutant per water sample collected near the discharge tubes of an island power plant has probability density function

$$f(y) = \begin{cases} \dfrac{1}{10} & \text{if } 0 < y < 10 \\ 0 & \text{elsewhere} \end{cases}$$

A new cleaning device has been developed to help reduce the amount of pollution discharged into the ocean. It is believed that the amount a of pollutant discharged when the device is operating will be related to y by

$$a = \begin{cases} \dfrac{y}{2} & \text{if } 0 < y < 5 \\ \dfrac{2y - 5}{2} & \text{if } 5 < y < 10 \end{cases}$$

Find the probability density function of a.

7.11 Researchers at the University of California (Berkeley) have developed a switched-capacitor circuit for generating pseudorandom signals (*International Journal of Circuit Theory and Applications,* May/June, 1990). The intensity of the signal (voltage), y, is modeled using the Rayleigh probability distribution with mean μ. This continuous distribution has density function:

$$f(y) = \frac{y}{\mu}\exp^{-y^2/(2\mu)} \quad (y > 0)$$

Find the density function of the random variable $w = y^2$. Can you name the distribution?

7.12 Use Theorem 7.1 to draw a random sample of $n = 5$ observations from a distribution with probability density function

$$f(y) = \begin{cases} e^y & \text{if } y < 0 \\ 0 & \text{elsewhere} \end{cases}$$

7.13 Use Theorem 7.1 to draw a random sample of $n = 5$ observations from a beta distribution with $\alpha = 2$ and $\beta = 1$.

OPTIONAL EXERCISE

7.14 The total time x (in minutes) from the time a computer job is submitted until its run is completed and the time y the job waits in the job queue before being run have the joint density function

$$f(x, y) = \begin{cases} e^{-x} & \text{if } 0 \le y \le x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

The CPU time for the job (i.e., the length of time the job is in control of the computer's central processing unit) is given by the difference $w = x - y$. Find the density function of a job's CPU time. [*Hint:* You may use the facts that

$$
\begin{aligned}
P(w \le w_0) &= P(w \le w_0,\ x > w_0) + P(w \le w_0,\ x \le w_0) \\
&= P(x - w_0 \le y \le x,\ w_0 < x < \infty) + P(0 \le y \le x,\ 0 \le x \le w_0) \\
&= \int_{w_0}^{\infty}\int_{x-w_0}^{x} e^{-x}\,dy\,dx + \int_{0}^{w}\int_{0}^{x} e^{-x}\,dy\,dx
\end{aligned}
$$

and $\int ye^{-y}\,dy = -ye^{-y} + \int e^{-y}\,dy$ in determining the density function.]

7.4 Approximating a Sampling Distribution by Simulation

We explained in Section 7.2 that a statistic w is a function of the n sample measurements, $y_1, y_2, \ldots, y_n$, and we have shown in Optional Section 7.3 how we can use probability theory and mathematics to find its sampling distribution. However, the mathematical problem of finding $f(w)$ is often very difficult to solve. When such a situation occurs, we may be able to find an approximation to $f(w)$ by computer simulation.

To illustrate the procedure, we will approximate the sampling distribution for the sum $w = y_1 + y_2$ of a sample of $n = 2$ observations from a uniform distribution over the interval $0 \le y \le 1$. Recall that we found an exact expression for this sampling distribution in Example 7.4. Thus, we will be able to compare our simulated sampling distribution with the exact form of the sampling distribution shown in Figure 7.5b.

To begin the simulation procedure, we used the computer to generate 10,000 pairs of random numbers, with each pair representing a sample (y_1, y_2) from the uniform distribution over the interval $0 \le y \le 1$. We then programmed the computer to calculate the sum $w = y_1 + y_2$ for each of the 10,000 pairs. A computer-generated relative frequency histogram for the 10,000 values of w is shown in Figure 7.7. By comparing Figures 7.5b and 7.7, you can see that the simulated sampling distribution provides a good approximation to the true probability distribution of the sum of a sample of $n = 2$ observations from a uniform distribution.

EXAMPLE 7.6

Simulate the sampling distribution of the sample mean

$$\bar{y} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

for a sample of $n = 5$ observations drawn from the uniform probability distribution shown in Figure 7.8 on page 304. Note that the uniform distribution has mean $\mu = .5$. Repeat the procedure for $n = 15, 25, 50,$ and 100. Interpret the results.

Solution

We first obtained 1,000 computer-generated random samples of size $n = 5$ from the uniform probability distribution, over the interval (0, 1), and programmed the computer (using SAS) to compute the mean

$$\bar{y} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

for each sample. The horizontal relative frequency histogram for the 1,000 values of $\bar{y}$ obtained from the uniform distribution is shown in Figure 7.9a on page 304. Note its shape for this small value of n.

The relative frequency histograms of $\bar{y}$ based on samples of size $n = 15$, 25, 50, and 100, also simulated by computer, are shown in Figures 7.9b–e, respectively. Note that the values of $\bar{y}$ tend to cluster about the mean of the uniform distribution, $\mu = .5$. Furthermore, as n increases, there is less variation in the sampling distribution. You can also see from the figures that as the sample size increases, the shape of the sampling distribution of $\bar{y}$ tends toward the shape of the normal distribution (symmetric and mound-shaped).

FIGURE 7.7 ▶
Simulated sampling distribution for the sum of two observations from a uniform (0, 1) distribution

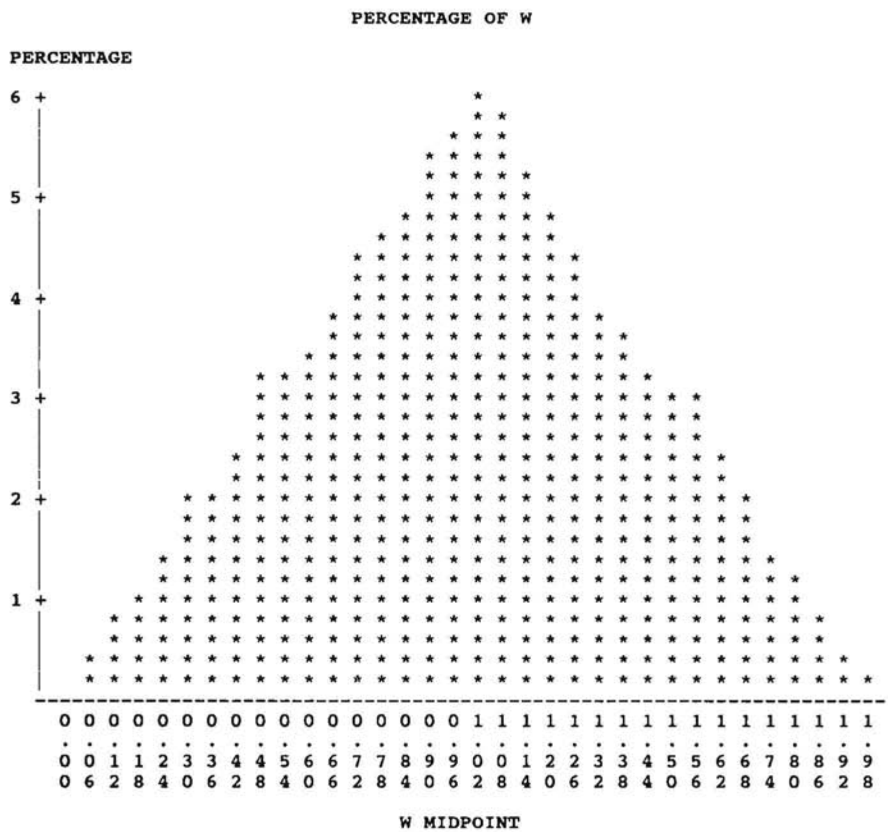
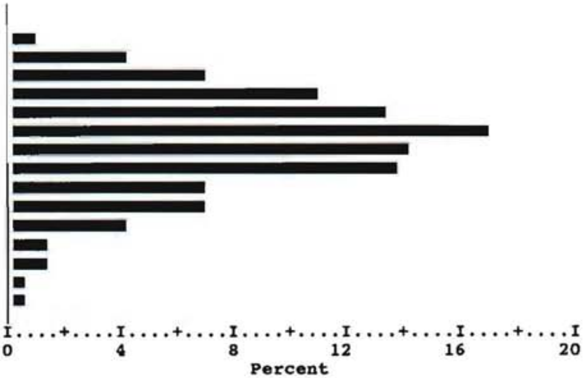


FIGURE 7.12 ▶
Continued

e. $n = 100$

XBAR

Count	Midpoint
1	.74
8	.78
38	.82
69	.86
108	.90
130	.94
168	.98
141	1.02
136	1.06
66	1.10
68	1.14
39	1.18
11	1.22
11	1.26
3	1.30
2	1.34
1	1.38



In Section 7.5, we generalize the results of Examples 7.6 and 7.7 in the form of a theorem.

EXERCISE

OPTIONAL EXERCISE

7.15 Use the computer to simulate the sampling distribution of s^2, the variance of a sample of $n = 100$ observations from a

a. Uniform distribution on the interval (0, 1)
b. Normal distribution, with mean 0 and variance 1
c. Exponential distribution with mean 1

7.5 The Sampling Distributions of Means and Sums

The simulation of the sampling distribution of the sample mean based on independent random samples from uniform, normal, and exponential distributions in Examples 7.6 and 7.7 illustrates the ideas embodied in one of the most important theorems in statistics. The following version of the theorem applies to the sampling distribution of the sample mean, $\bar{y}$.

Theorem 7.2: The Central Limit Theorem

If a random sample of n observations, $y_1, y_2, \ldots, y_n$, is drawn from a population with finite mean μ and variance σ^2, then, when n is sufficiently large, the sampling distribution of the sample mean $\bar{y}$ can be approximated by a normal density function.

The sampling distribution of $\bar{y}$, in addition to being approximately normal for large n, has other known characteristics, which are given in Definition 7.4.

Definition 7.4

Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from a population with finite mean μ and finite standard deviation σ. Then, the **mean and standard deviation of the sampling distribution** of $\bar{y}$, denoted $\mu_{\bar{y}}$ and $\sigma_{\bar{y}}$, respectively, are:

$$\mu_{\bar{y}} = \mu \qquad \sigma_{\bar{y}} = \sigma/\sqrt{n}$$

The significance of the central limit theorem and Definition 7.4 is that we can use the normal distribution to approximate the sampling distribution of the sample mean $\bar{y}$ as long as the population possesses a finite mean and variance, and the number n of measurements in the sample is sufficiently large. How large the sample size must be will depend on the nature of the sampled population. You can see from our simulated experiments in Examples 7.6 and 7.7 that the sampling distribution of $\bar{y}$ tends to become very nearly normal for sample sizes as small as $n = 25$ for the uniform, normal, and exponential population distributions. When the population distribution is symmetric about its mean, the sampling distribution of $\bar{y}$ will be mound-shaped and nearly normal for sample sizes as small as $n = 15$. In addition, if the sampled population possesses a normal distribution, then the sampling distribution of $\bar{y}$ will be a normal density function, regardless of the sample size. (This may be seen in Figure 7.11.) In fact, it can be shown that *the sampling distribution of* ***any linear function*** *of normally distributed random variables, even those that are correlated and have different means and variances, is a normal distribution*. This important result is presented (without proof) in Theorem 7.3 and illustrated in an example.

Theorem 7.3

Let $a_1, a_2, \ldots, a_n$ be constants and let $y_1, y_2, \ldots, y_n$ be n normally distributed random variables with $E(y_i) = \mu_i$, $V(y_i) = \sigma_i^2$, and $\text{Cov}(y_i, y_j) = \sigma_{ij}$ $(i = 1, 2, \ldots, n)$. Then the sampling distribution of a linear combination of the normal random variables

$$\ell = a_1 y_1 + a_2 y_2 + \cdots + a_n y_n$$

possesses a normal density function with mean and variance*

$$E(\ell) = \mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

and

$$\begin{aligned} V(\ell) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \\ + 2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + \cdots + 2a_1a_n\sigma_{1n} \\ + 2a_2a_3\sigma_{23} + \cdots + 2a_2a_n\sigma_{2n} \\ + \cdots + 2a_{n-1}a_n\sigma_{n-1,n} \end{aligned}$$

EXAMPLE 7.8

Suppose you select independent random samples from two normal populations, n_1 observations from population 1 and n_2 observations from population 2. If the means and variances for populations 1 and 2 are (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, and if $\bar{y}_1$ and $\bar{y}_2$ are the corresponding sample means, find the distribution of the difference $(\bar{y}_1 - \bar{y}_2)$.

Solution

Since $\bar{y}_1$ and $\bar{y}_2$ are both linear functions of normally distributed random variables, they will be normally distributed by Theorem 7.3. The means and variances of the sample means (see Example 6.13) are

$$E(\bar{y}_i) = \mu_i \quad \text{and} \quad V(\bar{y}_i) = \frac{\sigma_i^2}{n_i} \qquad (i = 1, 2)$$

Then, $\ell = \bar{y}_1 - \bar{y}_2$ is a linear function of two normally distributed random variables, $\bar{y}_1$ and $\bar{y}_2$. According to Theorem 7.3, ℓ will be normally distributed with

$$E(\ell) = \mu_\ell = E(\bar{y}_1) - E(\bar{y}_2) = \mu_1 - \mu_2$$
$$V(\ell) = \sigma_\ell^2 = (1)^2V(\bar{y}_1) + (-1)^2V(\bar{y}_2) + 2(1)(-1)\text{Cov}(\bar{y}_1, \bar{y}_2)$$

But, since the samples were independently selected, $\bar{y}_1$ and $\bar{y}_2$ are independent and $\text{Cov}(\bar{y}_1, \bar{y}_2) = 0$. Therefore,

$$V(\ell) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

*The formulas for the mean and variance of a linear function of any random variables, $y_1, y_2, \ldots, y_n$, were given in Theorem 6.7.

We have shown that $(\bar{y}_1 - \bar{y}_2)$ is a normally distributed random variable with mean $(\mu_1 - \mu_2)$ and variance $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$.

Typical applications of the central limit theorem, however, involve samples selected from nonnormal or unknown populations, as illustrated in Examples 7.9 and 7.10.

EXAMPLE 7.9

Engineers responsible for the design and maintenance of aircraft pavements traditionally use pavement-quality concrete. A study was conducted at Luton Airport (United Kingdom) to assess the suitability of concrete blocks as a surface for aircraft pavements (*Proceedings of the Institute of Civil Engineers*, Apr. 1986). The original pavement-quality concrete of the western end of the runway was overlaid with 80-mm-thick concrete blocks. A series of plate-bearing tests was carried out to determine the load classification number (LCN)—a measure of breaking strength—of the surface. Let $\bar{y}$ represent the mean LCN of a sample of 25 concrete block sections on the western end of the runway.

a. Prior to resurfacing, the mean LCN of the original pavement-quality concrete of the western end of the runway was known to be $\mu = 60$, and the standard deviation was $\sigma = 10$. If the mean strength of the new concrete block surface is no different from that of the original surface, describe the sampling distribution of $\bar{y}$.

b. If the mean strength of the new concrete block surface is no different from that of the original surface, find the probability that $\bar{y}$, the sample mean LCN of the 25 concrete block sections, exceeds 65.

c. The plate-bearing tests on the new concrete block surface resulted in $\bar{y} = 73$. Based on this result, what can you infer about the true mean LCN of the new surface?

Solution

a. Although we have no information about the shape of the relative frequency distribution of the breaking strengths (LCNs) for sections of the new surface, we can apply Theorem 7.2 to conclude that the sampling distribution of $\bar{y}$, the mean LCN of the sample, is approximately normally distributed. In addition, if $\mu = 60$ and $\sigma = 10$, the mean, $\mu_{\bar{y}}$, and the standard deviation, $\sigma_{\bar{y}}$, of the sampling distribution are given by

$$\mu_{\bar{y}} = \mu = 60$$

and

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

b. We want to calculate $P(\bar{y} > 65)$. Since $\bar{y}$ has an approximate normal distribution, we have

$$P(\bar{y} > 65) = P\left(\frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}} > \frac{65 - \mu_{\bar{y}}}{\sigma_{\bar{y}}}\right)$$
$$\approx P\left(z > \frac{65 - 60}{2}\right) = P(z > 2.5)$$

where z is a standard normal random variable. Using Table 4 of Appendix II, we obtain

$$P(z > 2.5) = .5 - .4938 = .0062$$

Therefore, $P(\bar{y} > 65) = .0062$.

c. If there is no difference between the true mean strengths of the new and original surfaces (i.e., $\mu = 60$ for both surfaces), the probability that we would obtain a sample mean LCN for concrete block of 65 or greater is only .0062. Observing $\bar{y} = 73$ provides strong evidence that the true mean breaking strength of the new surface exceeds $\mu = 60$. Our reasoning stems from the rare event philosophy of Chapter 3, which states that such a large sample mean ($\bar{y} = 73$) is very unlikely to occur if $\mu = 60$.

EXAMPLE 7.10

Consider a binomial experiment with n Bernoulli trials and probability of success p on each trial. The number y of successes divided by the number n of trials is called the **sample proportion of successes** and is denoted by the symbol $\hat{p} = y/n$. Explain why the random variable

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

has approximately a standard normal distribution for large values of n.

Solution

If we denote the outcome of the ith Bernoulli trial as y_i ($i = 1, 2, \ldots, n$), where

$$y_i = \begin{cases} 1 & \text{if outcome is a success} \\ 0 & \text{if outcome is a failure} \end{cases}$$

then the number y of successes in n trials is equal to the sum of n independent Bernoulli random variables:

$$\sum_{i=1}^{n} y_i$$

Therefore, $\hat{p} = y/n$ is a sample mean and, according to Theorem 7.2, $\hat{p}$ will be approximately normally distributed when the sample size n is large. To find the expected value and variance of $\hat{p}$, we can view $\hat{p}$ as a linear function of a single random variable y:

$$\hat{p} = \ell = a_1 y_1 = \left(\frac{1}{n}\right)y \qquad \text{where } a_1 = \frac{1}{n} \text{ and } y_1 = y$$

We now apply Theorem 6.7 to obtain $E(\ell)$ and $V(\ell)$:

$$E(\hat{p}) = \frac{1}{n}E(y) = \frac{1}{n}(np) = p$$

$$V(\hat{p}) = \left(\frac{1}{n}\right)^2 V(y) = \frac{1}{n^2}(npq) = \frac{pq}{n}$$

Therefore,

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

is equal to the deviation between a normally distributed random variable $\hat{p}$ and its mean p, expressed in units of its standard deviation, $\sqrt{pq/n}$. This satisfies the definition of a standard normal random variable given in Section 5.5.

The central limit theorem also applies to the sum of a sample of n measurements subject to the conditions stated in Theorem 7.2. The only difference is that the approximating normal distribution will have mean $n\mu$ and variance $n\sigma^2$.

The Sampling Distribution of a Sum of Random Variables

If a random sample of n observations, $y_1, y_2, \ldots, y_n$, is drawn from a population with finite mean μ and variance σ^2, then, when n is sufficiently large, the sampling distribution of the sum

$$\sum_{i=1}^{n} y_i$$

can be approximated by a normal density function with mean $\mu_{\Sigma y_i} = n\mu$ and $\sigma^2_{\Sigma y_i} = n\sigma^2$.

In Section 7.6, we apply the central limit theorem for sums to show that the normal density function can be used to approximate the binomial probability distribution when the number n of trials is large.

EXERCISES

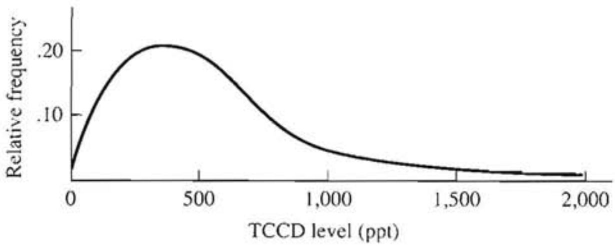
7.16 Let $\bar{y}_{25}$ represent the mean of a random sample of size $n = 25$ from a probability distribution with unknown density $f(y)$, mean $\mu = 17$, and standard deviation $\sigma = 10$. Similarly, let $\bar{y}_{100}$ represent the mean of a random sample of size $n = 100$ selected from the same probability distribution.

a. Describe the sampling distributions of $\bar{y}_{25}$ and $\bar{y}_{100}$.

b. Which of the probabilities, $P(15 < \bar{y}_{25} < 19)$ or $P(15 < \bar{y}_{100} < 19)$, would you expect to be larger?

c. Calculate approximations to the two probabilities of part **b**.

7.17 The National Institute for Occupational Safety and Health (NIOSH) recently completed a study to evaluate the level of exposure of workers to the chemical dioxin, 2,3,7,8-TCDD. The distribution of TCDD levels in parts per trillion (ppt) of production workers at a Newark, New Jersey, chemical plant had a mean of 293 ppt and a standard deviation of 847 ppt (*Chemosphere*, Vol. 20, 1990). A graph of the distribution is shown here.



In a random sample of $n = 50$ workers selected at the New Jersey plant, let $\bar{y}$ represent the sample mean TCDD level.

a. Find the mean and standard deviation of the sampling distribution of $\bar{y}$.

b. Draw a sketch of the sampling distribution of $\bar{y}$. Locate the mean on the graph.

c. Find the probability that $\bar{y}$ exceeds 550 ppt.

7.18 Studies by neuroscientists at the Massachusetts Institute of Technology (MIT) reveal that melatonin, which is secreted by the pineal gland in the brain, functions naturally as a sleep-inducing hormone (*Tampa Tribune*, Mar. 1, 1994). Male volunteers were given various doses of melatonin or placebos and then placed in a dark room at midday and told to close their eyes and fall asleep on demand. Of interest to the MIT researchers is the time y (in minutes) required for each volunteer to fall asleep. With the placebo (i.e., no hormone), the researchers found that the mean time to fall asleep was 15 minutes. Assume that with the placebo treatment $\mu = 15$ and $\sigma = 5$.

a. Consider a random sample of $n = 20$ men who are given the sleep-inducing hormone, melatonin. Let $\bar{y}$ represent the mean time to fall asleep for this sample. If the hormone is *not* effective in inducing sleep, describe the sampling distribution of $\bar{y}$.

b. Refer to part **a**. Find $P(\bar{y} \leq 6)$.

c. In the actual study, the mean time to fall asleep for the 20 volunteers was $\bar{y} = 5$. Use this result to make an inference about the true value of μ for those taking the melatonin.

7.19 *Cost estimation* is the term used to describe the process by which engineers estimate the cost of work contracts (e.g., road construction, building construction) that are to be awarded to the lowest bidder. The engineers' estimate is the baseline against which the low (winning) bid is compared. A recent study investigated the factors that affect the accuracy of engineers' estimates (*Cost Engineering*, Oct. 1988), where accuracy is measured as the percentage difference between the low bid and the engineers' estimate. One of the most important factors is number of bidders—the more bidders on the contract, the more likely the engineers are to overestimate the cost. For building contracts with five bidders, the mean percentage error was -7.02 and the standard deviation was 24.66. Consider a sample of 50 building contracts, each with 5 bidders.

a. Describe the sampling distribution of $\bar{y}$, the mean percentage difference between the low bid and the engineers' estimate, for the 50 contracts.

b. Find $P(\bar{y} < 0)$. (This is the probability of an overestimate.)

c. Suppose you observe $\bar{y} = -17.83$ for a sample of 50 building contracts. Based on the information given, are all these contracts likely to have five bidders? Explain.

7.20 Many species of terrestrial frogs that hibernate at or near the ground surface can survive prolonged exposure to low winter temperatures. In freezing conditions, the frog's body temperature, called its *supercooling temperature*, remains relatively higher because of an accumulation of glycerol in its body fluids. Studies have shown that the supercooling temperature of terrestrial frogs frozen at $-6°C$ has a relative frequency distribution with a mean of $-2.18°C$ and a standard deviation of $.32°C$ (*Science*, May 1983). Consider the mean supercooling temperature, $\bar{y}$, of a random sample of $n = 42$ terrestrial frogs frozen at $-6°C$.

a. Find the probability that $\bar{y}$ exceeds $-2.05°C$.

b. Find the probability that $\bar{y}$ falls between $-2.20°C$ and $-2.10°C$.

7.21 General trace organic monitoring describes the process in which water engineers analyze water samples for various types of organic material (e.g., contaminants). One such contaminant, commonly found in treated surface water, is the pesticide trihalomethane (THM). General trace organic monitoring at the Bedford (England) water treatment works revealed a mean THM level of 51 μg/l and a standard deviation of 14 μg/l (*Journal of the Institution of Water Engineers and Scientists*, Feb. 1986). Assume that these figures represent the population mean μ and standard deviation σ, respectively. Suppose we collect 45 water samples (called water "profiles") at the Bedford plant and measure the THM level in each.

a. Describe the sampling distribution of $\bar{y}$, the mean THM level of the 45 water profiles.

b. Find the probability that $\bar{y}$ exceeds 52 μg/l.

c. Find the probability that $\bar{y}$ falls between 49.5 and 50.5 μg/l.

7.22 The U.S. Army Engineering and Housing Support Center recently sponsored a study of the reliability, availability, and maintainability (RAM) characteristics of small diesel and gas-powered systems at commercial and military facilities (*IEEE Transactions on Industry Applications*, July/Aug. 1990). The study revealed that the time, y, to perform corrective maintenance on continuous diesel auxiliary systems has an approximate exponential distribution with an estimated mean of 1,700 hours.

a. Assuming $\mu = 1{,}700$, find the probability that the mean time to perform corrective maintenance for a sample of 70 continuous diesel auxiliary systems exceeds 2,500 hours.

b. If you observe $\bar{y} > 2{,}500$, what inference would you make about the value of μ?

7.23 An article in *Industrial Engineering* (Aug. 1990) discussed the importance of modeling machine downtime correctly in simulation studies. As an illustration, the researcher considered a single-machine-tool system with repair times (in minutes) that can be modeled by a gamma distribution with parameters $\alpha = 1$ and $\beta = 60$. Of interest is the mean repair time, $\bar{y}$, of a sample of 100 machine breakdowns.

a. Find $E(\bar{y})$ and $\text{Var}(\bar{y})$.

b. What probability distribution provides the best model of the sampling distribution of $\bar{y}$? Why?

c. Calculate the probability that the mean repair time, $\bar{y}$, is no longer than 30 minutes.

7.24 A large freight elevator can transport a maximum of 10,000 pounds (5 tons). Suppose a load of cargo containing 45 boxes must be transported via the elevator. Experience has shown that the weight y of a box of this type of cargo follows a probability distribution with mean $\mu = 200$ pounds and standard deviation $\sigma = 55$ pounds. What is the probability that all 45 boxes can be loaded onto the freight elevator and transported simultaneously? [*Hint:* Find $P(\Sigma_{i=1}^{45} y_i \leq 10{,}000)$.]

OPTIONAL EXERCISES

7.25 Let $\hat{p}_1$ be the sample proportion of successes in a binomial experiment with n_1 trials and let $\hat{p}_2$ be the sample proportion of successes in a binomial experiment with n_2 trials, conducted independently of the first. Let p_1 and p_2 be the corresponding population parameters. Show that

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}}$$

has approximately a standard normal distribution for large values of n_1 and n_2.

7.26 If y has a χ^2 distribution with n degrees of freedom (see Section 5.7), then y could be represented by $y = \Sigma_{i=1}^{n} x_i$, where the x_i's are independent χ^2 distributions, each with 1 degree of freedom.

a. Show that $z = (y - n)/\sqrt{2n}$ has approximately a standard normal distribution for large values of n.

b. If y has a χ^2 distribution with 30 degrees of freedom, find the approximate probability that y falls within 2 standard deviations of its mean, i.e., find $P(\mu - 2\sigma < y < \mu + 2\sigma)$.

7.6 Normal Approximation to the Binomial Distribution

Consider the binomial random variable y with parameters n and p. Recall that y has mean $\mu = np$ and variance $\sigma^2 = npq$. We showed in Example 7.10 that the number y of successes in n trials can be regarded as a sum consisting of n values of 0 and 1, with each 0 and 1 representing the outcome (failure or success, respectively) of a particular trial, i.e.,

$$y = \sum_{i=1}^{n} y_i \qquad \text{where } y_i = \begin{cases} 1 & \text{if success} \\ 0 & \text{if failure} \end{cases}$$

Then, according to the central limit theorem for sums, the binomial probability distribution $p(y)$ should become more nearly normal as n becomes larger. The normal approximation to a binomial probability distribution is reasonably good even for small samples—say, n as small as 10—when $p = .5$ and the distribution of y is therefore symmetric about its mean $\mu = np$. When p is near 0 (or 1), the binomial probability distribution will tend to be skewed to the right (or left), but this skewness will disappear as n becomes large. In general, the approximation will be good when n is large enough so that $\mu - 2\sigma = np - 2\sqrt{npq}$ and $\mu + 2\sigma = np + 2\sqrt{npq}$ both lie between 0 and n. It can be shown (proof omitted) that for both $\mu - 2\sigma$ and $\mu + 2\sigma$ to fall between 0 and n, both np and nq must be greater than or equal to 4.

Condition Required to Apply a Normal Approximation to a Binomial Probability Distribution

The approximation will be good if both $\mu - 2\sigma = np - 2\sqrt{npq}$ and $\mu + 2\sigma = np + 2\sqrt{npq}$ lie between 0 and n. This condition will be satisfied if both $np \geq 4$ and $nq \geq 4$.

EXAMPLE 7.11

Let y be a binomial probability distribution with $n = 10$ and $p = .5$.

a. Graph $p(y)$ and superimpose on the graph a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$.

b. Use Table 1 of Appendix II to find $P(y \leq 4)$.

c. Use the normal approximation to the binomial probability distribution to find an approximation to $P(y \leq 4)$.

Solution

a. The graphs of $p(y)$ and a normal distribution with

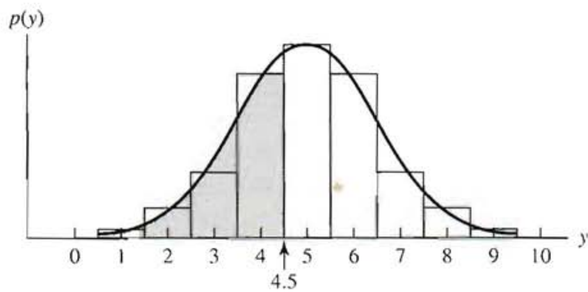
$$\mu = np = (10)(.5) = 5$$

and

$$\sigma = \sqrt{npq} = \sqrt{(10)(.5)(.5)} = 1.58$$

are shown in Figure 7.13. Note that both $np = 5$ and $nq = 5$ both exceed 4. Thus, the normal density function with $\mu = 5$ and $\sigma = 1.58$ provides a good approximation to $p(y)$.

FIGURE 7.13 ► A binomial probability distribution ($n = 10$, $p = .5$) and the approximating normal distribution ($\mu = np = 5$ and $\sigma = \sqrt{npq} = 1.58$)



b. From Table 1 of Appendix II, we obtain

$$\sum_{y=0}^{4} p(y) = .377$$

c. By examining Figure 7.13, you can see that $P(y \leq 4)$ is the area under the normal curve to the left of $y = 4.5$. Note that the area to the left of $y = 4$ would *not* be appropriate because it would omit half the probability rectangle corresponding to

$y = 4$. We need to add .5 to 4 before calculating the probability to correct for the fact that we are using a continuous probability distribution to approximate a discrete probability distribution. The value .5 is called the **continuity correction factor** for the normal approximation to the binomial probability (see the box). The z value corresponding to the corrected value $y = 4.5$ is

$$z = \frac{y - \mu}{\sigma} = \frac{4.5 - 5}{1.58} = \frac{-.5}{1.58} = -.32$$

The area between $z = 0$ and $z = .32$, given in Table 4 of Appendix II, is A = .1255. Therefore,

$$P(y \leq 4) \approx .5 - \text{A} = .5 - .1255 = .3745$$

Thus, the normal approximation to $P(y \leq 4) = .377$ is quite good, although n is as small as 10. The sample size would have to be larger to apply the approximation if p were not equal to .5.

Continuity Correction for the Normal Approximation to a Binomial Probability

Let y be a binomial random variable with parameters n and p, and let z be a standard random variable. Then

$$P(y \leq a) \approx P\left(z < \frac{(a + .5) - np}{\sqrt{npq}}\right)$$

$$P(y \geq a) \approx P\left(z > \frac{(a - .5) - np}{\sqrt{npq}}\right)$$

$$P(a \leq y \leq b) \approx P\left(\frac{(a - .5) - np}{\sqrt{npq}} < z < \frac{(b + .5) - np}{\sqrt{npq}}\right)$$

EXERCISES

7.27 Let y be a binomial random variable with $n = 15$ and $p = .3$.

a. Use Table 1 of Appendix II to find $P(y \leq 8)$.

b. Use the normal approximation to the binomial probability distribution to find an approximation to $P(y \leq 8)$. Compare to your answer in part a.

7.28 *Consumer Reports* (Feb. 1992) found widespread contamination of seafood in New York and Chicago supermarkets. For example, 40% of the swordfish pieces available for sale have a level of mercury above the Food and Drug Administration (FDA) limit. Consider a random sample of 20 swordfish pieces from New York and Chicago supermarkets.

a. Use the normal approximation to the binomial to calculate the probability that less than 2 of the 20 swordfish pieces have mercury levels exceeding the FDA limit.

b. Use the normal approximation to the binomial to calculate the probability that more than half of the 20 swordfish pieces have mercury levels exceeding the FDA limit.

c. Use the binomial tables to calculate the exact probabilities in parts **a** and **b**. Does the normal distribution provide a good approximation to the binomial distribution?

7.29 The merging process from an acceleration lane to the through lane of a freeway constitutes an important aspect of traffic operation at interchanges. A study of parallel interchange ramps in Israel revealed that many drivers do not use the entire length of parallel lanes for acceleration, but seek as soon as possible an appropriate gap in the major stream of traffic for merging (*Transportation Engineering*, Nov. 1985). At one site (Yavneh), 54% of the drivers use less than half the lane length available before merging. Suppose we plan to monitor the merging patterns of a random sample of 330 drivers at the Yavneh site.

a. What is the approximate probability that fewer than 100 of the drivers will use less than half the acceleration lane length before merging?

b. What is the approximate probability that 200 or more of the drivers will use less than half the acceleration lane length before merging?

7.30 *Occupational Outlook Quarterly* (Spring 1993) reported that 1% of all drywall installers employed in the construction industry are women.

a. Approximate the probability that more than 100 of a random sample of 500 drywall installers are women.

b. Approximate the probability that five or fewer of a random sample of 500 drywall installers are women.

7.31 One of the keys to developing successful information systems is to implement structured design and programming techniques. Computer-aided software engineering (CASE) technology provides several automated tools (e.g., data flow diagrams) that can facilitate structured techniques. The *Journal of Systems Management* (July 1989) reported that 60% of information systems (IS) professionals make extensive use of data flow diagrams in their work. In a sample of 150 IS professionals, what is the approximate probability that at least half make extensive use of data flow diagrams?

7.32 Quality control is a problem with items that are mass produced. The production process must be monitored to ensure that the rate of defective items is kept at an acceptably low level. One method of dealing with this problem is **lot acceptance sampling**, in which a random sample of items produced is selected and each item in the sample is carefully tested. The entire lot of items is then accepted or rejected, based on the number of defectives observed in the sample. Suppose a manufacturer of pocket calculators randomly chooses 200 stamped circuits from a day's production and determines y, the number of defective circuits in the sample. If a sample defective rate of 6% or less is considered acceptable and, unknown to the manufacturer, 8% of the entire day's production of circuits is defective, find the approximate probability that the lot of stamped circuits will be rejected.

7.33 How well does a college engineering degree prepare you for the workplace? A 2-year nationwide survey of engineers and engineering managers in "specific high-demand" industries revealed that only 34% believe that their companies make good use of their learned skills (*Chemical Engineering*, Feb. 3, 1986). In a random sample of 50 engineers and engineering managers, consider the number y who believe that their employer makes good use of their college engineering background. Find the approximate probability that:

a. $y \leq 10$ b. $y \geq 25$ c. $20 \leq y \leq 30$

7.7 Sampling Distributions Related to the Normal Distribution

In this section, we present the sampling distributions of several well-known statistics that are based on random samples of observations from a normal population. These statistics are the χ^2, t, and F statistics. In Chapter 8, we show how to use these statistics to estimate the values of certain population parameters. The following results are stated without proof. Proofs using the methodology of Chapter 6 can be found in the references at the end of this chapter.

Theorem 7.4

If a random sample of n observations, $y_1, y_2, \ldots, y_n$, is selected from a normal distribution with mean μ and variance σ^2, then the sampling distribution of

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square density function (see Section 5.7) with $\nu = (n-1)$ degrees of freedom.

Theorem 7.5

If χ_1^2 and χ_2^2 are independent chi-square random variables with ν_1 and ν_2 degrees of freedom, respectively, then the sum $(\chi_1^2 + \chi_2^2)$ has a chi-square distribution with $(\nu_1 + \nu_2)$ degrees of freedom.

Definition 7.5

Let z be a standard normal random variable and χ^2 be a chi-square random variable with ν degrees of freedom. If z and χ^2 are independent, then

$$t = \frac{z}{\sqrt{\chi^2/\nu}}$$

is said to possess a **Student's *t* distribution** (or, simply, ***t* distribution**) with ν degrees of freedom.

Definition 7.6

Let χ_1^2 and χ_2^2 be chi-square random variables with ν_1 and ν_2 degrees of freedom, respectively. If χ_1^2 and χ_2^2 are independent, then

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$$

is said to have an ***F* distribution** with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom.

Note: The sampling distributions for the t and F statistics can also be derived using the methods of Optional Section 7.3. Both sampling distributions are related to the density function for a beta-type random variable (see Section 5.9). It can be shown (proof omitted) that a t distribution with ν degrees of freedom is actually a special case of an F distribution with $\nu_1 = 1$ and $\nu_2 = \nu$ degrees of freedom. Neither of the cumulative distribution functions can be obtained in closed form. Consequently, we dispense with the equations of the density functions and present useful values of the statistics and corresponding areas in tabular form in Appendix II.

The following examples illustrate how these statistics can be used to make probability statements about population parameters.

EXAMPLE 7.12

Consider a cannery that produces 8-ounce cans of processed corn. Quality control engineers have determined that the process is operating properly when the true variation σ^2 of the fill amount per can is less than .0025. A random sample of $n = 10$ cans is selected from a day's production, and the fill amount (in ounces) recorded for each. Of interest is the sample variance, s^2. If, in fact, $\sigma^2 = .001$, find the probability that s^2 exceeds .0025. Assume that the fill amounts are normally distributed.

Solution

We want to calculate $P(s^2 > .0025)$. Assume the sample of 10 fill amounts is selected from a normal distribution. Theorem 7.4 states that the statistic

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square probability distribution with $\nu = (n - 1)$ degrees of freedom. Consequently, the probability we seek can be written

$$P(s^2 > .0025) = P\left[\frac{(n-1)s^2}{\sigma^2} > \frac{(n-1)(.0025)}{\sigma^2}\right]$$
$$= P\left[\chi^2 > \frac{(n-1)(.0025)}{\sigma^2}\right]$$

Substituting $n = 10$ and $\sigma^2 = .001$, we have

$$P(s^2 > .0025) = P\left(\chi^2 > \frac{9(.0025)}{.001}\right) = P(\chi^2 > 22.5)$$

Upper-tail areas of the chi-square distribution have been tabulated and are given in Table 8 of Appendix II, a portion of which is reproduced in Table 7.2. The table gives the values of χ^2, denoted χ^2_a, that locate an area (probability) a in the upper-tail of the distribution, i.e., $P(\chi^2 > \chi^2_a) = a$. In our example, we want to find the probability a such that $\chi^2_a > 22.5$.

Now, for $n = 10$, we have $\nu = n - 1 = 9$ degrees of freedom. Searching Table 7.2 in the row corresponding to $\nu = 9$, we find that $\chi^2_{.01} = 21.666$ and $\chi^2_{.005} = 23.5893$. (These values are shaded in Table 7.2.) Consequently, the probability that we seek falls between $a = .01$ and $a = .005$, i.e.,

$$.005 < P(\chi^2 > 22.5) < .01 \quad \text{(see Figure 7.14)}$$

Thus, the probability that the variance of the sample fill amounts exceeds .0025 is small (between .005 and .01) when the true population variance σ^2 equals .001.

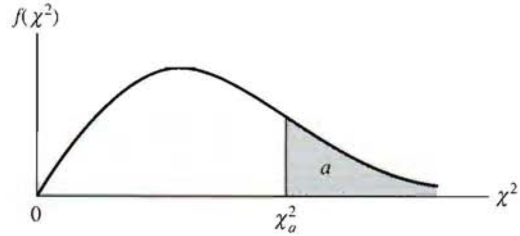


TABLE 7.2 Abbreviated Version of Table 8 of Appendix II: Tabulated Values of χ^2

Degrees of Freedom	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	2.70554	3.84146	5.02389	6.63490	7.87944
2	4.60517	5.99147	7.37776	9.21034	10.5966
3	6.25139	7.81473	9.34840	11.3449	12.8381
4	7.77944	9.48773	11.1433	13.2767	14.8602
5	9.23635	11.0705	12.8325	15.0863	16.7496
6	10.6446	12.5916	14.4494	16.8119	18.5476
7	12.0170	14.0671	16.0128	18.4753	20.2777
8	13.3616	15.5073	17.5346	20.0902	21.9550
9	14.6837	16.9190	19.0228	21.6660	23.5893
10	15.9871	18.3070	20.4831	23.2093	25.1882
11	17.2750	19.6751	21.9200	24.7250	26.7569
12	18.5494	21.0261	23.3367	26.2170	28.2995
13	19.8119	22.3621	24.7356	27.6883	29.8194
14	21.0642	23.6848	26.1190	29.1413	31.3193
15	22.3072	24.9958	27.4884	30.5779	32.8013
16	23.5418	26.2962	28.8454	31.9999	34.2672
17	24.7690	27.5871	30.1910	33.4087	35.7185
18	25.9894	28.8693	31.5264	34.8053	37.1564
19	27.2036	30.1435	32.8523	36.1908	38.5822

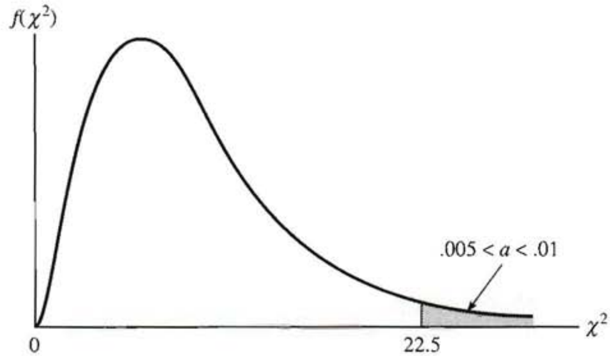


FIGURE 7.14 ▶ Finding $P(\chi^2 > 22.5)$ in Example 7.12

EXAMPLE 7.13

Suppose that $\bar{y}$ and s^2 are the mean and variance of a random sample of n observations from a normally distributed population with mean μ and variance σ^2. It can be shown (proof omitted) that $\bar{y}$ and s^2 are statistically independent when the sampled population has a normal distribution. Use this result to show that

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

possesses a t distribution with $\nu = (n - 1)$ degrees of freedom.*

Solution

We know from Theorem 7.3 that $\bar{y}$ is normally distributed with mean μ and variance σ^2/n. Therefore,

$$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal random variable. We also know from Theorem 7.4 that

$$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$$

is a χ^2 random variable with $\nu = (n - 1)$ degrees of freedom. Then, using Definition 7.4 and the information that $\bar{y}$ and s^2 are independent, we conclude that

$$t = \frac{z}{\sqrt{\chi^2/\nu}} = \frac{\dfrac{\bar{y} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n - 1)s^2}{\sigma^2}\Big/(n - 1)}} = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

has a Student's t distribution with $\nu = (n - 1)$ degrees of freedom. As we will learn in Chapter 8, the t distribution is useful for making inferences about the population

*The result was first published in 1908 by W. S. Gosset, who wrote under the pen name of Student. Thereafter, this statistic became known as Student's t.

mean μ when the population standard deviation σ is unknown (and must be estimated by s^2).

Theorem 7.4 and Examples 7.12 and 7.13 identify the sampling distributions of two statistics that will play important roles in statistical inference. Others are presented without proof in Tables 7.3a and 7.3b. All are based on random sampling from normally distributed populations. The results contained in Table 7.3 will be needed in Chapter 8.

TABLE 7.3a. Sampling Distributions of Statistics Based on Independent Random Samples of n_1 and n_2 Observations, Respectively, from Normally Distributed Populations with Parameters (μ_1, σ_1^2) and (μ_2, σ_2^2)

Statistic	Sampling Distribution	Additional Assumptions	Basis of Derivation of Sampling Distribution
$\chi^2 = \dfrac{(n_1 + n_2 - 2)s_p^2}{\sigma^2}$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	Chi-square with $\nu = (n_1 + n_2 - 2)$ degrees of freedom	$\sigma_1^2 = \sigma_2^2 = \sigma^2$	Theorems 7.4–7.5
$t = \dfrac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	Student's t with $\nu = (n_1 + n_2 - 2)$ degrees of freedom	$\sigma_1^2 = \sigma_2^2 = \sigma^2$	Theorems 7.3–7.4 and Definition 7.4
$F = \left(\dfrac{s_1^2}{s_2^2}\right)\left(\dfrac{\sigma_2^2}{\sigma_1^2}\right)$	F distribution with $\nu_1 = (n_1 - 1)$ numerator degrees of freedom and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom	None	Theorem 7.4 and Definition 7.6

TABLE 7.3b. Sampling Distributions of Statistics Based on a Random Sample from a Single Normally Distributed Population with Mean μ and Variance σ^2

Statistic	Sampling Distribution	Additional Assumptions	Basis of Derivation of Sampling Distribution
$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$	Chi-square with $\nu = (n-1)$ degrees of freedom	None	Methods of Section 7.2
$t = \frac{\bar{y}-\mu}{s/\sqrt{n}}$	Student's t with $\nu = (n-1)$ degrees of freedom	None	Theorems 7.3–7.4 and Definition 7.4

EXERCISES

7.34 Let $y_1, y_2, \ldots, y_n$, be a random sample of n observations from a normal distribution with mean μ and variance σ^2. Let s^2 be the variance of the sample. Use Table 8 of Appendix II to estimate the following probabilities:

a. $P(s^2 > 8)$ when $n = 10$, $\sigma^2 = 5$

b. $P(s^2 > 1.11)$ when $n = 5$, $\sigma^2 = .3$

c. $P(s^2 > 199)$ when $n = 22$, $\sigma^2 = 107$

7.35 *IEEE Transactions* (June 1990) presented a hybrid algorithm for solving polynomial 0–1 mathematical programming problems. The solution time (in seconds) for a randomly selected problem solved using the hybrid algorithm has a normal probability distribution with mean $\mu = .8$ second and $\sigma = 1.5$ seconds. Consider a random sample of $n = 30$ problems solved with the hybrid algorithm.

a. Describe the sampling distribution of s^2, the variance of the solution times for the 30 problems.

b. Find the approximate probability that s^2 will exceed 3.30.

OPTIONAL EXERCISES

7.36 Let $y_1, y_2, \ldots, y_n$ be a random sample of n_1 observations from a normal distribution with mean μ_1 and variance σ_1^2. Let $x_1, x_2, \ldots, x_{n_2}$ be a random sample of n_2 observations from a normal distribution with mean μ_2 and variance σ_2^2. Assuming the samples were independently selected, show that

$$F = \left(\frac{s_1^2}{s_2^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$$

has an F distribution with $\nu_1 = (n_1 - 1)$ numerator degrees of freedom and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom.

7.37 Let s_1^2 and s_2^2 be the variances of independent random samples of sizes n_1 and n_2 selected from normally distributed populations with parameters (μ_1, σ^2) and (μ_2, σ^2), respectively. Thus, the populations have different means, but a common variance σ^2. To estimate the common variance, we can combine information from both samples and use the **pooled estimator**

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Use Theorems 7.4 and 7.5 to show that $(n_1 + n_2 - 2)s^2/\sigma^2$ has a chi-square distribution with $\nu = (n_1 + n_2 - 2)$ degrees of freedom.

7.38 Let $\bar{y}_1$ and $\bar{y}_2$ be the means of independent random samples of sizes n_1 and n_2 selected from normally distributed populations with parameters (μ_1, σ^2) and (μ_2, σ^2), respectively. If

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

show that

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

has a Student's t distribution with $\nu = (n_1 + n_2 - 2)$ degrees of freedom.

7.39 The continuous random variable y is said to have a lognormal distribution with parameters μ and σ if its probability density function, $f(y)$, satisfies

$$f(y) = \frac{1}{\sigma y\sqrt{2\pi}} \exp\left\{-\frac{(\ln y - \mu)^2}{2\sigma^2}\right\} \qquad (y > 0)$$

Show that $x = \ln(y)$ has a normal distribution with mean μ and variance σ^2.

7.8 Summary

In the following chapters, we will use sample statistics to make inferences about population parameters; the properties of these statistics will be determined by their probability distributions. The probability distribution of a statistic is called its **sampling distribution**.

A **simulation procedure** may be used to approximate the sampling distribution for a statistic. Random samples of a fixed size are drawn from a known population of data. The value of some statistic—say, the sample mean $\bar{y}$—is computed for each sample. The relative frequency distribution of the values of the statistic, generated by repeated sampling, approximates the probability distribution of the statistic.

Evidence of the major role that the normal distribution plays in statistical inference is given by the **central limit theorem**, Theorem 7.3, and the related $\boldsymbol{\chi^2}$, $\boldsymbol{F}$, and $\boldsymbol{t}$ **distributions**. The central limit theorem explains why many statistics, especially those based on large samples, possess sampling distributions that can be approximated by a normal density function. Theorem 7.3, which states that linear functions of normally distributed random variables will be normally distributed, provides further explanation for the common occurrence of normally distributed sampling distributions. The χ^2, t, and F statistics are approximated when sampling from normally distributed populations. You will encounter them frequently in the statistical methodology to be developed in the following chapters.

SUPPLEMENTARY EXERCISES

7.40 Consider the density function

$$f(y) = \begin{cases} 3y^2 & \text{if } 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the density function of w, where:
a. $w = \sqrt{y}$ **b.** $w = 3 - y$ **c.** $w = -\ln(y)$

7.41 A supplier of home heating oil has a 250-gallon tank that is filled at the beginning of each week. Since the weekly demand for the oil increases steadily up to 100 gallons and then levels off between 100 and 250 gallons, the probability distribution of the weekly demand y (in hundreds of gallons) can be represented by

$$f(y) = \begin{cases} \frac{y}{2} & \text{if } 0 \le y \le 1 \\ \frac{1}{2} & \text{if } 1 \le y \le 2.5 \\ 0 & \text{elsewhere} \end{cases}$$

If the supplier's profit is given by $w = 10y - 2$, find the probability density function of w.

7.42 Dioxin, often described as the most toxic chemical known, is created as a by-product in the manufacture of herbicides such as Agent Orange. Scientists have found that .000005 gram (five-millionths of a gram) of dioxin—a dot barely visible to the human eye—is a lethal dose for experimental guinea pigs in more than half the animals tested, making dioxin 2,000 times more toxic than strychnine. Assume that the amount of dioxin required to kill a guinea pig has a relative frequency distribution with mean $\mu = .000005$ gram and standard deviation $\sigma = .000002$ gram. Consider an experiment in which the amount of dioxin required to kill each of $n = 50$ guinea pigs is measured, and the sample mean $\bar{y}$ is computed.
a. Calculate $\mu_{\bar{y}}$ and $\sigma_{\bar{y}}$.
b. Find the probability that the mean amount of dioxin required to kill the 50 guinea pigs is larger than .0000053 gram.

7.43 The determination of the percent canopy closure of a forest is essential for wildlife habitat assessment, watershed runoff estimation, erosion control, and other forest management activities. One way in which geoscientists estimate percent forest canopy closure is through the use of a satellite sensor called the Landsat Thematic Mapper. A study of the percent canopy closure in the San Juan National Forest (Colorado) was conducted by examining Thematic Mapper Simulator (TMS) data collected by aircraft at various forest sites (*IEEE Transactions on Geoscience and Remote Sensing*, Jan. 1986). The mean and standard deviation of the readings obtained from TMS Channel 5 were found to be 121.74 and 27.52, respectively.

a. Let $\bar{y}$ be the mean TMS reading for a sample of 32 forest sites. Assuming the figures given are population values, describe the sampling distribution of $\bar{y}$.
b. Use the sampling distribution of part **a** to find the probability that $\bar{y}$ falls between 118 and 130.

7.44 Refer to Exercise 7.43. Let s^2 be the variance of the TMS readings for the 32 sampled forest sites. Assuming the sample is from a normal population, estimate the probability that s^2 exceeds 1,311.

7.45 Use Theorem 7.1 to draw a random sample of $n = 5$ observations from a population with probability density function given by

$$f(y) = \begin{cases} 2ye^{-y^2} & \text{if } 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

7.46 Use Theorem 7.1 to draw a random sample of $n = 5$ observations from a population with probability density function given by

$$f(y) = \begin{cases} 2(y-1) & \text{if } 1 \le y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

7.47 This year a large architectural and engineering consulting firm began a program of compensating its management personnel for sick days not used. The firm decided to pay each manager a bonus for every unused sick day. In past years, the number y of sick days used per manager per year had a probability distribution with mean $\mu = 9.2$ and variance $\sigma^2 = 3.24$. To determine whether the compensation program has effectively reduced the mean number of sick days used, the firm randomly sampled $n = 80$ managers and recorded y, the number of sick days used by each at year's end.
a. Assuming the compensation program was *not* effective in reducing the average number of sick days used, find the probability that $\bar{y}$, the mean number of sick days used by the sample of 80 managers, is less than 8.80 days, i.e., find $P(\bar{y} < 8.80)$.
b. If you observe $\bar{y} < 8.80$, what inference would you make about the effectiveness of the compensation program?

7.48 To determine whether a metal lathe that produces machine bearings is properly adjusted, a random sample of 36 bearings is collected and the diameter of each is measured. Assume that the standard deviation of the diameter of the machine bearings measured over a long period of time is .001 inch.
a. What is the probability that the mean diameter $\bar{y}$ of the sample of 36 bearings will lie within .0001 inch of the population mean diameter of the bearings?
b. Suppose the mean diameter of the bearings produced by the machine is supposed to be .5 inch. The company decides to use the sample mean to decide whether the process is in control—i.e., whether it is producing bearings with a mean diameter of .5 inch. The machine will be considered out of control if the mean of the sample of $n = 36$ diameters is less than .4994 inch or larger than .5006 inch. If the true mean diameter of the bearings produced by the machine is .501 inch, what is the probability that the test will fail to imply that the process is out of control?

7.49 Refer to the problem of transporting neutral particles in a nuclear fusion reactor, described in Exercise 3.25. Recall that particles released into a certain type of evacuated duct collide with the inner duct wall and are either scattered (reflected) with probability .16 or absorbed with probability .84 (*Nuclear Science and Engineering*, May 1986). Suppose 2,000 neutral particles are released into an unknown type of evacuated duct in a nuclear fusion reactor. Of these, 280 are reflected. What is the approximate probability that as few as 280 (i.e., 280 or fewer) of the 2,000 neutral particles would be reflected off the inner duct wall if the reflection probability of the evacuated duct is .16?

7.50 Shear block tests on epoxy-repaired timber indicate that the probability distribution of the bond strengths of parallel grain, mill lumber specimens has a mean of 1,312 pounds per square inch (psi) and a standard deviation of 422 psi (*Journal of Structural Engineering*, Feb. 1986). Suppose a sample of 100 epoxy-repaired timber specimens is randomly selected and the bond strength of each is determined.

a. Describe the sampling distribution of $\bar{y}$, the mean bond strength of the sample of 100 epoxy-repaired timber specimens.

b. Compute $P(\bar{y} \geq 1{,}418)$.

c. If the actual sample mean is computed to be $\bar{y} = 1{,}418$, what would you infer about the shear block test results?

7.51 Refer to Exercise 7.50.

a. Describe the sampling distribution of s^2, the variance of the bond strengths of the 100 sampled epoxy-repaired timber specimens. Assume the sample is from a normal population.

b. Estimate $P(s > 500)$.

OPTIONAL SUPPLEMENTARY EXERCISES

7.52 The waiting time y until delivery of a new component for a data-processing unit is uniformly distributed over the interval from 1 to 5 days. The cost c (in hundreds of dollars) of this delay to the purchaser is given by $c = (2y^2 + 3)$. Find the probability that the cost of delay is at least \$2,000, i.e., compute $P(c \geq 20)$.

7.53 Let y_1 and y_2 be a sample of $n = 2$ observations from a gamma random variable with parameters $\alpha = 1$ and arbitrary β, and corresponding density function

$$f(y_i) = \begin{cases} \dfrac{1}{\beta}e^{-y_i/\beta} & \text{if } y_i > 0 \quad (i = 1, 2) \\ 0 & \text{elsewhere} \end{cases}$$

Show that the sum $w = (y_1 + y_2)$ is also a gamma random variable with parameters $\alpha = 2$ and β. [*Hint:* You may use the result

$$P(w \leq w_0) = P(0 < y_2 \leq w - y_1,\ 0 \leq y_1 < w) = \int_0^w \int_0^{w-y_1} f(y_1, y_2)dy_2\, dy_1$$

Then use the fact that

$$f(y_1, y_2) = f(y_1)f(y_2)$$

since y_1 and y_2 are independent.]

7.54 Let y have an exponential density with mean β. Show that $w = 2y/\beta$ has a χ^2 density with $\nu = 2$ degrees of freedom.

7.55 The lifetime y of an electronic component of a home minicomputer has a *Rayleigh density*, given by

$$f(y) = \begin{cases} \left(\dfrac{2y}{\beta}\right)e^{-y^2/\beta} & \text{if } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probability density function for $w = y^2$, and identify the type of density function. [*Hint:* You may use the result

$$\int \frac{2y}{\beta} e^{-y^2/\beta}\, dy = -e^{-y^2/\beta}$$

in determining the density function for w.]

7.56 Let y_1 and y_2 be a random sample of $n = 2$ observations from a normal distribution with mean μ and variance σ^2.

a. Show that

$$z = \frac{y_1 - y_2}{\sqrt{2}\sigma}$$

has a standard normal distribution.

b. Given the result in part **a**, show that z^2 possesses a χ^2 distribution with 1 degree of freedom. [*Hint:* First show that $s^2 = (y_1 - y_2)^2/2$; then apply Theorem 7.4.]

7.57 Refer to Exercise 7.12. Use the computer to generate a random sample of $n = 100$ observations from a distribution with probability density

$$f(y) = \begin{cases} e^y & \text{if } y < 0 \\ 0 & \text{elsewhere} \end{cases}$$

Repeat the procedure 1,000 times and compute the sample mean $\bar{y}$ for each of the 1,000 samples of size $n = 100$. Then generate (by computer) a relative frequency histogram for the 1,000 sample means. Does your result agree with the theoretical sampling distribution described by the central limit theorem?

COMPUTER LAB: Generating Random Samples

Most statistical computer software packages have built-in algorithms for generating random samples of observations from a variety of probability distributions. The SAS and MINITAB commands for generating random samples of size 50 from the uniform distribution are given in the following programs. Table 7.4 gives the corresponding commands for generating samples from the normal, binomial, Poisson, exponential, and gamma distributions.

TABLE 7.4 Random Number Generators for SAS and MINITAB

Probability Distribution	SAS	MINITAB
Uniform (0, 1)	Y = RANUNI(seed);	UNIFORM.
Uniform (A, B)	Y = A+B*RANUNI(seed);	UNIFORM A B.
Normal (mean = 0, std.dev. = 1)	Y = RANNOR(seed);	NORMAL.
Normal (mean = M, std.dev. = S)	Y = M+S*RANNOR(seed);	NORMAL M S.
Binomial (N, P)	Y = RANBIN(seed,N,P);	BINOMIAL N P.
Exponential (mean = 1)	Y = RANEXP(seed);	EXPONENTIAL.
Exponential (mean = B)	Y = RANEXP(seed)/B;	EXPONENTIAL B.
Gamma (A, 1)	Y = RANGAM(seed,A);	GAMMA A 1.
Gamma (A, B)	Y = B*RANGAM(seed,A);	GAMMA A B.
Chi-square (df = V)	Y = 2*RANGAM(seed,V/2);	CHISQUARE V.
Poisson (mean = L)	Y = RANPOI(seed,L)	POISSON L.
Beta (A, B)	Y1 = RANGAM(seed,A); Y2 = RANGAM(seed,B); Y = Y1/(Y1+Y2);	BETA A B.
Weibull (A, B)	Not available	WEIBULL A B.

SAS

Command line		
1	`DATA SAMPLE;`	Data entry instruction
2	`DO N = 1 TO 50;`	Generates 50 random numbers
3	`Y = RANUNI(213);`	
4	`X = 1 + 2000*RANUNI(6);`	
5	`X = ROUND(X, 1);`	
6	`OUTPUT;`	
7	`END;`	
8	`PROC PRINT;`	Prints the random numbers

COMMAND 3 RANUNI generates uniform random numbers in the interval (0, 1). The numerical "seed" (i.e., the number in parentheses following RANUNI) can be any integer value.

COMMANDS 4–5 Multiplying the uniform random number by 2,000 and adding 1 will generate a random number between 1 and 2,000. The ROUND function (Command 5) will round the resulting random number to the nearest integer.

NOTE: The output from this SAS program is displayed in Figure 7.15 on page 334. SAS commands for the random number generators of several other distributions are provided in Table 7.4.

MINITAB

Command line		
1	`RANDOM 50 C1;`	Generates 50 random numbers
2	`UNIFORM.`	
3	`RANDOM 50 C2;`	
4	`UNIFORM 1 2000.`	
5	`PRINT C1 C2`	Prints the random numbers

COMMANDS 1–2 RANDOM with the UNIFORM subcommand generates uniform random numbers in the interval (0, 1).

COMMANDS 3–4 The UNIFORM subcommand followed by the values 1 and 2000 will generate a random number between 1 and 2000.

NOTE: The output from this MINITAB program is displayed in Figure 7.16. Minitab subcommands for the random number generators of several other probability distributions are provided in Table 7.4.

FIGURE 7.15 ▶
SAS printout for Computer Lab

N	Y	X
1	0.39703	1259
2	0.16258	505
3	0.30135	913
4	0.65456	502
5	0.60613	358
6	0.51486	1170
7	0.77299	622
8	0.84608	606
9	0.06631	1911
10	0.05092	493
11	0.59439	1838
12	0.97203	359
13	0.34312	1965
14	0.15364	1496
15	0.08987	609
16	0.14101	1814
17	0.34850	1806
18	0.59765	618
19	0.29204	1256
20	0.73898	1607
21	0.47006	1092
22	0.64217	158
23	0.80029	303
24	0.55323	1762
25	0.91071	546
26	0.51053	1306
27	0.22638	1059
28	0.59268	1011
29	0.44032	591
30	0.68000	1031
31	0.26740	275
32	0.83772	1691
33	0.59476	1994
34	0.69763	1696
35	0.99709	624
36	0.27695	1930
37	0.61072	561
38	0.35919	1072
39	0.43803	1035
40	0.67097	1067
41	0.26741	471
42	0.07262	1105
43	0.41368	10
44	0.87680	1626
45	0.68880	65
46	0.57243	1721
47	0.89295	965
48	0.64618	269
49	0.75235	180
50	0.68174	216

FIGURE 7.16 ▶
MINITAB printout for Computer Lab

ROW	C1	C2
1	0.262667	170.25
2	0.833319	1166.89
3	0.164871	1808.98
4	0.608885	111.94
5	0.110577	1874.77
6	0.822181	339.32
7	0.772671	312.60
8	0.583832	969.61
9	0.978965	1137.29
10	0.370586	108.39
11	0.323264	1430.28
12	0.408010	750.34
13	0.001304	1714.99
14	0.163034	357.31
15	0.379198	561.50
16	0.399769	99.04
17	0.971068	262.59
18	0.383489	715.27
19	0.936076	1329.24
20	0.009521	113.97
21	0.190185	129.56
22	0.773176	78.92
23	0.646953	1745.57
24	0.869129	181.87
25	0.641075	620.47
26	0.134382	1472.91
27	0.797779	81.93
28	0.722374	121.72
29	0.296690	1097.71
30	0.086229	1157.76
31	0.778634	667.47
32	0.329307	1350.76
33	0.163365	805.58
34	0.420638	623.57
35	0.579784	1860.12
36	0.473011	506.76
37	0.126340	1251.50
38	0.792532	392.08
39	0.066484	909.82
40	0.310511	1659.03
41	0.813840	1358.03
42	0.729962	1713.86
43	0.245253	215.31
44	0.656617	802.93
45	0.077151	291.86
46	0.643837	376.03
47	0.479582	903.08
48	0.947811	817.06
49	0.476421	59.64
50	0.552677	1333.67

References

Freedman, D., Pisani, R., and Purves, R. *Statistics*. New York: W. W. Norton and Co., 1978.

Hogg, R. V., and Craig, A. T. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.

Mendenhall, W. *Introduction to Probability and Statistics*, 8th ed. Boston: Duxbury, 1990.

Mendenhall, W., Wackerly, D. D., and Scheaffer, R. L. *Mathematical Statistics with Applications*, 4th ed. Boston: Duxbury, 1989.

Snedecor, G. W., and Cochran, W. G. *Statistical Methods*, 7th ed. Ames, Iowa: Iowa State University Press, 1980.

CHAPTER EIGHT

Estimation

Objective

To explain the basic concepts of statistical estimation; to present some estimators and to illustrate their use in practical sampling situations involving one or two samples

Contents

8.1 Estimators

An inference about a population parameter can be made in either of two ways—we can estimate the unknown parameter value or we can make a decision about a hypothesized value of the parameter. To illustrate, we can estimate the mean number μ of jobs submitted per hour to a data-processing center or we might want to decide whether the mean μ exceeds some value—say, 60. The method for making a decision about one or more population parameters, called a **statistical test of a hypothesis**, is the topic of Chapter 9. This chapter will be concerned with **estimation**.

Suppose we want to estimate some population parameter, which we denote by θ. For example, θ could be a population mean μ, a population variance σ^2, or the probability $F(a)$ that an observation selected from the population is less than or equal to the value a. A **point estimator**, designated by the symbol $\hat{\theta}$ (i.e., we place a "hat" over the symbol of a parameter to denote its estimator), is a rule or formula that tells us how to use the observations in a sample to compute a single number (a point) that serves as an **estimate** of the value of θ. For example, the mean $\bar{y}$ of a random sample of n observations, $y_1, y_2, \ldots, y_n$, selected from a population is a point estimator of the population mean μ—i.e., $\hat{\mu} = \bar{y}$. Similarly, the sample variance s^2 is a point estimator of σ^2—i.e., $\hat{\sigma}^2 = s^2$.

> **Definition 8.1**
>
> A **point estimator** is a rule or formula that tells us how to calculate a numerical estimate based on the measurements contained in a sample. The single number that results from the calculation is called a **point estimate**.

Another way to estimate the value of a population parameter θ is to use an interval estimator. An **interval estimator** is a rule, usually expressed as a formula, for calculating two points from the sample data. The objective is to form an interval that contains θ with a high degree of confidence. For example, if we estimate the mean number μ of jobs submitted to a data-processing center to be between 40 and 60 jobs per hour, then the interval 40 to 60 is an interval estimate of μ.

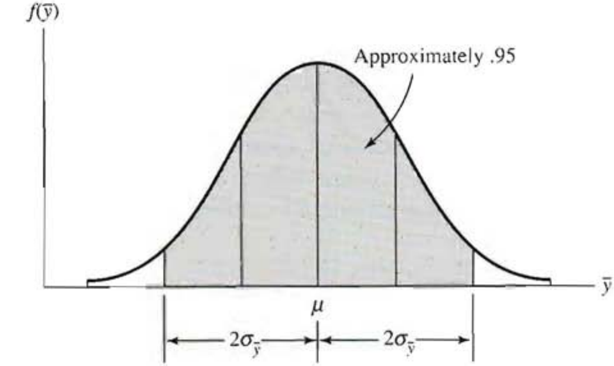
> **Definition 8.2**
>
> An **interval estimator** is a formula that tells us how to use sample data to calculate an interval that estimates a population parameter.

In this chapter, we will identify desirable properties of point and interval estimators, explain how to compare two or more estimators for a single parameter, and show how to measure how good a single estimate actually is. In addition, we will present methods for finding both point and interval estimators, give the formulas for some useful estimators, and show how they can be used in practical situations.

8.2 Properties of Point Estimators

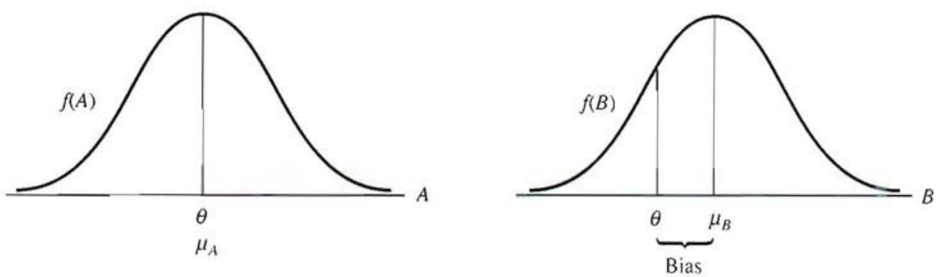
Since a point estimator is calculated from a sample, it possesses a sampling distribution. The sampling distribution of a point estimator completely describes its properties. For example, according to the central limit theorem, the sampling distribution for a sample mean will be approximately normally distributed for large sample sizes, say, $n = 30$ or more, with mean μ and standard error $\sigma/\sqrt{n}$ (see Figure 8.1). The figure shows that a sample mean $\bar{y}$ is equally likely to fall above or below μ and that the probability is approximately .95 that it will not deviate from μ by more than $2\sigma_{\bar{y}} = 2\sigma/\sqrt{n}$.

FIGURE 8.1 ▶ Sampling distribution of a sample mean for large samples



The characteristics exhibited in Figure 8.1 identify the two most desirable properties of estimators. First, we would like the sampling distribution of an estimator to be centered over the parameter being estimated. If the mean of the sampling distribution of an estimator $\hat{\theta}$ is equal to the estimated parameter θ, then the estimator is said to be **unbiased**. If not, the estimator is said to be **biased**. The sample mean is an unbiased estimator of the population mean μ. Sampling distributions for unbiased and biased estimators are shown in Figures 8.2a and 8.2b, respectively.

FIGURE 8.2 ▶ Sampling distributions for unbiased and biased estimators of θ



a. Estimator A is unbiased. **b.** Estimator B is biased.

Definition 8.3

An estimator $\hat{\theta}$ of a parameter θ is **unbiased** if $E(\hat{\theta}) = \theta$. If $E(\hat{\theta}) \neq \theta$, the estimator is said to be **biased**.

Definition 8.4

The **bias** B of an estimator $\hat{\theta}$ is equal to the difference between the mean $E(\hat{\theta})$ of the sampling distribution of $\hat{\theta}$ and θ, i.e.,

$$B = E(\hat{\theta}) - \theta$$

In addition to unbiasedness, we would like the sampling distribution of an estimator to have **minimum variance**, i.e., we want the spread of the sampling distribution to be as small as possible so that estimates will tend to fall close to θ.

Figure 8.3 portrays the sampling distributions of two unbiased estimators, A and B, with A having smaller variance than B. An unbiased estimator that has the minimum variance among all unbiased estimators is called the **minimum variance unbiased estimator** (**MVUE**). For example, $\bar{y}$ is the MVUE for μ. That is, $\text{Var}(\bar{y}) = \sigma^2/n$ is the smallest variance among all unbiased estimators of μ. (Proof omitted.)

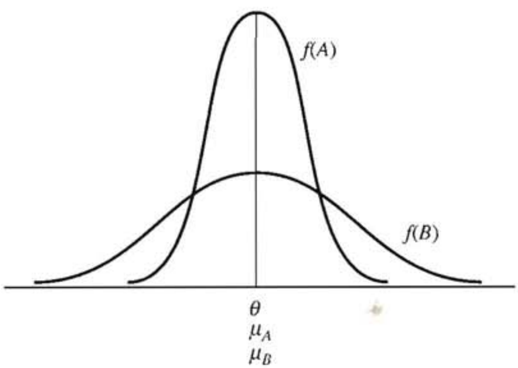


FIGURE 8.3 ▶ Sampling distributions for two unbiased estimators of θ with different variances

Definition 8.5

The **minimum variance unbiased estimator (MVUE)** of a parameter θ is the estimator $\hat{\theta}$ that has the smallest variance of all unbiased estimators.

Sometimes we cannot achieve both unbiasedness and minimum variance in the same estimator. For example, Figure 8.4 shows a biased estimator A with slight bias, but with a smaller variance than the MVUE B. In such a case, we prefer the estimator that minimizes the **mean squared error**, the mean of the squared deviations between $\hat{\theta}$ and θ:

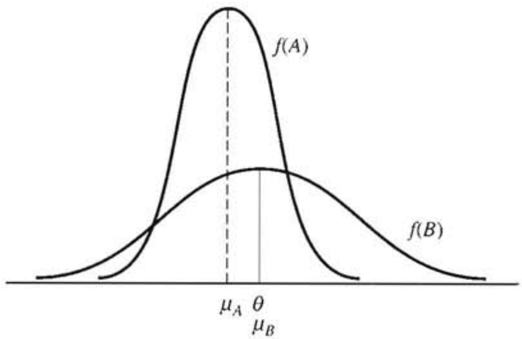
$$\text{Mean squared error for } \hat{\theta}: \quad E[(\hat{\theta} - \theta)^2]$$

It can be shown (proof omitted) that

$$E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + B^2$$

Therefore, if $\hat{\theta}$ is unbiased, i.e., if $B = 0$, then the mean squared error is equal to $V(\hat{\theta})$. Furthermore, when $B = 0$, the estimator $\hat{\theta}$ that yields the smallest mean squared error is also the MVUE for θ.

FIGURE 8.4 ▶
Sampling distributions of biased estimator A and MVUE B



EXAMPLE 8.1

Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from a normal distribution with mean μ and variance σ^2. Show that the sample variance s^2 is an unbiased estimator of the population variance σ^2 when:

a. The sampled population has a normal distribution.
b. The distribution of the sampled population is unknown.

Solution

a. From Theorem 7.4, we know that when sampling from a normal distribution,

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2$$

where χ^2 is a chi-square random variable with $\nu = (n - 1)$ degrees of freedom. Rearranging terms yields

$$s^2 = \frac{\sigma^2}{(n-1)}\chi^2$$

from which it follows that

$$E(s^2) = E\left[\frac{\sigma^2}{(n-1)}\chi^2\right]$$

Applying Theorem 5.2, we obtain

$$E(s^2) = \frac{\sigma^2}{(n-1)}E(\chi^2)$$

We know from Section 5.6 that $E(\chi^2) = \nu$ and $V(\chi^2) = 2\nu$; thus

$$E(s^2) = \frac{\sigma^2}{(n-1)}\nu = \frac{\sigma^2}{(n-1)}(n-1) = \sigma^2$$

Therefore, by Definition 8.3, we conclude that s^2 is an unbiased estimator of σ^2.

b. By the definition of sample variance, we have

$$s^2 = \frac{1}{(n-1)}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum y_i\right)^2}{n}\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2\right]$$

From Theorem 4.4, $\sigma^2 = E(y^2) - \mu^2$. Consequently, $E(y^2) = \sigma^2 + \mu^2$ for a random variable y. Since each y value, $y_1, y_2, \ldots, y_n$, was randomly selected from a population with mean μ and variance σ^2, it follows that

$$E(y_i^2) = \sigma^2 + \mu^2 \quad (i = 1, 2, \ldots, n)$$

and

$$E(\bar{y}^2) = \sigma_{\bar{y}}^2 + (\mu_{\bar{y}})^2 = \sigma^2/n + \mu^2$$

Taking the expected value of s^2 and substituting these expressions, we obtain

$$\begin{aligned}
E(s^2) &= E\left\{\frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2\right]\right\} \\
&= \frac{1}{n-1}\left\{E\left[\sum_{i=1}^{n} y_i^2\right] - E[n(\bar{y})^2]\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n} E[y_i^2] - nE[(\bar{y})^2]\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n} (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\} \\
&= \frac{1}{n-1}[(n\sigma^2 + n\mu^2) - \sigma^2 - n\mu^2] \\
&= \frac{1}{n-1}[n\sigma^2 - \sigma^2] \\
&= \left(\frac{n-1}{n-1}\right)\sigma^2 = \sigma^2
\end{aligned}$$

This shows that, regardless of the nature of the sampled population, s^2 is an unbiased estimator of σ^2.

EXERCISES

OPTIONAL EXERCISES

8.1 Let y_1, y_2, y_3 be a random sample from an exponential distribution with mean θ, i.e., $E(y_i) = \theta$, $i = 1, 2, 3$. Consider three estimators of θ:

$$\hat{\theta}_1 = \bar{y} \qquad \hat{\theta}_2 = y_1 \qquad \hat{\theta}_3 = \frac{y_1 + y_2}{2}$$

a. Show that all three estimators are unbiased.
b. Which of the estimators has the smallest variance? [*Hint:* Recall that, for an exponential distribution, $V(y_i) = \theta^2$.]

8.2 Let $y_1, y_2, y_3, \ldots, y_n$ be a random sample from a Poisson distribution with mean λ, i.e., $E(y_i) = \lambda$, $i = 1, 2, \ldots, n$. Consider four estimators of λ:

$$\hat{\lambda}_1 = \bar{y} \qquad \hat{\lambda}_2 = n(y_1 + y_2 + \cdots + y_n)$$

$$\hat{\lambda}_3 = \frac{y_1 + y_2}{2} \qquad \hat{\lambda}_4 = \frac{y_1}{n}$$

a. Which of the four estimators are unbiased?
b. Of the unbiased estimators, which has the smallest variance? [*Hint:* Recall that, for a Poisson distribution, $V(y_i) = \lambda$.]

8.3 Suppose y has a binomial distribution with parameters n and p.
a. Show that $\hat{p} = y/n$ is an unbiased estimator of p.
b. Find the variance of $\hat{p}$.

8.4 Let $y_1, y_2, \ldots, y_n$ be a random sample from a gamma distribution with parameters $\alpha = 2$ and β unknown.
a. Show that $\bar{y}$ is a biased estimator of β. Compute the bias.
b. Show that $\hat{\beta} = \bar{y}/2$ is an unbiased estimator of β.
c. Find the variance of $\hat{\beta} = \bar{y}/2$. [*Hint:* Recall that, for a gamma distribution, $E(y_i) = 2\beta$ and $V(y_i) = 2\beta^2$.]

8.5 Show that $E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + B^2$, where the bias $B = E(\hat{\theta}) - \theta$. [*Hint:* Write $(\hat{\theta} - \theta) = [\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta]$.]

8.6 Let y_1 be a sample of size 1 from a uniform distribution over the interval from 2 to θ.
a. Show that y_1 is a biased estimator of θ and compute the bias.
b. Show that $2(y_1 - 1)$ is an unbiased estimator of θ.
c. Find the variance of $2(y_1 - 1)$.

8.7 Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal distribution, with mean μ and variance σ^2. Shcw that the variance of the sampling distribution of s^2 is $2\sigma^4/(n - 1)$.

8.3 Finding Point Estimators: Methods of Estimation

There are a number of different methods for finding point estimators of parameters. Two classical methods, the **method of moments** and the **method of maximum likelihood**, are the main topics of this section. These techniques produce the estimators of the population parameters encountered in Sections 8.5–8.11. A discussion of other methods for finding point estimators is beyond the scope of this text; we give a brief description of these other methods and refer you to the references given at the end of this chapter.

METHOD OF MOMENTS The method of estimation that we have employed thus far is to use sample numerical descriptive measures to estimate their population parameters. For example, we used the sample mean $\bar{y}$ to estimate the population mean μ. From Definition 4.7, we know that the parameter $E(y) = \mu$ is the first moment about the origin or, as it is sometimes called, the **first population moment**. Similarly, we define the **first sample moment** as

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

The general technique of using sample moments to estimate their corresponding population moments is called the **method of moments**. For the parameters discussed in this chapter, the method of moments yields estimators that have the two desired properties mentioned earlier, i.e., unbiased estimators and estimators with minimum variance.

Definition 8.6

Let $y_1, y_2, \ldots, y_n$ represent a random sample of size n from some probability distribution (discrete or continuous). The ***k*th population moment** and ***k*th sample moment** are defined as follows:

kth population moment: $E(y^k)$

kth sample moment: $m^k = \frac{\sum_{i=1}^{n} y_i^k}{n}$

For the case $k = 1$, the first population moment is $E(y) = \mu$ and the first sample moment is $m = \bar{y}$.

Definition 8.7

Let $y_1, y_2, \ldots, y_n$ represent a random sample of size n from a probability distribution (discrete or continuous) with parameters $\theta_1, \theta_2, \ldots, \theta_m$. Then the **moment estimators,** $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$, are obtained by equating the first m sample moments to the corresponding first m population moments:

$$E(y) = \frac{1}{n}\sum y_i$$

$$E(y^2) = \frac{1}{n}\sum y_i^2$$

$$\vdots$$

$$E(y^m) = \frac{1}{n}\sum y_i^m$$

and solving for $\theta_1, \theta_2, \ldots, \theta_m$. (Note that the first m population moments will be functions of $\theta_1, \theta_2, \ldots, \theta_m$.)

For the special case $m = 1$, the moment estimator of θ is some function of the sample mean $\bar{y}$.

EXAMPLE 8.2

The response rate y of auditory nerve fibers in cats has an approximate Poisson distribution with unknown mean λ (*Journal of the Acoustical Society of America*, Feb. 1986). Suppose the auditory nerve fiber response rate (recorded as number of spikes per 200 milliseconds of noise burst) was measured in each of a random sample of 10 cats. The data follow:

15.1 14.6 12.0 19.2 16.1 15.5 11.3 18.7 17.1 17.2

Calculate a point estimate for the mean response rate λ using the method of moments.

Solution

We have only one parameter, λ, to estimate; therefore, the moment estimator is found by setting the first population moment, $E(y)$, equal to the first sample moment, $\bar{y}$. For the Poisson distribution, $E(y) = \lambda$; hence, the moment estimator is

$$\hat{\lambda} = \bar{y}$$

For this example,

$$\bar{y} = \frac{15.1 + 14.6 + \cdots + 17.2}{10} = 15.68$$

Thus, our estimate of the mean auditory nerve fiber response rate λ is 15.68 spikes per 200 milliseconds of noise burst.

EXAMPLE 8.3 (OPTIONAL) The time y until failure from fatigue cracks for underground cable possesses an approximate gamma probability distribution with parameters α and β (*IEEE Transactions on Energy Conversion*, Mar. 1986). Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations on the random variable y. Find the moment estimators of α and β.

Solution Since we must estimate two parameters, α and β, the method of moments requires that we set the first two population moments equal to their corresponding sample moments. From Section 5.6, we know that for the gamma distribution

$$\mu = E(y) = \alpha\beta$$
$$\sigma^2 = \alpha\beta^2$$

Also, from Theorem 4.4, $\sigma^2 = E(y^2) - \mu^2$. Thus, $E(y^2) = \sigma^2 + \mu^2$. Then for the gamma distribution, the first two population moments are

$$E(y) = \alpha\beta$$
$$E(y^2) = \sigma^2 + \mu^2 = \alpha\beta^2 + (\alpha\beta)^2$$

Setting these equal to their respective sample moments, we have

$$\hat{\alpha}\hat{\beta} = \bar{y}$$
$$\hat{\alpha}\hat{\beta}^2 + (\hat{\alpha}\hat{\beta})^2 = \frac{\sum y_i^2}{n}$$

Substituting $\bar{y}$ for $\hat{\alpha}\hat{\beta}$ in the second equation, we obtain

$$\bar{y}\hat{\beta} + (\bar{y})^2 = \frac{\sum y_i^2}{n}$$

or,

$$\begin{aligned}\bar{y}\hat{\beta} &= \frac{\sum y_i^2}{n} - (\bar{y})^2 \\ &= \frac{\sum y_i^2 - n(\bar{y})^2}{n} = \frac{\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}}{n} \\ &= \frac{(n-1)s^2}{n}\end{aligned}$$

Our two equations are now reduced to

$$\hat{\alpha}\hat{\beta} = \bar{y}$$
$$\bar{y}\hat{\beta} = \left(\frac{n-1}{n}\right)s^2$$

Solving these equations simultaneously, we obtain the moment estimators

$$\hat{\beta} = \left(\frac{n-1}{n}\right)\frac{s^2}{\bar{y}} \quad \text{and} \quad \hat{\alpha} = \left(\frac{n}{n-1}\right)\frac{\bar{y}^2}{s^2} = \left(\frac{n}{n-1}\right)\left(\frac{\bar{y}}{s}\right)^2$$

METHOD OF MAXIMUM LIKELIHOOD The method of maximum likelihood and an exposition of the properties of maximum likelihood estimators are the results of work by Sir Ronald A. Fisher (1890–1962). Fisher's logic can be seen by considering the following example: If we randomly select a sample of n observations, $y_1, y_2, \ldots, y_n$, of a discrete random variable y and if the probability distribution $p(y)$ is a function of a single parameter θ, then the probability of observing these n independent values of y is

$$p(y_1, y_2, \ldots, y_n) = p(y_1)p(y_2) \cdots p(y_n)$$

Fisher called this joint probability of the sample values, $y_1, y_2, \ldots, y_n$, the **likelihood** L of the sample, and suggested that one should choose as an estimate of θ the value of θ that maximizes L. If the likelihood L of the sample is a function of two parameters, say, θ_1 and θ_2, then the maximum likelihood estimates of θ_1 and θ_2 are the values that maximize L. The notion is easily extended to the situation in which L is a function of more than two parameters.

Definition 8.8

a. The **likelihood** L of a sample of n observations, $y_1, y_2, \ldots, y_n$, is the joint probability function $p(y_1, y_2, \ldots, y_n)$ when $y_1, y_2, \ldots, y_n$ are discrete random variables.
b. The **likelihood** L of a sample of n observations, $y_1, y_2, \ldots, y_n$, is the joint density function $f(y_1, y_2, \ldots, y_n)$ when $y_1, y_2, \ldots, y_n$ are continuous random variables.

Note: For fixed values of $y_1, y_2, \ldots, y_n$, L will be a function of θ.

Theorem 8.1 follows directly from the definition of independence and Definitions 6.8 and 6.9.

Theorem 8.1

a. Let $y_1, y_2, \ldots, y_n$ represent a random sample of n observations on a random variable y. Then $L = p(y_1)p(y_2) \cdots p(y_n)$ when y is a discrete random variable with probability distribution $p(y)$.
b. Let $y_1, y_2, \ldots, y_n$ represent a random sample of n observations on a random variable y. Then $L = f(y_1)f(y_2) \cdots f(y_n)$ when y is a continuous random variable with density function $f(y)$.

Definition 8.9

Let L be the likelihood of a sample, where L is a function of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the **maximum likelihood estimators** of $\theta_1, \theta_2, \ldots, \theta_k$ are the values of $\theta_1, \theta_2, \ldots, \theta_k$ that maximize L.

Fisher showed that maximum likelihood estimators of population means and proportions possess some very desirable properties. As the sample size n becomes larger and larger, the sampling distribution of a maximum likelihood estimator $\hat{\theta}$ tends to become more and more nearly normal, with mean equal to θ and a variance that is equal to or less than the variance of *any other* estimator. Although these properties of maximum likelihood estimators pertain only to estimates based on large samples, they tend to provide support for the maximum likelihood method of estimation. The properties of maximum likelihood estimators based on small samples can be acquired by using the methods of Chapters 4, 5, and 6 to derive their sampling distributions or, at the very least, to acquire their means and variances.

To simplify our explanation of how to find a maximum likelihood estimator, we will assume that L is a function of a single parameter θ. Then, from differential calculus, we know that the value of θ that maximizes (or minimizes) L is the value for which $\frac{dL}{d\theta} = 0$. Obtaining this solution, which always yields a maximum (proof omitted), can be difficult because L is usually the product of a number of quantities involving θ. Differentiating a sum is easier than differentiating a product, so we attempt to maximize the logarithm of L rather than L itself. Since the logarithm of L is a monotonically increasing function of L, L will be maximized by the same value of θ that maximizes its logarithm. We illustrate the procedure in Examples 8.4 and 8.5.

EXAMPLE 8.4

Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations on a random variable y with the exponential density function

$$f(y) = \begin{cases} \dfrac{e^{-y/\beta}}{\beta} & \text{if } 0 \le y < \infty \\ 0 & \text{elsewhere} \end{cases}$$

Determine the maximum likelihood estimator of β.

Solution

Since $y_1, y_2, \ldots, y_n$ are independent random variables, we have

$$\begin{aligned} L &= f(y_1)f(y_2)\cdots f(y_n) \\ &= \left(\frac{e^{-y_1/\beta}}{\beta}\right)\left(\frac{e^{-y_2/\beta}}{\beta}\right)\cdots\left(\frac{e^{-y_n/\beta}}{\beta}\right) \\ &= \frac{e^{-\sum_{i=1}^{n} y_i/\beta}}{\beta^n} \end{aligned}$$

Taking the natural logarithm of L yields

$$\ln(L) = \ln(e^{-\sum_{i=1}^{n} y_i/\beta}) - n\ln(\beta) = -\frac{\sum_{i=1}^{n} y_i}{\beta} - n\ln(\beta)$$

Then

$$\frac{d\ln(L)}{d\beta} = \frac{\sum_{i=1}^{n} y_i}{\beta^2} - \frac{n}{\beta}$$

Setting this derivative equal to 0 and solving for $\hat{\beta}$, we obtain

$$\frac{\sum_{i=1}^{n} y_i}{\hat{\beta}^2} - \frac{n}{\hat{\beta}} = 0 \qquad \text{or} \qquad n\hat{\beta} = \sum_{i=1}^{n} y_i$$

This yields

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}$$

Therefore, the maximum likelihood estimator (MLE) of β is the sample mean $\bar{y}$, i.e., $\hat{\beta} = \bar{y}$.

EXAMPLE 8.5 (OPTIONAL) Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations on the random variable y, where $f(y)$ is a normal density function with mean μ and variance σ^2. Find the maximum likelihood estimators of μ and σ^2.

Solution Since $y_1, y_2, \ldots, y_n$ are independent random variables, it follows that

$$\begin{aligned} L &= f(y_1)f(y_2)\cdots f(y_n) \\ &= \left(\frac{e^{-(y_1-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}\right)\left(\frac{e^{-(y_2-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}\right)\cdots\left(\frac{e^{-(y_n-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}\right) \\ &= \frac{e^{-\sum_{i=1}^{n}(y_i-\mu)^2/(2\sigma^2)}}{\sigma^n(2\pi)^{n/2}} \end{aligned}$$

and

$$\ln(L) = -\frac{\sum_{i=1}^{n}(y_i-\mu)^2}{2\sigma^2} - \frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln(2\pi)$$

Taking derivatives of ln(L) with respect to μ and σ and setting them equal to 0 yields

$$\frac{d\ \ln(L)}{d\mu} = \frac{\sum_{i=1}^{n} 2(y_i - \hat{\mu})}{2\hat{\sigma}^2} - 0 - 0 = 0$$

and

$$\frac{d\ \ln(L)}{d\sigma^2} = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu})^2}{2\hat{\sigma}^4} - \frac{n}{2}\left(\frac{1}{\hat{\sigma}^2}\right) - 0 = 0$$

The values of μ and σ^2 that maximize L [and hence $\ln(L)$] will be the simultaneous solution of these two equations. The first equation reduces to

$$\sum_{i=1}^{n} (y_i - \hat{\mu}) = 0 \quad \text{or} \quad \sum_{i=1}^{n} y_i - n\hat{\mu} = 0$$

and it follows that

$$n\hat{\mu} = \sum_{i=1}^{n} y_i \quad \text{and} \quad \hat{\mu} = \bar{y}$$

Substituting $\hat{\mu} = \bar{y}$ into the second equation and multiplying by $2\hat{\sigma}^2$, we obtain

$$\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\hat{\sigma}^2} = n \quad \text{or} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}$$

Therefore, the maximum likelihood estimators of μ and σ^2 are

$$\hat{\mu} = \bar{y} \quad \text{and} \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}$$

Note that the maximum likelihood estimator of σ^2 is equal to the sum of squares of deviations $\Sigma_{i=1}^{n} (y_i - \bar{y})^2$ divided by n, whereas the sample variance s^2 uses a divisor of $(n - 1)$. We showed in Example 8.1 that s^2 is an unbiased estimator of σ^2. Therefore, the maximum likelihood estimator

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n} = \frac{(n - 1)}{n} s^2$$

is a biased estimator of σ^2.

METHOD OF LEAST SQUARES Another useful technique for finding point estimators is the **method of least squares**. This method finds the estimate of θ that minimizes the mean squared error (MSE):

$$\text{MSE} = E(\hat{\theta} - \theta)^2$$

The method of least squares—a widely used estimation technique—is discussed in detail in Chapter 11. Several other estimation methods are briefly described here; consult the references at the end of this chapter if you want to learn more about their use.

JACKKNIFE ESTIMATORS Tukey (1958) developed a "leave-one-out-at-a-time" approach to estimation, called the **jackknife**,* that is gaining increasing acceptance among practitioners. Let $y_1, y_2, \ldots, y_n$ be a sample of size n from a population with parameter θ. An estimate $\hat{\theta}_{(i)}$ is obtained by omitting the ith observation (i.e., y_i) and computing the estimate based on the remaining $(n - 1)$ observations. This calculation is performed for each observation in the data set, and the procedure results in n estimates of θ: $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \ldots, \hat{\theta}_{(n)}$. The **jackknife estimator** of θ is then some suitably chosen linear combination (e.g., a weighted average) of the n estimates. Application of the jackknife is suggested for situations where we are likely to have outliers or biased samples, or find it difficult to assess the variability of the more traditional estimators.

ROBUST ESTIMATORS Many of the estimators discussed in Sections 8.5–8.11 are based on the assumption that the sampled population is approximately normal. When the distribution of the sampled population deviates greatly from normality, such estimators do not have desirable properties (e.g., unbiasedness and minimum variance). An estimator that performs well for a very wide range of probability distributions is called a **robust estimator**. For example, a robust estimate of the population mean μ, called the ***M*-estimator**, compares favorably to the sample mean $\bar{y}$ when the sampled population is normal and is considerably better than $\bar{y}$ when the population is heavy-tailed. See Mosteller and Tukey (1977) and Devore (1987) for a good practical discussion of robust estimation techniques.

BAYES ESTIMATORS The classical approach to estimation is based on the concept that the unknown parameter θ is a constant. All the information available to us about θ is contained in the random sample $y_1, y_2, \ldots, y_n$ selected from the relevant population. In contrast, the **Bayesian** approach to estimation regards θ as a random variable with some known (**prior**) probability distribution $g(\theta)$. The sample information is used to modify the prior distribution on θ to obtain the **posterior** distribution, $f(\theta \mid y_1, y_2, \ldots, y_n)$. The **Bayes estimator** of θ is then the mean of the posterior probability distribution [see Mendenhall, Wackerly, and Scheaffer (1989)].

OPTIONAL EXERCISES

8.8 A binomial experiment consisting of n trials resulted in Bernoulli observations $y_1, y_2, \ldots, y_n$, where

$$y_i = \begin{cases} 1 & \text{if the } i\text{th trial was a success} \\ 0 & \text{if not} \end{cases}$$

and $P(y_i = 1) = p$, $P(y_i = 0) = 1 - p$. Let $y = \Sigma_{i=1}^{n} y_i$ be the number of successes in n trials.

*The procedure derives its name from the Boy Scout jackknife; like the jackknife, the procedure serves as a handy tool in a variety of situations when specialized techniques may not be available.

a. Find the moment estimator of p.
b. Is the moment estimator unbiased?
c. Find the maximum likelihood estimator of p. [*Hint:* $L = p^y(1 - p)^{n-y}$.]
d. Is the maximum likelihood estimator unbiased?

8.9 Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from a Poisson distribution with probability function

$$p(y) = \frac{e^{-\lambda}\lambda^y}{y!} \qquad (y = 0, 1, 2, \ldots)$$

a. Find the maximum likelihood estimator of λ.
b. Is the maximum likelihood estimator unbiased?

8.10 Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations on a random variable y, where $f(y)$ is a gamma density function with $\alpha = 2$ and unknown β:

$$f(y) = \begin{cases} \dfrac{ye^{-y/\beta}}{\beta^2} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the maximum likelihood estimator of β.
b. Find $E(\hat{\beta})$ and $V(\hat{\beta})$.

8.11 Refer to Exercise 8.10.
a. Find the moment estimator of β.
b. Find $E(\hat{\beta})$ and $V(\hat{\beta})$.

8.12 Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from a normal distribution with mean 0 and unknown variance σ^2. Find the maximum likelihood estimator of σ^2.

8.13 Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from an exponential distribution with density

$$f(y) = \begin{cases} \dfrac{1}{\beta}e^{-y/\beta} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the moment estimator of β.
b. Is the moment estimator unbiased?
c. Find $V(\hat{\beta})$.

8.4 Finding Interval Estimators: The Pivotal Method

In Section 8.1, we defined an interval estimator as a rule that tells how to use the sample observations to calculate two numbers that define an interval that will enclose the estimated parameter with a high degree of confidence. The resulting random interval (random, because the sample observations used to calculate the endpoints of the interval are random variables) is called a **confidence interval**, and the probability (prior to sampling) that it contains the estimated parameter is called its **confidence coefficient**. If a confidence interval has a confidence coefficient equal to .95, we call it a 95% confidence interval. If the confidence coefficient is .99, the interval is said to be a 99% confidence interval, etc. A more practical interpretation of the confidence coefficient for a confidence interval is given later in this section.

Definition 8.10

The **confidence coefficient** for a confidence interval is equal to the probability that the random interval, prior to sampling, will contain the estimated parameter.

One way to find a confidence interval for a parameter θ is to acquire a **pivotal statistic**, a statistic that is a function of the sample values and the single parameter θ. Because many statistics are approximately normally distributed when the sample size n is large (central limit theorem), we can construct confidence intervals for their expected values using the standard normal random variable z as a pivotal statistic.

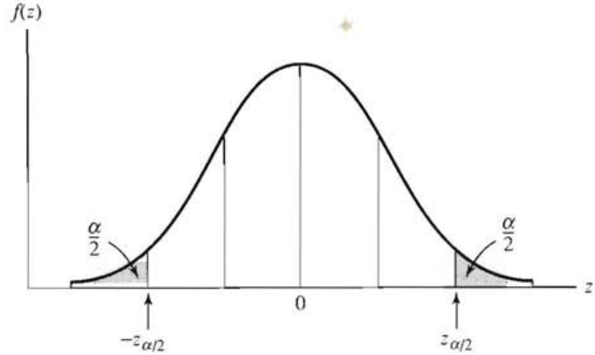
To illustrate, let $\hat{\theta}$ be a statistic with a sampling distribution that is approximately normally distributed for large samples with mean $E(\hat{\theta}) = \theta$ and standard error $\sigma_{\hat{\theta}}$. Then,

$$z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$$

is a standard normal random variable. Since z is also a function of only the sample statistic $\hat{\theta}$ and the parameter θ, we will use it as a pivotal statistic. To derive a confidence interval for θ, we first make a probability statement about the pivotal statistic. To do this, we locate values $z_{\alpha/2}$ and $-z_{\alpha/2}$ that place a probability of $\alpha/2$ in each tail of the z distribution (see Figure 8.5), i.e., $P(z > z_{\alpha/2}) = \alpha/2$. It can be seen from Figure 8.5 that

$$P(-z_{\alpha/2} \leq z \leq z_{\alpha/2}) = 1 - \alpha$$

FIGURE 8.5 ► Locating $z_{\alpha/2}$ for a confidence interval



Substituting the expression for z into the probability statement and using some simple algebraic operations on the inequality, we obtain

$$
\begin{aligned}
P(-z_{\alpha/2} \le z \le z_{\alpha/2}) &= P\left(-z_{\alpha/2} \le \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \le z_{\alpha/2}\right) \\
&= P(-z_{\alpha/2}\sigma_{\hat{\theta}} \le \hat{\theta} - \theta \le z_{\alpha/2}\sigma_{\hat{\theta}}) \\
&= P(-\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le -\theta \le -\hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) \\
&= P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \le \theta \le \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1 - \alpha
\end{aligned}
$$

Therefore, the probability that the interval formed by

$$\text{LCL} = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \quad \text{to} \quad \text{UCL} = \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$$

will enclose θ is equal to $(1 - \alpha)$. The quantities LCL and UCL are called the **lower** and **upper confidence limits**, respectively, for the confidence interval. The confidence coefficient for the interval will be $(1 - \alpha)$.

The derivation of a large-sample $(1 - \alpha)100\%$ confidence interval for θ is summarized in Theorem 8.2.

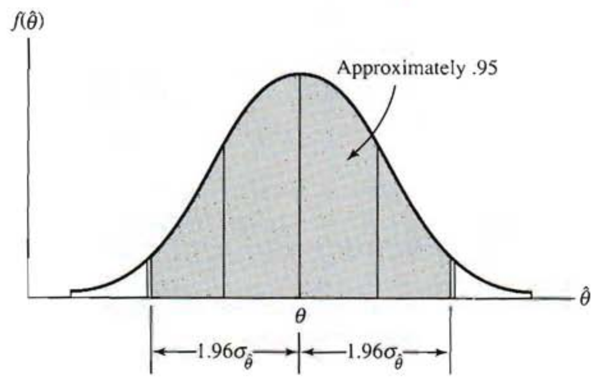
Theorem 8.2

Let $\hat{\theta}$ be normally distributed for large samples with $E(\hat{\theta}) = \theta$ and standard error $\sigma_{\hat{\theta}}$. Then a $(1 - \alpha)100\%$ confidence interval for θ is

$$\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \quad \text{to} \quad \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}$$

The large-sample confidence interval can also be acquired intuitively by examining Figure 8.6. The z value corresponding to an area $A = .475$—i.e., the z value that places area $\alpha/2 = .025$ in the upper tail of the z distribution—is (see Table 4 of Appendix II) $z_{.025} = 1.96$. Therefore, the probability that $\hat{\theta}$ will lie within $1.96\sigma_{\hat{\theta}}$ of θ is .95. You can see from Figure 8.6 that whenever $\hat{\theta}$ falls within the interval $\theta \pm 1.96\sigma_{\hat{\theta}}$, then the interval $\hat{\theta} \pm 1.96\sigma_{\hat{\theta}}$ will enclose θ. Therefore, $\hat{\theta} \pm 1.96\sigma_{\hat{\theta}}$ yields a 95% confidence interval for θ.

FIGURE 8.6 ►
The sampling distribution of θ for large samples



We may encounter one slight difficulty when we attempt to apply this confidence interval in practice. It is often the case that $\sigma_{\hat{\theta}}$ is a function of the parameter θ that we are attempting to estimate. However, when the sample size n is large (which we have assumed throughout the derivation), we can substitute the estimate $\hat{\theta}$ for the parameter θ to obtain an approximate value for $\sigma_{\hat{\theta}}$.

In Example 8.6 we will use a pivotal statistic to find a confidence interval for μ when the sample size is small, say, $n < 30$.

EXAMPLE 8.6

Let $\bar{y}$ and s^2 be the sample mean and variance based on a random sample of n observations ($n < 30$) from a normal distribution with mean μ and variance σ^2. Find a 95% confidence interval for μ.

Solution

A pivotal statistic for μ can be constructed using the t statistic of Chapter 7. By Definition 7.5,

$$t = \frac{z}{\sqrt{\chi^2/\nu}}$$

where z and χ^2 are independent random variables and χ^2 is based on ν degrees of freedom. We know that $\bar{y}$ is normally distributed and that

$$z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal random variable. From Theorem 7.4, it follows that

$$\frac{(n-1)s^2}{\sigma^2} = \chi^2$$

is a chi-square random variable with $\nu = (n-1)$ degrees of freedom. We state (without proof) that $\bar{y}$ and s^2 are independent when they are based on a random sample selected from a normal distribution. Therefore, z and χ^2 will be independent random variables. Substituting the expressions for z and χ^2 into the formula for t, we obtain

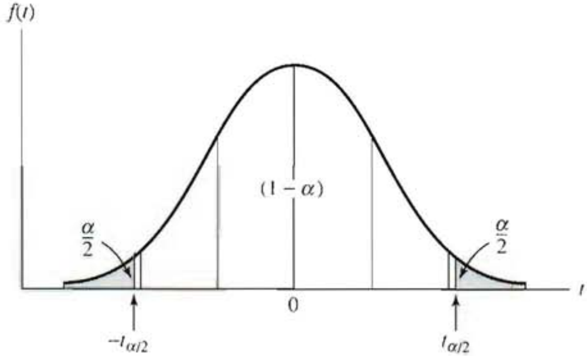
$$t = \frac{z}{\sqrt{\chi^2/\nu}} = \frac{\dfrac{\bar{y} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)s^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

Note that the pivotal statistic is a function only of μ and the sample statistics $\bar{y}$ and s^2.

The next step in finding a confidence interval for μ is to make a probability statement about the pivotal statistic t. We will select two values of t, call them $t_{\alpha/2}$ and $-t_{\alpha/2}$, that correspond to probabilities of $\alpha/2$ in the upper and lower tails, respectively, of the t distribution (see Figure 8.7). From Figure 8.7, it can be seen that

$$P(-t_{\alpha/2} \le t \le t_{\alpha/2}) = 1 - \alpha$$

FIGURE 8.7 ► The location of $t_{\alpha/2}$ and $-t_{\alpha/2}$ for a Student's t distribution



Substituting the expression for t into the probability statement, we obtain

$$P(-t_{\alpha/2} \le t \le t_{\alpha/2}) = P\left(-t_{\alpha/2} \le \frac{\bar{y} - \mu}{s/\sqrt{n}} \le t_{\alpha/2}\right) = 1 - \alpha$$

Multiplying the inequality within the brackets by $s/\sqrt{n}$, we obtain

$$P\left[-t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \le \bar{y} - \mu \le t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right] = 1 - \alpha$$

Subtracting $\bar{y}$ from each part of the inequality yields

$$P\left[-\bar{y} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \le -\mu \le -\bar{y} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right] = 1 - \alpha$$

Finally, we multiply each term of the inequality by (-1), thereby reversing the inequality signs. The result is

$$P\left[\bar{y} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{y} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)\right] = 1 - \alpha$$

Therefore, a $(1 - \alpha)100\%$ confidence interval for μ when n is small is

$$\bar{y} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \quad \text{to} \quad \bar{y} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

We now apply the confidence interval derived in Example 8.6 to a practical situation.

EXAMPLE 8.7

Chemical plants must be regulated to prevent the poisoning of fish in nearby rivers or streams. One of the measurements made on fish to evaluate the potential toxicity of chemicals is the length of mature fish. If a river or stream is inhabited by an abundance of mature fish with lengths less than the average length of mature members of their species, we have strong evidence that the river is being chemically contaminated. A chemical plant, under investigation for chlorine poisoning of a stream, has hired a biologist to estimate the mean length of fathead minnows (the main inhabitants of the stream) exposed to 20 micrograms of chlorine per liter of water. The biologist captures 20 newborn fathead minnows from the stream and rears them in aquaria with this chlorine concentration. The length of each (in millimeters) is measured after a 10-week maturation period, with the following results:

$$\bar{y} = 27.5$$

$$s = 2.6$$

Construct a 95% confidence interval for the true mean length of fathead minnows reared in chlorine-contaminated water. Assume that the lengths of the fathead minnows are approximately normal.

Solution

Recall that the sampling distribution of the t statistic depends on its degrees of freedom, ν. The tabulated values t_a, such that $P(t \geq t_a) = a$, are given in Table 7 of Appendix II, for values of ν from 1 to 29, as well as the value of t_a when ν becomes infinitely large. An abbreviated version of this table is shown in Table 8.1 on page 358. For example, suppose a t statistic is based on $\nu = 4$ degrees of freedom (df) and we want to find the value t_a that places probability $a = .025$ in the upper tail of the t distribution. The appropriate value, shaded in Table 8.1, is $t_{.025} = 2.776$.

For our example, $n = 20$ and t will possess $(n - 1) = 19$ degrees of freedom. Since we want to find a $95\% = (1 - \alpha)100\%$ confidence interval for the mean length μ of fathead minnows, $\alpha = .05$; we must find the value $t_{.025}$ corresponding to $a = .025$ and 19 degrees of freedom. This value is given in Table 7 of Appendix II as $t_{\alpha/2} = t_{.025} = 2.093$. Then the confidence interval is

$$\bar{y} \pm t_{.025}\left(\frac{s}{\sqrt{n}}\right) = 27.5 \pm 2.093\left(\frac{2.60}{\sqrt{20}}\right)$$

$$= 27.5 \pm 1.22 \quad \text{or} \quad (26.28,\ 28.72)$$

Since the confidence coefficient is .95, we say that we are 95% confident that the interval from 26.28 to 28.72 millimeters contains the true mean length, μ, of fathead minnows reared in chlorine-contaminated water.

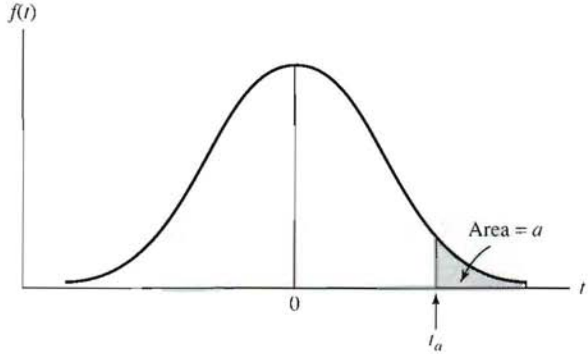


TABLE 8.1 An Abbreviated Version of Table 7 of Appendix II

Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947

To demonstrate the interpretation of a confidence interval, we programmed a computer to draw 1,000 samples of size $n = 10$ from a normal distribution with mean $\mu = 10$ and variance $\sigma^2 = 1$. A 95% confidence interval for μ was computed for each of the 1,000 samples. These are shown in Table 8.2. Only the 50 intervals that are starred (*) fail to enclose the mean $\mu = 10$. The proportion that enclose μ, .95, is exactly equal to the confidence coefficient. This explains why we are reasonably confident that the interval calculated in Example 8.7 (26.28, 28.72), encloses the true value of μ. *If we were to employ our interval estimator on repeated occasions, 95% of the intervals constructed would contain* μ.

TABLE 8.2 One Thousand 95% Confidence Intervals for the Mean of a Normal Distribution ($\mu = 10$, $\sigma^2 = 1$). [*Note:* Starred (*) intervals fail to include $\mu = 10$.]

SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL
1	(9.574	, 11.183)	2	(9.370	, 11.137)	3	(9.425	, 10.403)	4	(9.356	, 10.777)	5	(9.956	, 11.455)
6	(8.807	, 10.589)	7	(9.188	, 10.588)	8	(9.798	, 11.390)	9	(10.057	, 11.561)*	10	(9.793	, 11.009)
11	(9.180	, 10.848)	12	(9.735	, 10.399)	13	(9.518	, 10.785)	14	(9.872	, 11.003)	15	(9.028	, 10.907)
16	(9.414	, 11.107)	17	(9.603	, 10.816)	18	(9.469	, 10.896)	19	(8.758	, 9.889)*	20	(9.617	, 11.285)
21	(8.926	, 10.389)	22	(9.710	, 10.512)	23	(8.847	, 10.667)	24	(9.148	, 10.675)	25	(9.722	, 11.017)
26	(9.575	, 11.221)	27	(8.820	, 10.664)	28	(9.222	, 11.015)	29	(9.525	, 10.717)	30	(9.036	, 10.802)
31	(8.758	, 10.151)	32	(9.043	, 10.650)	33	(9.819	, 10.958)	34	(9.085	, 10.610)	35	(8.847	, 10.692)
36	(9.590	, 10.551)	37	(9.826	, 11.063)	38	(9.506	, 10.615)	39	(9.322	, 10.401)	40	(8.892	, 10.239)
41	(9.519	, 10.812)	42	(9.023	, 10.640)	43	(9.560	, 10.651)	44	(8.885	, 10.327)	45	(9.901	, 11.353)
46	(9.348	, 10.324)	47	(9.188	, 10.766)	48	(9.173	, 10.563)	49	(9.039	, 10.567)	50	(9.234	, 10.864)
51	(8.799	, 10.503)	52	(9.870	, 11.382)	53	(8.706	, 10.830)	54	(9.690	, 11.002)	55	(9.340	, 10.864)
56	(9.244	, 10.666)	57	(9.910	, 11.487)	58	(8.690	, 10.111)	59	(9.149	, 10.661)	60	(9.320	, 10.740)
61	(9.052	, 11.085)	62	(9.593	, 10.985)	63	(9.103	, 10.768)	64	(9.429	, 10.023)	65	(9.275	, 10.032)
66	(9.139	, 11.033)	67	(9.621	, 10.916)	68	(9.464	, 10.571)	69	(9.717	, 10.938)	70	(8.895	, 10.638)
71	(9.496	, 11.260)	72	(9.124	, 10.437)	73	(9.416	, 10.718)	74	(8.516	, 10.296)	75	(8.991	, 10.290)
76	(9.225	, 10.554)	77	(8.641	, 10.451)	78	(9.598	, 11.359)	79	(9.443	, 11.217)	80	(9.384	, 11.055)
81	(9.088	, 10.592)	82	(9.777	, 11.142)	83	(9.160	, 10.683)	84	(8.969	, 10.122)	85	(9.275	, 10.372)
86	(9.676	, 10.772)	87	(9.075	, 10.507)	88	(8.425	, 10.059)	89	(9.382	, 10.761)	90	(8.646	, 10.648)
91	(9.414	, 10.410)	92	(9.661	, 10.909)	93	(9.240	, 10.679)	94	(9.138	, 10.469)	95	(9.497	, 10.824)
96	(8.871	, 10.627)	97	(9.402	, 10.923)	98	(9.241	, 10.695)	99	(8.832	, 10.790)	100	(9.546	, 10.472)
101	(9.604	, 10.705)	102	(9.315	, 10.519)	103	(9.306	, 10.572)	104	(9.129	, 10.750)	105	(9.566	, 10.658)
106	(9.490	, 10.558)	107	(9.052	, 10.387)	108	(9.464	, 10.969)	109	(9.899	, 10.919)	110	(9.045	, 10.284)
111	(9.112	, 10.341)	112	(9.593	, 10.836)	113	(8.726	, 10.515)	114	(9.511	, 10.916)	115	(9.535	, 10.938)
116	(9.526	, 10.690)	117	(8.848	, 10.395)	118	(8.765	, 10.126)	119	(9.151	, 10.696)	120	(8.764	, 10.382)
121	(9.598	, 10.578)	122	(8.988	, 10.286)	123	(9.435	, 10.890)	124	(8.833	, 10.335)	125	(9.644	, 10.744)
126	(9.589	, 10.533)	127	(9.284	, 10.174)	128	(9.110	, 10.522)	129	(9.502	, 10.728)	130	(9.144	, 11.044)
131	(9.344	, 10.851)	132	(9.915	, 11.372)	133	(9.252	, 10.399)	134	(9.833	, 11.188)	135	(9.268	, 10.219)
136	(9.681	, 10.804)	137	(9.082	, 10.719)	138	(9.374	, 10.198)	139	(9.303	, 10.781)	140	(9.046	, 10.329)
141	(9.191	, 10.640)	142	(9.777	, 10.812)	143	(8.622	, 10.513)	144	(9.175	, 10.931)	145	(9.227	, 10.863)
146	(9.167	, 10.455)	147	(9.323	, 11.067)	148	(9.148	, 10.282)	149	(9.169	, 10.360)	150	(9.635	, 10.989)
151	(9.877	, 10.723)	152	(8.950	, 10.546)	153	(8.911	, 9.986)*	154	(9.039	, 10.502)	155	(9.765	, 10.749)
156	(9.354	, 10.900)	157	(8.566	, 9.779)*	158	(9.218	, 10.508)	159	(9.312	, 10.971)	160	(9.262	, 10.390)
161	(9.584	, 10.838)	162	(9.621	, 11.306)	163	(9.114	, 10.142)	164	(8.680	, 10.564)	165	(9.249	, 10.030)
166	(9.518	, 10.579)	167	(9.740	, 10.844)	168	(9.183	, 10.795)	169	(8.686	, 10.580)	170	(9.442	, 10.740)
171	(9.129	, 11.088)	172	(9.542	, 11.062)	173	(9.201	, 10.730)	174	(9.485	, 11.251)	175	(8.640	, 10.463)
176	(9.382	, 10.690)	177	(9.241	, 10.585)	178	(9.495	, 10.810)	179	(9.859	, 11.136)	180	(9.356	, 10.975)
181	(9.385	, 10.516)	182	(8.884	, 10.759)	183	(9.011	, 10.829)	184	(9.401	, 10.731)	185	(8.637	, 10.468)
186	(8.769	, 10.727)	187	(9.925	, 11.065)	188	(9.427	, 11.199)	189	(9.550	, 11.040)	190	(9.729	, 10.887)
191	(9.157	, 10.439)	192	(9.277	, 10.485)	193	(9.691	, 10.611)	194	(9.358	, 10.997)	195	(9.229	, 10.976)
196	(9.014	, 11.275)	197	(8.475	, 10.120)	198	(9.390	, 10.469)	199	(9.485	, 11.083)	200	(9.238	, 10.721)
201	(9.153	, 10.465)	202	(9.121	, 10.960)	203	(9.596	, 10.440)	204	(9.595	, 11.125)	205	(9.136	, 10.383)
206	(8.930	, 10.040)	207	(9.045	, 10.714)	208	(8.894	, 10.830)	209	(9.029	, 10.821)	210	(9.320	, 10.355)
211	(9.380	, 10.707)	212	(9.283	, 10.236)	213	(9.280	, 10.681)	214	(8.606	, 10.343)	215	(9.585	, 11.183)
216	(8.958	, 9.899)*	217	(9.511	, 10.891)	218	(9.733	, 10.805)	219	(9.037	, 10.317)	220	(9.777	, 10.910)
221	(9.312	, 10.209)	222	(9.349	, 10.646)	223	(9.252	, 10.943)	224	(9.682	, 11.676)	225	(8.773	, 10.697)
226	(9.202	, 10.266)	227	(9.756	, 10.685)	228	(9.544	, 10.478)	229	(9.523	, 10.876)	230	(9.444	, 10.866)
231	(9.316	, 10.725)	232	(9.534	, 11.093)	233	(9.455	, 10.602)	234	(8.962	, 10.305)	235	(9.374	, 10.571)
236	(9.138	, 10.986)	237	(9.778	, 11.217)	238	(9.597	, 11.095)	239	(8.985	, 10.318)	240	(8.915	, 10.549)
241	(9.830	, 10.732)	242	(9.810	, 10.923)	243	(8.951	, 10.315)	244	(10.451	, 11.345)*	245	(9.170	, 10.339)
246	(9.020	, 10.123)	247	(9.296	, 10.288)	248	(9.510	, 10.359)	249	(9.031	, 10.354)	250	(9.428	, 11.241)
251	(9.445	, 10.761)	252	(9.697	, 10.937)	253	(9.494	, 11.246)	254	(9.200	, 10.861)	255	(8.904	, 10.378)
256	(9.129	, 10.713)	257	(9.383	, 10.142)	258	(9.879	, 10.792)	259	(8.852	, 10.912)	260	(9.725	, 10.517)
261	(9.626	, 11.264)	262	(8.700	, 10.547)	263	(8.911	, 10.488)	264	(9.289	, 10.694)	265	(9.229	, 10.756)
266	(9.188	, 10.753)	267	(9.396	, 11.143)	268	(9.225	, 11.391)	269	(9.005	, 10.153)	270	(9.196	, 10.505)
271	(9.208	, 10.526)	272	(8.902	, 10.389)	273	(8.742	, 11.019)	274	(9.069	, 10.410)	275	(9.501	, 10.632)
276	(9.324	, 10.635)	277	(9.488	, 11.056)	278	(9.277	, 10.547)	279	(9.408	, 10.679)	280	(9.329	, 10.839)
281	(8.636	, 9.820)*	282	(9.646	, 10.639)	283	(9.403	, 10.742)	284	(9.216	, 10.454)	285	(8.598	, 9.849)*
286	(9.266	, 11.348)	287	(9.208	, 10.449)	288	(9.113	, 10.901)	289	(8.934	, 10.334)	290	(9.306	, 10.454)
291	(9.573	, 11.202)	292	(9.063	, 10.685)	293	(10.229	, 11.040)*	294	(9.254	, 11.018)	295	(9.137	, 10.709)
296	(8.815	, 10.211)	297	(9.007	, 10.592)	298	(8.787	, 10.315)	299	(9.260	, 10.962)	300	(9.319	, 10.876)
301	(8.995	, 10.614)	302	(9.104	, 10.095)	303	(9.306	, 10.346)	304	(9.239	, 10.968)	305	(9.113	, 9.934)*
306	(9.390	, 10.522)	307	(9.639	, 10.848)	308	(9.209	, 10.601)	309	(8.866	, 11.200)	310	(9.597	, 11.396)
311	(8.391	, 10.287)	312	(8.964	, 10.791)	313	(9.645	, 10.769)	314	(9.373	, 10.817)	315	(9.420	, 10.751)
316	(9.312	, 11.081)	317	(8.901	, 10.141)	318	(9.730	, 10.525)	319	(9.636	, 11.284)	320	(9.291	, 10.781)
321	(9.723	, 10.775)	322	(9.249	, 10.688)	323	(9.113	, 10.160)	324	(9.109	, 10.434)	325	(9.329	, 11.007)
326	(8.959	, 10.226)	327	(9.664	, 10.516)	328	(9.856	, 11.101)	329	(9.345	, 10.956)	330	(8.908	, 10.843)
331	(9.693	, 11.499)	332	(9.423	, 11.238)	333	(8.978	, 10.768)	334	(9.534	, 11.050)	335	(9.576	, 11.082)
336	(9.371	, 10.638)	337	(8.950	, 11.271)	338	(9.276	, 10.557)	339	(9.310	, 10.619)	340	(9.073	, 10.076)
341	(9.281	, 10.795)	342	(9.744	, 10.505)	343	(9.542	, 10.813)	344	(8.913	, 10.316)	345	(9.414	, 11.246)
346	(9.661	, 10.722)	347	(8.724	, 10.361)	348	(9.224	, 11.179)	349	(9.354	, 10.569)	350	(9.318	, 10.665)
351	(9.174	, 10.895)	352	(9.615	, 11.003)	353	(9.121	, 10.696)	354	(9.517	, 10.884)	355	(8.730	, 10.328)
356	(8.942	, 10.684)	357	(8.890	, 10.399)	358	(9.640	, 11.146)	359	(9.349	, 11.253)	360	(9.522	, 10.452)
361	(9.136	, 10.555)	362	(8.923	, 10.764)	363	(9.737	, 11.513)	364	(9.159	, 10.257)	365	(9.736	, 11.236)
366	(9.156	, 10.699)	367	(9.515	, 11.037)	368	(9.175	, 10.724)	369	(9.475	, 10.408)	370	(9.108	, 10.933)
371	(9.135	, 10.924)	372	(9.144	, 10.631)	373	(8.854	, 10.319)	374	(8.680	, 10.513)	375	(9.529	, 10.391)

(continued)

TABLE 8.2 (continued)

SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL
376	9.886	11.506	377	9.633	10.961	378	9.252	11.411	379	9.218	10.469	380	8.648	10.653
381	9.193	10.900	382	9.331	10.868	383	9.072	10.634	384	8.974	10.078	385	9.170	10.652
386	9.063	10.490	387	9.047	10.484	388	9.202	10.194	389	9.475	10.655	390	9.553	10.720
391	9.396	11.115	392	9.236	10.644	393	8.785	10.183	394	9.145	10.765	395	9.340	11.055
396	9.123	10.534	397	9.828	11.058	398	9.386	10.229	399	9.234	10.500	400	9.072	10.405
401	9.680	10.752	402	9.487	10.849	403	9.539	11.137	404	9.795	11.293	405	9.566	10.792
406	8.983	10.842	407	9.410	10.964	408	9.892	10.949	409	9.097	11.117	410	9.229	11.201
411	9.451	10.924	412	9.530	10.756	413	9.328	11.021	414	9.512	10.590	415	9.027	10.805
416	8.982	10.591	417	9.059	10.856	418	8.971	10.620	419	9.236	10.456	420	9.085	10.768
421	8.900	10.358	422	9.604	11.044	423	10.265	11.443*	424	9.101	10.972	425	9.229	10.903
426	9.092	10.530	427	8.971	10.457	428	10.116	11.071*	429	9.579	11.107	430	9.066	10.596
431	8.892	10.710	432	9.684	11.258	433	8.919	10.350	434	9.226	11.093	435	9.012	10.969
436	8.582	10.107	437	9.106	9.925*	438	8.820	10.324	439	9.031	10.282	440	9.206	10.572
441	9.473	10.449	442	9.075	10.210	443	9.500	11.252	444	9.513	10.446	445	8.878	10.616
446	8.818	10.675	447	9.399	11.045	448	8.961	10.221	449	9.866	10.829	450	9.655	10.463
451	9.835	10.905	452	9.397	10.600	453	9.073	10.202	454	8.961	10.150	455	9.322	10.270
456	9.596	10.959	457	9.479	11.106	458	9.978	11.042	459	9.006	10.141	460	8.759	10.644
461	9.574	11.400	462	9.186	10.611	463	9.646	11.112	464	9.637	10.857	465	8.907	10.885
466	9.211	10.779	467	9.169	10.604	468	9.094	10.083	469	9.587	10.604	470	9.661	11.289
471	9.798	11.143	472	9.358	11.244	473	8.485	10.234	474	8.755	10.473	475	9.358	11.257
476	9.325	10.594	477	8.857	10.668	478	9.487	10.684	479	9.044	10.299	480	9.382	10.697
481	9.298	10.465	482	9.594	10.939	483	9.297	10.780	484	8.896	10.333	485	9.074	9.904*
486	9.284	10.777	487	9.125	10.438	488	9.081	10.442	489	9.242	10.483	490	9.420	10.687
491	9.284	10.471	492	9.462	10.845	493	9.274	10.676	494	8.961	10.458	495	9.796	11.205
496	9.080	10.854	497	9.399	10.606	498	8.593	10.554	499	9.527	10.925	500	9.340	10.413
501	8.907	10.531	502	9.243	10.401	503	9.270	10.787	504	9.331	11.183	505	9.200	10.602
506	9.422	10.835	507	9.279	10.687	508	8.884	10.683	509	9.190	10.255	510	8.733	10.404
511	9.441	10.889	512	9.070	10.392	513	8.946	9.772*	514	8.799	10.236	515	9.070	10.830
516	8.984	10.365	517	9.183	10.387	518	9.236	10.863	519	9.826	11.197	520	8.906	10.672
521	9.755	10.996	522	9.400	10.880	523	9.374	11.276	524	9.288	11.211	525	9.412	11.000
526	9.028	10.300	527	8.647	10.285	528	9.190	10.072	529	9.238	10.465	530	10.003	11.190*
531	9.644	10.974	532	9.679	10.738	533	8.559	10.014	534	9.895	11.131	535	9.653	11.007
536	9.769	11.144	537	8.837	10.136	538	9.939	10.836	539	9.553	10.853	540	9.351	10.552
541	9.532	11.320	542	9.262	10.728	543	8.864	11.341	544	9.052	10.482	545	9.551	10.610
546	9.564	11.060	547	9.699	10.912	548	8.915	10.042	549	8.801	10.648	550	9.111	9.913*
551	9.811	11.558	552	8.593	10.128	553	8.612	9.854*	554	9.555	10.986	555	9.567	10.666
556	9.399	10.979	557	9.168	11.195	558	9.270	11.456	559	9.197	10.554	560	8.985	10.492
561	9.067	10.622	562	8.843	10.484	563	9.346	11.128	564	8.692	10.475	565	9.413	10.583
566	9.416	10.678	567	9.451	11.099	568	9.339	10.862	569	8.933	10.698	570	9.212	10.368
571	8.894	10.438	572	9.161	10.964	573	9.841	11.306	574	8.990	10.541	575	8.530	10.038
576	9.687	11.005	577	9.131	10.759	578	9.167	10.734	579	9.301	10.507	580	9.379	10.872
581	9.298	10.917	582	8.407	10.030	583	9.080	10.523	584	9.043	10.509	585	9.636	11.032
586	9.562	10.527	587	9.224	10.279	588	9.439	10.305	589	9.320	10.482	590	9.363	11.047
591	9.136	10.521	592	9.059	10.320	593	8.686	10.278	594	9.280	10.267	595	9.251	10.964
596	8.693	10.114	597	8.712	10.394	598	9.340	10.710	599	8.244	9.684*	600	9.583	10.992
601	9.232	10.346	602	9.014	10.458	603	9.861	11.485	604	9.139	11.097	605	9.060	10.269
606	9.712	11.648	607	8.963	10.055	608	8.991	10.548	609	9.540	10.769	610	9.822	11.243
611	9.338	10.357	612	8.632	10.201	613	9.371	10.898	614	9.155	10.582	615	8.806	10.919
616	9.182	10.488	617	9.403	10.755	618	9.199	10.527	619	9.016	10.844	620	9.321	11.077
621	9.475	10.651	622	9.481	10.701	623	9.661	10.590	624	9.358	10.812	625	9.046	10.679
626	9.948	10.907	627	8.649	9.996*	628	9.201	10.332	629	9.195	10.908	630	9.460	10.435
631	9.222	10.772	632	9.757	10.880	633	9.926	10.885	634	9.027	10.425	635	8.436	10.011
636	9.160	10.474	637	9.723	11.075	638	8.597	10.879	639	10.024	10.931*	640	8.475	10.397
641	8.712	10.702	642	10.038	11.678*	643	9.706	10.962	644	9.028	10.275	645	9.395	10.414
646	9.283	10.641	647	8.628	10.107	648	9.456	10.820	649	9.999	11.336	650	8.587	10.063
651	9.616	11.090	652	9.403	10.537	653	10.263	11.277*	654	9.325	10.717	655	9.795	10.737
656	9.669	11.778	657	9.739	10.636	658	9.285	10.965	659	9.210	10.552	660	9.384	10.962
661	9.041	10.347	662	9.380	10.846	663	9.950	10.597	664	9.602	10.584	665	9.092	10.439
666	9.475	10.844	667	9.192	10.844	668	10.134	11.385*	669	8.523	10.431	670	9.657	11.222
671	8.710	10.470	672	8.854	10.039	673	8.833	10.174	674	9.500	10.956	675	9.546	10.782
676	9.115	10.545	677	9.005	10.434	678	9.783	11.217	679	9.384	10.647	680	9.783	11.586
681	9.160	10.040	682	8.822	10.238	683	9.374	11.064	684	8.895	10.274	685	8.986	10.854
686	8.665	10.523	687	8.630	10.270	688	9.914	10.898	689	8.787	10.323	690	9.483	10.850
691	9.369	10.797	692	9.271	10.776	693	8.715	9.846*	694	8.764	10.481	695	8.934	10.053
696	9.280	10.143	697	8.354	9.985*	698	9.599	11.275	699	8.488	10.224	700	9.278	10.213
701	9.247	10.552	702	9.043	10.327	703	9.578	10.914	704	8.815	10.387	705	8.786	9.901*
706	9.029	10.654	707	9.731	10.545	708	9.143	10.562	709	9.254	10.501	710	9.045	10.718
711	9.552	10.664	712	10.039	11.511*	713	9.670	10.482	714	9.491	10.669	715	9.589	10.869
716	8.900	10.986	717	9.557	10.872	718	8.845	10.578	719	9.316	11.041	720	9.420	10.645
721	9.839	10.896	722	9.264	10.440	723	10.020	11.239*	724	9.235	11.200	725	9.194	11.026
726	9.497	10.654	727	9.212	10.904	728	9.328	10.619	729	9.563	10.672	730	9.646	11.070
731	9.082	10.194	732	9.171	10.781	733	9.016	10.914	734	8.944	10.604	735	9.028	10.664
736	9.074	9.924*	737	9.050	10.604	738	8.641	10.155	739	10.052	11.660*	740	9.806	10.831
741	9.051	10.307	742	9.180	10.632	743	9.181	10.306	744	9.751	10.867	745	9.162	10.229
746	9.088	10.658	747	9.184	9.982*	748	8.697	10.393	749	9.292	10.612	750	9.331	10.615

(continued)

TABLE 8.2 (continued)

SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL	SAMPLE	LCL	UCL
751	9.138	10.361	752	9.604	11.201	753	8.921	10.326	754	8.943	10.219	755	9.222	10.216
756	9.530	10.981	757	9.248	10.720	758	9.646	10.700	759	8.895	10.036	760	9.618	10.742
761	9.290	10.929	762	9.504	10.942	763	9.053	10.474	764	9.754	10.946	765	9.198	10.351
766	9.146	10.468	767	9.180	10.399	768	9.177	10.305	769	9.130	10.580	770	9.960	11.238
771	8.694	10.742	772	9.463	10.594	773	9.348	11.102	774	9.224	10.726	775	9.229	11.217
776	9.082	10.291	777	9.352	10.366	778	9.604	11.415	779	8.366	9.595*	780	9.622	11.160
781	10.024	11.043*	782	9.247	10.508	783	10.053	11.078*	784	8.640	10.792	785	9.278	10.767
786	9.486	11.021	787	9.215	10.090	788	9.647	11.227	789	8.559	10.444	790	8.498	10.529
791	9.867	10.967	792	9.095	10.364	793	8.815	10.275	794	8.648	10.216	795	9.859	11.008
796	8.862	10.274	797	9.218	10.439	798	9.299	10.668	799	9.015	10.139	800	8.873	10.581
801	9.502	11.150	802	9.598	11.290	803	9.843	11.204	804	9.377	10.387	805	9.388	10.640
806	8.571	9.804*	807	9.369	10.523	808	8.432	10.584	809	9.305	10.629	810	9.263	10.718
811	9.253	9.991*	812	9.060	10.301	813	9.323	11.395	814	9.261	10.791	815	9.655	10.995
816	9.425	10.722	817	9.166	10.566	818	9.511	10.630	819	9.185	10.674	820	9.612	10.713
821	9.795	11.330	822	9.491	11.104	823	9.133	10.491	824	9.459	10.787	825	9.197	10.451
826	9.276	10.493	827	9.528	10.964	828	8.961	10.897	829	8.814	10.037	830	9.439	10.769
831	9.430	10.786	832	10.506	11.206*	833	9.033	10.450	834	9.641	11.223	835	9.383	10.561
836	9.046	10.512	837	9.281	10.414	838	8.707	10.181	839	9.870	11.157	840	9.321	10.426
841	9.058	10.378	842	9.480	11.349	843	8.897	10.717	844	9.611	10.216	845	8.722	9.934*
846	9.350	10.886	847	9.411	10.844	848	8.984	10.566	849	8.968	10.537	850	9.081	10.380
851	9.054	10.647	852	8.873	9.791*	853	10.021	11.515*	854	9.554	11.099	855	8.524	10.378
856	8.781	10.739	857	9.385	10.910	858	8.945	10.416	859	9.183	10.624	860	9.462	10.607
861	9.099	10.434	862	9.331	10.806	863	9.771	10.995	864	9.327	10.731	865	8.963	10.438
866	9.259	11.270	867	9.211	10.519	868	9.821	11.420	869	9.335	10.513	870	9.078	10.210
871	10.080	10.769*	872	9.375	10.590	873	8.535	9.890*	874	9.414	10.751	875	8.877	9.994*
876	9.587	10.795	877	9.121	10.960	878	9.486	10.822	879	10.293	11.456*	880	9.514	10.926
881	9.058	10.909	882	8.990	10.079	883	9.580	11.051	884	9.185	10.505	885	8.812	10.421
886	9.301	10.096	887	9.194	10.273	888	9.278	11.004	889	8.658	10.170	890	9.367	10.074
891	8.630	10.978	892	9.842	11.724	893	9.504	10.998	894	9.287	10.866	895	9.234	10.570
896	9.986	10.907	897	9.758	11.048	898	9.687	10.993	899	9.381	10.822	900	9.518	10.493
901	9.114	10.575	902	8.869	10.508	903	9.363	10.595	904	9.252	10.618	905	9.784	10.718
906	9.147	10.241	907	9.448	10.569	908	9.330	10.693	909	9.096	10.499	910	9.780	10.687
911	9.047	10.283	912	9.036	10.381	913	9.655	11.262	914	9.400	9.964*	915	9.368	11.079
916	9.456	10.747	917	8.768	10.250	918	9.270	10.158	919	9.419	10.101	920	9.159	10.773
921	9.736	11.113	922	9.445	10.763	923	9.423	10.674	924	8.777	10.774	925	9.155	10.204
926	9.087	10.368	927	9.079	10.049	928	9.245	10.969	929	9.096	10.402	930	9.106	10.613
931	9.603	10.961	932	9.511	11.157	933	9.650	10.768	934	9.149	10.002	935	10.015	11.540*
936	9.676	10.788	937	9.700	11.167	938	9.615	11.085	939	9.555	10.694	940	9.382	10.570
941	8.498	9.897*	942	9.216	10.700	943	9.140	10.459	944	9.543	10.540	945	8.824	10.129
946	9.523	10.824	947	9.147	10.406	948	9.068	10.536	949	9.119	10.172	950	8.709	10.638
951	9.850	11.410	952	9.729	10.940	953	9.067	10.090	954	9.599	11.064	955	9.753	10.920
956	9.501	10.523	957	9.598	10.705	958	9.220	10.626	959	8.391	9.950*	960	9.629	10.594
961	9.105	10.574	962	9.504	10.543	963	9.137	10.475	964	9.303	10.910	965	9.563	10.871
966	9.161	10.453	967	9.487	10.752	968	9.531	11.014	969	8.920	10.599	970	9.058	10.440
971	9.409	10.760	972	8.981	10.788	973	9.097	10.186	974	8.674	10.776	975	9.010	10.745
976	8.714	10.521	977	9.176	10.301	978	9.263	10.555	979	8.700	10.244	980	9.334	10.959
981	9.577	10.873	982	9.383	10.970	983	9.462	10.826	984	9.367	10.726	985	8.657	10.496
986	9.436	10.970	987	9.532	11.605	988	9.309	10.876	989	9.536	10.799	990	9.827	10.698
991	8.834	9.807*	992	8.672	10.247	993	8.974	10.373	994	9.169	10.891	995	8.704	10.044
996	9.713	10.932	997	9.169	10.769	998	9.595	10.769	999	9.648	10.762	1000	9.029	10.684

Theoretical Interpretation of the Confidence Coefficient $(1 - \alpha)$

If we were to repeatedly collect a sample of size n from the population and construct a $(1 - \alpha)100\%$ confidence interval for each sample, then we expect $(1 - \alpha)100\%$ of the intervals to enclose the true parameter value.

Confidence intervals for population parameters other than the population mean can be derived using the pivotal method outlined in this section. The estimators and pivotal statistics for many of these parameters are well known. In Sections 8.5–8.11, we give the confidence interval formulas for several population parameters that are commonly encountered in practice.

EXERCISES

8.14 Use Table 7 of Appendix II to determine the values of $t_{\alpha/2}$ that would be used in the construction of a confidence interval for a population mean for each of the following combinations of confidence coefficient and sample size:

a. Confidence coefficient .99, $n = 18$
b. Confidence coefficient .95, $n = 10$
c. Confidence coefficient .90, $n = 15$

8.15 It can be shown (proof omitted) that as the sample size n increases, the t distribution tends to normality and the value t_a, such that $P(t > t_a) = a$, approaches the value z_a, such that $P(z > z_a) = a$. Use Table 7 of Appendix II to verify that as the sample size n gets infinitely large, $t_{.05} = z_{.05}$, $t_{.025} = z_{.025}$, and $t_{.01} = z_{.01}$.

8.16 Let y be the number of successes in a binomial experiment with n trials and probability of success p. Assuming that n is large, use the sample proportion of successes $\hat{p} = y/n$ to form a confidence interval for p with confidence coefficient $(1 - \alpha)$. [*Hint:* Start with the pivotal statistic

$$z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

and use the fact (proof omitted) that for large n, z is approximately a standard normal random variable.]

8.17 Let $y_1, y_2, \ldots, y_n$ be a random sample from a Poisson distribution with mean λ. Suppose we use $\bar{y}$ as an estimator of λ. Derive a $(1 - \alpha)100\%$ confidence interval for λ. [*Hint:* Start with the pivotal statistic

$$z = \frac{\bar{y} - \lambda}{\sqrt{\lambda/n}}$$

and show that for large samples, z is approximately a standard normal random variable. Then substitute $\bar{y}$ for λ in the denominator (why can you do this?) and follow the pivotal method of Example 8.6.]

8.18 Let $y_1, y_2, \ldots, y_n$ be a random sample of n observations from an exponential distribution with mean β. Derive a large-sample confidence interval for β. [*Hint:* Start with the pivotal statistic

$$z = \frac{\bar{y} - \beta}{\beta/\sqrt{n}}$$

and show that for large samples, z is approximately a standard normal random variable. Then substitute $\bar{y}$ for β in the denominator (why can you do this?) and follow the pivotal method of Example 8.6.]

OPTIONAL EXERCISES

8.19 Let $\bar{y}_1$ and s_1^2 be the sample mean and sample variance, respectively, of n_1 observations randomly selected from a population with mean μ_1 and variance σ_1^2. Similarly, define $\bar{y}_2$ and s_2^2 for an independent random sample of n_2 observations from a population with mean μ_2 and σ_2^2. Derive a large-sample confidence interval for $(\mu_1 - \mu_2)$. [*Hint:* Start with the pivotal statistic

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

and show that for large samples, z is approximately a standard normal random variable. Substitute s_1^2 for σ_1^2 and s_2^2 for σ_2^2 (why can you do this?) and follow the pivotal method of Example 8.6.]

8.20 Let $(\bar{y}_1, s_1^2)$ and $(\bar{y}_2, s_2^2)$ be the means and variances of two independent random samples of sizes n_1 and n_2, respectively, selected from normal populations with different means, μ_1 and μ_2, but with a common variance, σ^2.

a. Show that $E(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$.

b. Show that

$$V(\bar{y}_1 - \bar{y}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

c. Explain why

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

is a standard normal random variable.

8.21 Refer to Exercise 8.20. According to Theorem 7.4,

$$\chi_1^2 = \frac{(n_1 - 1)s_1^2}{\sigma^2} \quad \text{and} \quad \chi_2^2 = \frac{(n_2 - 1)s_2^2}{\sigma^2}$$

are independent chi-square random variables with $(n_1 - 1)$ and $(n_2 - 1)$ df, respectively. Show that

$$\chi^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{\sigma^2}$$

is a chi-square random variable with $(n_1 + n_2 - 2)$ df.

8.22 Refer to Exercises 8.20 and 8.21. The pooled estimator of the common variance σ^2 is given by

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Show that

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

has a Student's t distribution with $(n_1 + n_2 - 2)$ df. [*Hint:* Recall that $t = z/\sqrt{\chi^2/\nu}$ has a Student's t distribution with ν df and use the results of Exercises 8.20c and 8.21.]

8.23 Use the pivotal statistic t given in Exercise 8.22 to derive a $(1 - \alpha)100\%$ small-sample confidence interval for $(\mu_1 - \mu_2)$.

8.5 Estimation of a Population Mean

From our discussions in Section 8.3, we already know that a useful point estimate of the population mean μ is $\bar{y}$, the sample mean. According to the central limit theorem (Theorem 7.2), we also know that for sufficiently large n, the sampling distribution of the sample mean $\bar{y}$ is approximately normal with $E(\bar{y}) = \mu$ and $V(\bar{y}) = \sigma^2/n$. The fact that $E(\bar{y}) = \mu$ implies that $\bar{y}$ is an unbiased estimator of μ. Furthermore, it can be shown (proof omitted) that $\bar{y}$ has the smallest variance among all unbiased estimators of μ. Hence, $\bar{y}$ is the MVUE for μ. Therefore, it is not surprising that $\bar{y}$ is considered the best estimator of μ.

Since $\bar{y}$ is approximately normal for large n, we can apply Theorem 8.2 to construct a large-sample $(1 - \alpha)100\%$ confidence interval for μ. Substituting $\hat{\theta} = \bar{y}$ and $\sigma_{\hat{\theta}} = \sigma/\sqrt{n}$ into the confidence interval formula given in Theorem 8.2, we obtain the formula given in the following box.

Large-Sample $(1 - \alpha)100\%$ Confidence Interval for the Population Mean, μ

$$\bar{y} \pm z_{\alpha/2}\sigma_{\bar{y}} = \bar{y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \bar{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

where $z_{\alpha/2}$ is the z value that locates an area of $\alpha/2$ to its right, σ is the standard deviation of the population from which the sample was selected, n is the sample size, and $\bar{y}$ is the value of the sample mean.

[*Note:* When the value of σ is unknown (as will usually be the case), the sample standard deviation s may be used to approximate σ in the formula for the confidence interval. The approximation is generally quite satisfactory when $n \geq 30$.]

Assumptions: None (since the central limit theorem guarantees that $\bar{y}$ is approximately normal regardless of the distribution of the sampled population)

Note: The value of the sample size n required for the sampling distribution of $\bar{y}$ to be approximately normal will vary depending on the shape (distribution) of the target population (see Examples 7.6 and 7.7). As a general rule of thumb, a sample size n of 30 or more will be considered *sufficiently large* for the central limit theorem to apply.

EXAMPLE 8.8

Suppose a regional computer center wants to evaluate the performance of its disk memory system. One measure of performance is the average time between failures of its disk drive. To estimate this value, the center recorded the time between failures for a random sample of 45 disk-drive failures. The following sample statistics were computed:

$$\bar{y} = 1{,}762 \text{ hours} \qquad s = 215 \text{ hours}$$

a. Estimate the true mean time between failures with a 90% confidence interval.

b. If the disk memory system is running properly, the true mean time between failures will exceed 1,700 hours. Based on the interval, part **a**, what can you infer about the disk memory system?

Solution

a. For a confidence coefficient of $1 - \alpha = .90$, we have $\alpha = .10$ and $\alpha/2 = .05$; therefore, a 90% confidence interval for μ is given by

$$\begin{aligned} \bar{y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) &= \bar{y} \pm z_{.05}\left(\frac{\sigma}{\sqrt{n}}\right) \\ &\approx \bar{y} \pm z_{.05}\left(\frac{s}{\sqrt{n}}\right) \\ &= 1{,}762 \pm z_{.05}\left(\frac{215}{\sqrt{45}}\right) \end{aligned}$$

where $z_{.05}$ is the z value corresponding to an upper-tail area of .05. From Table 4 of Appendix II, $z_{.05} = 1.645$. Then the desired interval is

$$\begin{aligned} 1{,}762 \pm z_{.05}\left(\frac{215}{\sqrt{45}}\right) &= 1{,}762 \pm 1.645\left(\frac{215}{\sqrt{45}}\right) \\ &= 1{,}762 \pm 52.7 \end{aligned}$$

or 1,709.3 to 1,814.7 hours. We are 90% confident that the interval (1,709.3, 1,814.7) encloses μ, the true mean time between disk failures.

b. Since all values within the 90% confidence interval exceed 1,700 hours, we can infer (with 90% confidence) that the disk memory system is running properly.

Sometimes, time or cost limitations may restrict the number of sample observations that may be obtained for estimating μ. In the case of small samples, (say, $n < 30$), the following two problems arise:

1. Since the central limit theorem applies only to large samples, we are not able to assume that the sampling distribution of $\bar{y}$ is approximately normal. Therefore, we cannot apply Theorem 8.2. For small samples, the sampling distribution of $\bar{y}$ depends on the particular form of the relative frequency distribution of the population being sampled.

2. The sample standard deviation s may not be a satisfactory approximation to the population standard deviation σ if the sample size is small.

Fortunately, we may proceed with estimation techniques based on small samples if we can assume that the population from which the sample is selected has an approximate normal distribution. If this assumption is valid, then we can use the procedure of Example 8.6 to construct a confidence interval for μ. The general form of a small-sample confidence interval for μ, based on the Student's t distribution, is as shown in the next box.

Small-Sample $(1 - \alpha)100\%$ Confidence Interval for the Population Mean, μ

$$\bar{y} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

where the distribution of t is based on $(n - 1)$ degrees of freedom.

Assumption: The population from which the sample is selected has an approximate normal distribution.

EXAMPLE 8.9

The Geothermal Loop Experimental Facility, located in the Salton Sea in southern California, is a U.S. Department of Energy operation for studying the feasibility of generating electricity from the hot, highly saline water of the Salton Sea. Operating experience has shown that these brines leave silica scale deposits on metallic plant piping, causing excessive plant outages. Jacobsen et al. (*Journal of Testing and Evaluation*, Mar. 1981) have found that scaling can be reduced somewhat by adding chemical solutions to the brine. In one screening experiment, each of five antiscalants was added to an aliquot of brine, and the solutions were filtered. A silica determination (parts per million of silicon dioxide) was made on each filtered sample after a holding time of 24 hours, with the following results:

229 255 280 203 229

Estimate the mean amount of silicon dioxide present in the five antiscalant solutions. Use a 99% confidence interval.

Solution

The first step in constructing the confidence interval is to compute the mean, $\bar{y}$, and standard deviation, s, of the sample of five silicon dioxide amounts. These values, $\bar{y} = 239.2$ and $s = 29.3$, are shaded in the MINITAB printout, Figure 8.8.

FIGURE 8.8 ▶
MINITAB descriptive statistics for Example 8.9

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
PPM	5	239.2	229.0	239.2	29.3	13.1

	MIN	MAX	Q1	Q3
PPM	203.0	280.0	216.0	267.5

For a confidence coefficient of $1 - \alpha = .99$, we have $\alpha = .01$ and $\alpha/2 = .005$. Since the sample size is small ($n = 5$), our estimation technique requires the assumption that the amount of silicon dioxide present in an antiscalant solution has an approximately normal distribution (i.e., the sample of 5 silicon amounts is selected from a normal population).

Substituting the values for $\bar{y}$, s, and n into the formula for a small-sample confidence interval for μ, we obtain

$$\bar{y} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = \bar{y} \pm t_{.005}\left(\frac{s}{\sqrt{n}}\right)$$
$$= 239.2 \pm t_{.005}\left(\frac{29.3}{\sqrt{5}}\right)$$

where $t_{.005}$ is the value corresponding to an upper-tail area of .005 in the Student's t distribution based on $(n - 1) = 4$ degrees of freedom. From Table 7 of Appendix II, the required t value is $t_{.005} = 4.604$. Substitution of this value yields

$$239.2 \pm t_{.005}\left(\frac{29.3}{\sqrt{5}}\right) = 239.2 \pm (4.604)\left(\frac{29.3}{\sqrt{5}}\right)$$
$$= 239.2 \pm 60.3$$

or, 178.9 to 299.5 ppm. Thus, if the distribution of silicon dioxide amounts is approximately normal, we can be 99% confident that the interval (178.9, 299.5) encloses μ, the true mean amount of silicon dioxide present in an antiscalant solution.

The 99% confidence interval can also be obtained with a statistical software package. Figure 8.9 shows a MINITAB printout of the analysis. You can see that the computer-generated interval (shaded in Figure 8.9) is identical to the one we calculated.

FIGURE 8.9 ►
MINITAB confidence interval for Example 8.9

```
TEST OF MU = 300.000 VS MU N.E. 300.000

         N      MEAN     STDEV   SE MEAN   99.0 PERCENT C.I.
ppm      5     239.2      29.3      13.1   (   178.9,    299.5)
```

Before we conclude this section, two comments are necessary. The first concerns the assumption that the sampled population is normally distributed. In the real world, we rarely know whether a sampled population has an exact normal distribution. However, empirical studies indicate that moderate departures from this assumption do not seriously affect the confidence coefficients for small-sample confidence intervals. For example, if the population of silicon dioxide amounts for the antiscalant solutions of Example 8.9 has a distribution that is mound-shaped but nonnormal, it is likely that the actual confidence coefficient for the 99% confidence interval will be close to .99—at least close enough to be of practical use. As a consequence, the small-sample confidence interval given in the box is frequently used by experimenters when estimating the population mean of a nonnormal distribution as long as the distribution is mound-shaped and only moderately skewed. For populations that depart greatly from normality, other estimation techniques (such as robust estimation) or methods that are distribution-free (called **nonparametrics**) are recommended. Nonparametric statistics are the topic of Chapter 15.

The second comment focuses on whether σ is known or unknown. We have shown (Example 7.7) that when σ is known and the sampled population is normally distributed, the sampling distribution of $\bar{y}$ is normal regardless of the size of the sample. That is, if you know the value of σ and you know that the sample comes from a normal population, then you can use the z distribution rather than the t distribution to form confidence intervals. In reality, however, σ is rarely (if ever) known. Consequently, you will always be using s in place of σ in the confidence interval formulas, and the sampling distribution of $\bar{y}$ will be a t distribution. This is why the formula for a large-sample confidence interval given earlier in this section is only approximate; in the large-sample case, $t \approx z$. Many statistical software packages give the results for *exact* confidence intervals when σ is unknown; thus, these results are based on the t distribution. For practical reasons, however, we will continue to distinguish between z and t confidence intervals based on whether the sample size is large or small.

EXERCISES

8.24 Chemical engineers at the University of Murcia (Spain) conducted a series of experiments to determine the most effective membrane to use in a passive sampler (*Environmental Science & Technology*, Vol. 27, 1993). The effectiveness of a passive sampler was measured by the sampling rate, recorded in cubic centimeters per minute. In one experiment, six passive samplers were positioned with their faces parallel to the air flow and with an air velocity of 90 centimeters per second. After 6 hours, the sampling rate of each was determined. Based on the results, a 95% confidence interval for the mean sampling rate was calculated to be (49.66, 51.48).

a. What is the confidence coefficient for the interval?
b. Give a theoretical interpretation of the confidence coefficient, part **a**.
c. Give a practical interpretation of the confidence interval.
d. What assumptions, if any, are required for the interval to yield valid inferences?

8.25 The theoretical relationship between heat flux and temperature gradient for homogeneous materials is well known and described by a Fourier equation. However, this relationship does not hold for nonhomogeneous materials such as porous-capillary bodies, cellular systems, suspensions, and pastes. An experiment was conducted to estimate the mean thermal relaxation time (defined as the mean time needed for accumulating the thermal energy required for propagative transfer of heat) for several nonhomogeneous materials (*Journal of Heat Transfer*, Aug. 1990). A 95% confidence interval for the mean thermal relaxation time of sand was found to be 20.0 ± 6.4 seconds.

a. Give a practical interpretation of the 95% confidence interval.
b. Give a theoretical interpretation of the 95% confidence interval.

8.26 Unusual rocks at "The Seven Islands," located along the lower St. Lawrence River in Canada, have attracted geologists to the area for over a century. A major geological survey of "The Seven Islands" was recently completed for the purpose of developing a three-dimensional gravity model of the area (*Canadian Journal of Earth Sciences*, Vol. 27, 1990). One of the keys to an objective model is obtaining an accurate estimate of the rock densities. Based on samples of several varieties of rock, the following information on rock density (grams per cubic centimeter) was obtained.

Type of Rock	*Sample Size*	*Mean Density*	*Standard Deviation*
Late gabbro	36	3.04	.13
Massive gabbro	148	2.83	.11
Cumberlandite	135	3.05	.31

Source: Loncarevic, B. D., Feninger, T., and Lefebvre, D. "The Sept-Îles layered mafic intrusion: Geophysical expression." *Canadian Journal of Earth Sciences*, Vol. 27, Aug. 1990, p. 505.

a. For each rock type, estimate the mean density with a 90% confidence interval.
b. Interpret the intervals, part **a**.

8.27 An evaluation of trace metal chemistry and cycling in an acidic Adirondack lake was reported in *Environmental Science & Technology* (Dec. 1985). Twenty-four (24) water samples were collected from Darts Lake, New York, and analyzed for concentration of both lead and aluminum particulates.

a. The lead concentration measurements had a mean of 9.9 nmol/l and a standard deviation of 8.4 nmol/l. Calculate a 99% confidence interval for the true mean lead concentration in water samples collected from Darts Lake.
b. The aluminum concentration measurements had a mean of 6.7 nmol/l and a standard deviation of 10.8 nmol/l. Calculate a 99% confidence interval for the true mean aluminum concentration in water samples collected from Darts Lake.
c. What assumptions are necessary for the intervals of parts **a** and **b** to be valid?

8.28 According to one study, "The majority of people who die from fire and smoke in compartmented fire-resistive buildings—the type used for hotels, motels, apartments, and other health care facilities—die in the attempt to evacuate" (*Risk Management*, Feb. 1986). The accompanying data represent the numbers of victims who attempted to evacuate for a sample of 14 recent fires at compartmented fire-resistive buildings reported in the study.

Fire	*Died in Attempt to Evacuate*
Las Vegas Hilton (Las Vegas)	5
Inn on the Park (Toronto)	5
Westchase Hilton (Houston)	8
Holiday Inn (Cambridge, Ohio)	10
Conrad Hilton (Chicago)	4
Providence College (Providence)	8
Baptist Towers (Atlanta)	7
Howard Johnson (New Orleans)	5
Cornell University (Ithaca, New York)	9
Wesport Central Apartments (Kansas City, Missouri)	4
Orrington Hotel (Evanston, Illinois)	0
Hartford Hospital (Hartford, Connecticut)	16
Milford Plaza (New York)	0
MGM Grand (Las Vegas)	36

Source: Macdonald, J. N. "Is evacuation a fatal flaw in fire fighting philosophy?" *Risk Management*, Vol. 33, No. 2, Feb. 1986, p. 37.

a. State the assumption, in terms of the problem, that is required for a small-sample confidence interval technique to be valid.
b. Use the information in the accompanying MINITAB printout to construct a 98% confidence interval for the true mean number of victims per fire who die in an attempt to evacuate compartmented fire-resistive buildings.
c. Interpret the interval constructed in part b.

```
               N        MEAN      STDEV    SE MEAN      98.0 PERCENT C.I.
numdied       14        8.36       8.94       2.39    (     2.02,     14.69)
```

8.29 The *Journal of the American Medical Association* (Apr. 21, 1993) reported on the results of a National Health Interview Survey designed to determine the prevalence of smoking among U.S. adults. Over 40,000 adults responded to questions such as "Have you smoked at least 100 cigarettes in your lifetime?" and "Do you smoke cigarettes now?" Current smokers (over 11,000 adults in the survey) were also asked: "On the average, how many cigarettes do you now smoke a day?" The results yielded a mean of 20.0 cigarettes per day with an associated 95% confidence interval of (19.7, 20.3).
a. Interpret the 95% confidence interval.
b. State any assumptions about the target population of current cigarette smokers that must be satisfied for inferences derived from the interval to be valid.
c. A tobacco industry researcher claims that the mean number of cigarettes smoked per day by regular cigarette smokers is less than 15. Comment on this claim.

8.30 Tropical swarm-founding wasps, like ants and bees, rely on workers to raise their offspring. Interestingly, the workers of this species of wasp are mostly female, capable of producing offspring of their own. Instead, they rear the young of others in the brood. One possible explanation for this strange behavior is inbreeding, which increases relatedness among the wasps and makes it easier for the workers to pick out and aid their closest relatives. To test this theory, 197 swarm-founding wasps were captured in Venezuela, frozen at −70°C, and then subjected to a series of genetic tests (*Science*, Nov. 1988). The data were used to generate an inbreeding coefficient, x, for each wasp specimen, with the following results: $\bar{x} = .044$ and $s = .884$.
a. Construct a 90% confidence interval for the mean inbreeding coefficient of this species of wasp.
b. A coefficient of 0 implies that the wasp has no tendency to inbreed. Use the confidence interval, part a, to make an inference about the tendency for this species of wasp to inbreed.

8.31 The data for Exercise 2.57 are reproduced here. The numbers in the table represent the CPU solution times (in seconds) for 52 random polynomial 0–1 mathematical problems solved using a hybrid algorithm. A stem-and-leaf display and descriptive statistics for the data set are provided in the accompanying SAS printout. Use this information to estimate, with 95% confidence, the mean solution time for the hybrid algorithm. Interpret the result.

.045	.136	8.788	.079	3.985	1.267	.379	.327
.136	.130	.036	.136	.600	.209	.506	.064
.088	.194	.118	.258	4.170	.554	.412	.045
.361	.049	.070	1.639	.258	.670	.567	
.182	1.055	.091	.579	1.894	.291	.445	
.179	.336	.145	.394	1.070	.227	.258	
.182	.242	.209	.333	.912	3.046	3.888	

Source: Snyder, W. S., and Chrissis, J. W. "A hybrid algorithm for solving zero–one mathematical programming problems." *IIE Transactions*, Vol. 22, No. 2, June 1990, p. 166 (Table 1).

```
                         UNIVARIATE PROCEDURE

Variable=SOLTIME

                               Moments

               N                52  Sum Wgts          52
               Mean       0.812192  Sum           42.234
               Std Dev     1.50476  Variance    2.264303
               Skewness    3.65356  Kurtosis    15.73264
               USS        149.7816  CSS         115.4795
               CV         185.2714  Std Mean    0.208673
               T:Mean=0   3.892183  Prob>|T|      0.0003
               Sgn Rank        689  Prob>|S|      0.0001
               Num ^= 0         52
               W:Normal   0.530623  Prob<W           0.0

                           Quantiles(Def=5)

                100% Max     8.788       99%     8.788
                 75% Q3     0.5895       95%     3.985
                 50% Med    0.2745       90%     1.894
                 25% Q1      0.136       10%      0.07
                  0% Min     0.036        5%     0.045
                                          1%     0.036
                Range        8.752
                Q3-Q1       0.4535
                Mode         0.136

                               Extremes

                  Lowest    Obs     Highest    Obs
                   0.036(     22)    3.046(     49)
                   0.045(     50)    3.888(     35)
                   0.045(      1)    3.985(     33)
                   0.049(     51)     4.17(     43)
                   0.064(     45)    8.788(     17)

          Stem Leaf                                   #         Boxplot
             8 8                                      1            *
             8
             7
             7
             6
             6
             5
             5
             4
             4 02                                     2            *
             3 9                                      1            *
             3 0                                      1            *
             2
             2
             1 69                                     2            0
             1 113                                    3            |
             0 5666679                                7         +--+--+
             0 00001111111111122222222333333344444   35         *-----*
               ----+----+----+----+----+----+----+
```

8.6 Estimation of the Difference Between Two Population Means: Independent Samples

In Section 8.5, we learned how to estimate the parameter μ from a single population. We now proceed to a technique for using the information in two samples to estimate the difference between two population means, $(\mu_1 - \mu_2)$, when the samples are collected independently. For example, we may want to compare the mean starting salaries for college graduates with mechanical engineering and civil engineering degrees, or the mean operating costs of automobiles with rotary engines and standard engines, or the mean failure times of two electronic components. The technique to be presented is a straightforward extension of that used for estimation of a single population mean.

Suppose we select independent random samples of sizes n_1 and n_2 from populations with means μ_1 and μ_2, respectively. Intuitively, we want to use the difference between the sample means, $(\bar{y}_1 - \bar{y}_2)$, to estimate $(\mu_1 - \mu_2)$. In Example 7.5, we showed that

$$E(\bar{y}_1 - \bar{y}_2) = \mu_1 - \mu_2$$
$$V(\bar{y}_1 - \bar{y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

You can see that $(\bar{y}_1 - \bar{y}_2)$ is an unbiased estimator for $(\mu_1 - \mu_2)$. Further, it can be shown (proof omitted) that $V(\bar{y}_1 - \bar{y}_2)$ is smallest among all unbiased estimators, i.e., $(\bar{y}_1 - \bar{y}_2)$ is the MVUE for $(\mu_1 - \mu_2)$.

According to the central limit theorem, $(\bar{y}_1 - \bar{y}_2)$ will also be approximately normal for large n_1 and n_2 regardless of the distributions of the sampled populations. Thus, we can apply Theorem 8.2 to construct a large-sample confidence interval for $(\mu_1 - \mu_2)$. The procedure for forming a large-sample confidence interval for $(\mu_1 - \mu_2)$ appears in the box.

Large-Sample $(1 - \alpha)100\%$ Confidence Interval for $(\mu_1 - \mu_2)$: Independent Samples

$$(\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2}\sigma_{(\bar{y}_1 - \bar{y}_2)} = (\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
$$\approx (\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

[*Note:* We have used the sample variances s_1^2 and s_2^2 as approximations to the corresponding population parameters.]

Assumptions:
1. The two random samples are selected in an independent manner from the target populations. That is, the choice of elements in one sample does not affect, and is not affected by, the choice of elements in the other sample.
2. The sample sizes n_1 and n_2 are sufficiently large for the central limit theorem to apply. (We recommend $n_1 \geq 30$ and $n_2 \geq 30$.)

EXAMPLE 8.10

We want to estimate the difference between the mean starting salaries for recent graduates with mechanical engineering and civil engineering degrees from the University of Florida (UF). The following information is available:*

1. A random sample of 59 starting salaries for UF mechanical engineering graduates produced a sample mean of \$32,675 and a standard deviation of \$4,430.
2. A random sample of 30 starting salaries for UF civil engineering graduates produced a sample mean of \$27,460 and a standard deviation of \$4,286.

Solution

We will let the subscript 1 refer to the mechanical engineering graduates and the subscript 2 to the civil engineering graduates. We will also define the following notation:

μ_1 = Population mean starting salary of all recent UF mechanical engineering graduates

μ_2 = Population mean starting salary of all recent UF civil engineering graduates

Similarly, let $\bar{y}_1$ and $\bar{y}_2$ denote the respective sample means; s_1 and s_2, the respective sample standard deviations; and n_1 and n_2, the respective sample sizes. The given information is summarized in Table 8.3.

TABLE 8.3 Summary of Information for Example 8.10

	Mechanical Engineers	Civil Engineers
Sample Size	$n_1 = 59$	$n_2 = 30$
Sample Mean	$\bar{y}_1 = 32{,}675$	$\bar{y}_2 = 27{,}460$
Sample Standard Deviation	$s_1 = \$4{,}430$	$s_2 = 4{,}286$

Source: Career Resource Center, University of Florida.

The general form of a 95% confidence interval for $(\mu_1 - \mu_2)$, based on large, independent samples from the target populations, is given by

$$(\bar{y}_1 - \bar{y}_2) \pm z_{.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Recall that $z_{.025} = 1.96$ and use the information in Table 8.3 to make the following substitutions to obtain the desired confidence interval:

$$\begin{aligned}&(32{,}675 - 27{,}460) \pm 1.96\sqrt{\sigma_1^2/59 + \sigma_2^2/30}\\ &\quad\approx (32{,}675 - 27{,}460) \pm 1.96\sqrt{(4{,}430)^2/59 + (4{,}286)^2/30}\\ &\quad\approx 5{,}215 \pm 1{,}905\end{aligned}$$

or (\$3,310, \$7,120).

*The information for this example was extracted from a 1990 survey of graduates conducted by the Career Resource Center, University of Florida.

If we were to use this method of estimation repeatedly to produce confidence intervals for $(\mu_1 - \mu_2)$, the difference between population means, we would expect 95% of the intervals to enclose $(\mu_1 - \mu_2)$. Hence, we can be reasonably confident that the mean starting salary of mechanical engineering graduates of UF was between \$3,310 and \$7,120 higher than the mean starting salary of civil engineering graduates.

A confidence interval for $(\mu_1 - \mu_2)$, based on small samples from each population, is derived using Student's t distribution. As was the case when estimating a single population mean from information in a small sample, we must make specific assumptions about the relative frequency distributions of the two populations, as indicated in the box. These assumptions are required if either sample is small (i.e., if either $n_1 < 30$ or $n_2 < 30$).

Small-Sample $(1 - \alpha)100\%$ Confidence Interval for $(\mu_1 - \mu_2)$: Independent Samples and $\sigma_1^2 = \sigma_2^2$

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

and the value of $t_{\alpha/2}$ is based on $(n_1 + n_2 - 2)$ degrees of freedom.

Assumptions:
1. Both of the populations from which the samples are selected have relative frequency distributions that are approximately normal.
2. The variances σ_1^2 and σ_2^2 of the two populations are equal.
3. The random samples are selected in an independent manner from the two populations.

Note that this procedure requires that the samples be selected from two normal populations that have equal variances (i.e., $\sigma_1^2 = \sigma_2^2 = \sigma^2$). Since we are assuming the variances are equal, we construct an estimate of σ^2 based on the information contained in *both* samples. This **pooled estimate** is denoted by s_p^2 and is computed as shown in the previous box. You will notice that s_p^2 is a weighted average of the two sample variances, s_1^2 and s_2^2, with the weights proportional to the respective sample sizes.

EXAMPLE 8.11

The *Journal of Testing and Evaluation* (July 1981) reported on the results of laboratory tests conducted to investigate the stability and permeability of open-graded asphalt concrete. In one part of the experiment, four concrete specimens were prepared for asphalt contents of 3% and 7% by total weight of mix. The water permeability of each concrete specimen was determined by flowing deaerated water across the specimen and measuring the amount of water loss. The permeability measurements (recorded in inches per hour) for the eight concrete specimens are shown in Table 8.4. Find a 95% confidence interval for the difference between the mean permeabilities of concrete made with asphalt contents of 3% and 7%. Interpret the interval.

TABLE 8.4 Permeability Measurements for 3% and 7% Asphalt Concrete, Example 8.11

Asphalt Content	3%	1,189	840	1,020	980
	7%	853	900	733	785

Source: Woelfl, G., Wei, I., Faulstich, C., and Litwack, H. "Laboratory testing of asphalt concrete for porous pavements." *Journal of Testing and Evaluation*, Vol. 9, No. 4, July 1981, pp. 175–181. Copyright American Society for Testing and Materials.

Solution

First, we calculate the means and variances of the two samples, using the computer. A SAS printout giving descriptive statistics for the two samples is shown in Figure 8.10. For the 3% asphalt, $\bar{y}_1 = 1{,}007.25$ and $s_1^2 = 20{,}636.92$; for the 7% asphalt, $\bar{y}_2 = 817.75$ and $s_2^2 = 5{,}420.92$.

FIGURE 8.10 ► SAS descriptive statistics for Example 8.11

```
N Obs  Variable  N          Mean      Variance       Std Dev
-------------------------------------------------------------
    4  ASPH3PCT  4       1007.25      20636.92   143.6555487
       ASPH7PCT  4   817.7500000       5420.92    73.6268746
-------------------------------------------------------------
```

Since both samples are small ($n_1 = n_2 = 4$), the procedure requires the assumption that the two samples of permeability measurements are independently and randomly selected from normal populations with equal variances. The 95% small-sample confidence interval is

$$(\bar{y}_1 - \bar{y}_2) \pm t_{.025}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= (1{,}007.25 - 817.75) \pm t_{.025}\sqrt{s_p^2\left(\frac{1}{4} + \frac{1}{4}\right)}$$

where $t_{.025} = 2.447$ is obtained from the t distribution (Table 7 of Appendix II) based on $n_1 + n_2 - 2 = 4 + 4 - 2 = 6$ degrees of freedom, and

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{3(20{,}636.92) + 3(5{,}420.92)}{6}$$

$$= 13{,}028.92$$

is the pooled sample variance. Substitution yields the interval

$$(1{,}007.25 - 817.75) \pm 2.447\sqrt{13{,}028.92\left(\frac{1}{4} + \frac{1}{4}\right)}$$
$$= 189.5 \pm 197.50$$

or, −8.00 to 387.00. This interval could also be obtained using the computer. The MINITAB-generated 95% confidence interval is displayed in Figure 8.11. Our calculated interval agrees with the MINITAB result. The interval is interpreted as follows: We are 95% confident that the interval (−8, 387) encloses the true difference between the mean permeabilities of the two types of concrete. Since the interval includes 0, we are unable to conclude that the two means differ.

FIGURE 8.11 ▶ MINITAB printout for Example 8.11

```
TWOSAMPLE T FOR asph3pct VS asph7pct
            N       MEAN      STDEV   SE MEAN
asph3pct    4       1007        144        72
asph7pct    4      817.8       73.6        37

95 PCT CI FOR MU asph3pct - MU asph7pct: (-8, 387)

TTEST MU asph3pct = MU asph7pct (VS NE): T= 2.35  P=0.057  DF= 6

POOLED STDEV =        114
```

As with the one-sample case, the assumptions required for estimating $(\mu_1 - \mu_2)$ with small samples do not have to be satisfied exactly for the interval estimate to be useful in practice. Slight departures from these assumptions do not seriously affect the level of confidence in the procedure. For example, when the variances σ_1^2 and σ_2^2 of the sampled populations are unequal, researchers have found that the formula for the small-sample confidence interval for $(\mu_1 - \mu_2)$ still yields valid results in practice as long as the two populations are normal and the sample sizes are equal, i.e., $n_1 = n_2$.

This situation occurs in Example 8.11. The sample variances given in Figure 8.10 are $s_1^2 = 20{,}636.92$ and $s_2^2 = 5{,}420.92$. Thus, it is very likely that the population variances, σ_1^2 and σ_2^2, are unequal.* However, since $n_1 = n_2 = 4$, the inference derived from this interval is still valid if we use s_1^2 and s_2^2 as estimates for the population variances (rather than using the pooled sample variance, s_p^2).

In the case where $\sigma_1^2 \neq \sigma_2^2$ and $n_1 \neq n_2$, an approximate confidence interval for $(\mu_1 - \mu_2)$ can be constructed by modifying the degrees of freedom associated with the t distribution, and, again, substituting s_1^2 for σ_1^2 and s_2^2 for σ_2^2. These modifications are shown in the box.

*A method for comparing two population variances is presented in Section 8.11.

Approximate Small-Sample Inferences for $(\mu_1 - \mu_2)$ when $\sigma_1^2 \neq \sigma_2^2$

To obtain approximate confidence intervals and tests for $(\mu_1 - \mu_2)$ when $\sigma_1^2 \neq \sigma_2^2$, make the following modifications to the degrees of freedom ν used in the t distribution and the estimated standard error:

$$n_1 = n_2 = n: \quad \nu = n_1 + n_2 - 2 = 2(n-1) \qquad \hat{\sigma}_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{1}{n}(s_1^2 + s_2^2)}$$

$$n_1 \neq n_2: \quad \nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} \qquad \hat{\sigma}_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

[*Note:* In the case of $n_1 \neq n_2$, the value of ν will not generally be an integer. Round ν down to the nearest integer to use the t table.

Assumptions:
1. Both of the populations from which the samples are selected have relative frequency distributions that are approximately normal.
2. The random samples are selected in an independent manner from the two populations.

EXERCISES

8.32 Epidemiologists have theorized that the risk of coronary heart disease can be reduced by an increased consumption of fish. One study, begun in 1960, monitored the diet and health of a random sample of middle-age Dutchmen (*New England Journal of Medicine*, May 1985). The men were divided into groups according to the number of grams of fish consumed per day. Twenty years later, the level of dietary cholesterol (one of the risk factors for coronary disease) present in each was recorded. The results for two groups of subjects, the "no fish consumption" group (0 grams per day) and the "high fish consumption" group (greater than 45 grams per day), are summarized in the table. (Dietary cholesterol is measured in milligrams per 1,000 calories.)

	No Fish Consumption 0 grams/day	*High Fish Consumption 45 grams/day*
Sample Size	159	79
Mean	146	158
Standard Deviation	66	75

Source: Kromhout, D., Bosschieter, E. B., and Coulander, C. L. "The inverse relationship between fish consumption and 20-year mortality from coronary heart disease." *New England Journal of Medicine*, May 9, 1985, Vol. 312, No. 19, pp. 1205–1209. Reprinted by permission.

a. Calculate an approximate 99% confidence interval for the difference between the mean levels of dietary cholesterol present in the two groups.

b. Based on the interval constructed in part **a**, what can you infer about the true difference? Explain.

8.33 Marine biochemists at the University of Tokyo studied the properties of crustacean skeletal muscles (*The Journal of Experimental Zoology*, Sept. 1993). It is well known that certain muscles contract faster than others. The main purpose of the experiment was to compare the biochemical properties of fast and slow muscles of crayfish. Using crayfish obtained from a local supplier, twelve fast-muscle fiber bundles were extracted and each fiber bundle tested for uptake of the protein Ca^{2+}. Twelve slow-muscle fiber bundles were extracted from a second sample of crayfish, and Ca^{2+} uptake measured. The results of the experiment are summarized here. (All Ca^{2+} measurements are in moles per milligram.) Analyze the data using a 95% confidence interval. Make an inference about the difference between the protein uptake means of fast and slow muscles.

Fast Muscle	*Slow Muscle*
$n_1 = 12$	$n_2 = 12$
$\bar{y}_1 = .57$	$\bar{y}_2 = .37$
$s_1 = .104$	$s_2 = .035$

Source: Ushio, H., and Watabe, S. "Ultrastructural and biochemical analysis of the sarcoplasmic reticulum from crayfish fast and slow striated muscles." The *Journal of Experimental Zoology*, Vol. 267, Sept. 1993, p. 16 (Table 1).

8.34 Refer to the Harris Corporation/University of Florida study to determine whether a manufacturing process performed at a remote location can be established locally, Exercise 2.12. Test devices (pilots) were set up at both the old and new locations and voltage readings on 30 production runs at each location were obtained. The data are reproduced in the table. Descriptive statistics are displayed in the accompanying SAS printout. [*Note:* Larger voltage readings are better than smaller voltage readings.]

Old Location			*New Location*		
9.98	10.12	9.84	9.19	10.01	8.82
10.26	10.05	10.15	9.63	8.82	8.65
10.05	9.80	10.02	10.10	9.43	8.51
10.29	10.15	9.80	9.70	10.03	9.14
10.03	10.00	9.73	10.09	9.85	9.75
8.05	9.87	10.01	9.60	9.27	8.78
10.55	9.55	9.98	10.05	8.83	9.35
10.26	9.95	8.72	10.12	9.39	9.54
9.97	9.70	8.80	9.49	9.48	9.36
9.87	8.72	9.84	9.37	9.64	8.68

Source: Harris Corporation, Melbourne, Fla.

```
Analysis Variable : VOLTAGE

---------------------------- LOCATION=OLD ----------------------

N Obs   N        Minimum        Maximum           Mean       Std Dev
---------------------------------------------------------------------
   30  30      8.0500000     10.5500000      9.8036667     0.5409155
---------------------------------------------------------------------

---------------------------- LOCATION=NEW ----------------------

N Obs   N        Minimum        Maximum           Mean       Std Dev
---------------------------------------------------------------------
   30  30      8.5100000     10.1200000      9.4223333     0.4788757
---------------------------------------------------------------------
```

a. Compare the mean voltage readings at the two locations using a 90% confidence interval.
b. Based on the interval, part **a**, does it appear that the manufacturing process can be established locally?

8.35 The methodology for conducting a stress analysis of newly designed timber structures is well known. However, few data are available on the actual or allowable stress for repairing damaged structures. Consequently, design engineers often propose a repair scheme (e.g., gluing) without any knowledge of its structural effectiveness. To partially fill this void, a stress analysis was conducted on epoxy-repaired truss joints (*Journal of Structural Engineering*, Feb. 1986). Tests were conducted on epoxy-bonded truss joints made of various species of wood to determine actual glue-line shear stress recorded in pounds per square inch (psi). Summary information for independent random samples of southern pine and ponderosa pine truss joints is given in the accompanying table. Estimate the difference between the mean shear strengths of epoxy-repaired truss joints for the two species of wood with a 90% confidence interval.

	Southern Pine	*Ponderosa Pine*
Sample Size	100	47
Mean Shear Stress, psi	1,312	1,352
Standard Deviation	422	271

Source: Avent, R. R. "Design criteria for epoxy repair of timber structures." *Journal of Structural Engineering*, Vol. 112, No. 2, Feb. 1986, pp. 232.

8.36 To investigate the possible link between fluoride content of drinking water and cancer, Yiamouyiannis and Burk (1977) recorded cancer death rates (number of deaths per 100,000 population) from 1952–1969 in 20 selected U.S. cities—the 10 largest fluoridated cities and the 10 largest cities not fluoridated by 1969. Maritz and Jarrett (*Applied Statistics*, Feb. 1983) used the data collected by Yiamouyiannis and Burk to calculate for each city the annual rate of increase in cancer death rate over this 18-year period for each of four age groups: under 25, 25–44, 45–64, and 65 or older. The data for the 45–64 age group are reproduced in the table at the top of page 380, followed by a MINITAB analysis of the data.

Fluoridated City	*Fluoridated* Annual Increase in Cancer Death Rate	*Nonfluoridated* City	*Nonfluoridated* Annual Increase in Cancer Death Rate
Chicago	1.0640	Los Angeles	.8875
Philadelphia	1.4118	Boston	1.7358
Baltimore	2.1115	New Orleans	1.0165
Cleveland	1.9401	Seattle	.4923
Washington	3.8772	Cincinnati	4.0155
Milwaukee	−.4561	Atlanta	−1.1744
St. Louis	4.8359	Kansas City	2.8132
San Francisco	1.8875	Columbus	1.7451
Pittsburgh	4.4964	Newark	−.5676
Buffalo	1.4045	Portland	2.4471

Source: Maritz, J. S., and Jarrett, R. G. "The use of statistics to examine the association between fluoride in drinking water and cancer death rates." *Applied Statistics*, Vol. 32, No. 2, 1983, pp. 97–101.

```
TWOSAMPLE T FOR fluorat VS nonflrat
            N      MEAN      STDEV    SE MEAN
fluorat    10      2.26       1.66       0.52
nonflrat   10      1.34       1.56       0.49

95 PCT CI FOR MU fluorat - MU nonflrat: (-0.60, 2.43)

TTEST MU fluorat = MU nonflrat (VS NE): T= 1.27  P=0.22  DF=  18

POOLED STDEV =        1.61
```

a. Find a 95% confidence interval for the difference between the mean annual increases in cancer death rates for fluoridated and nonfluoridated cities.
b. Interpret the interval obtained in part a.
c. What assumptions are necessary for the validity of the interval estimation procedure and any inferences derived from it? Do you think these assumptions are satisfied?

8.37 Agricultural experts in Israel have developed a new method of irrigation, called *fertigation*, in which fertilizer is added to water and the mixture is dripped periodically onto the roots of the plants. Very little water—a precious commodity in Israel—is wasted, and the nutrients go directly where they are needed. To test this new process, 100 acres were randomly selected and their historical yields were recorded. The fertigation process was then applied to the new crop and the new yields were recorded. The accompanying table summarizes the results.

	Before Fertigation	*After Fertigation*
Sample Size	100	100
Mean Yield	40%	75%
Standard Deviation	8%	6%

a. Estimate the difference between the true mean yields before and after fertigation. Use a 90% confidence interval.
b. Interpret the confidence interval of part a.

8.38 *Sintering*, one of the most important techniques of materials science, is used to convert powdered material into a porous solid body. The following two measures characterize the final product:

$$V_V = \text{Percentage of total volume of final product that is solid}$$

$$= \left(\frac{\text{Solid volume}}{\text{Porous volume} + \text{Solid volume}}\right) \cdot 100$$

$$S_V = \text{Solid-pore interface area per unit volume of the product}$$

When $V_V = 100\%$, the product is completely solid—i.e., it contains no pores. Both V_V and S_V are estimated by a microscopic examination of polished cross sections of sintered material. The accompanying table gives the mean and standard deviation of the values of S_V (in squared centimeters per cubic centimeter) and V_V (percentage) for $n = 100$ specimens of sintered nickel for two different sintering times.

	S_V		V_V	
Time	$\bar{y}$	s	$\bar{y}$	s
10 minutes	736.0	181.9	96.73	2.1
150 minutes	299.5	161.0	97.82	1.5

Data and experimental information provided by Guoquan Liu while visiting at the University of Florida in 1983.

a. Find a 95% confidence interval for the mean change in S_V between sintering times of 10 minutes and 150 minutes. What inference would you make concerning the difference in mean sintering times?

b. Repeat part a for V_V.

8.7 Estimation of the Difference Between Two Population Means: Matched Pairs

The large- and small-sample procedures for estimating the difference between two population means presented in Section 8.6 were based on the assumption that the samples were randomly and independently selected from the target populations. Sometimes we can obtain more information about the difference between population means, $(\mu_1 - \mu_2)$, by selecting **paired observations**.

For example, suppose you want to compare two methods for drying concrete using samples of five cement mixes with each method. One method of sampling would be to randomly select 10 mixes (say, A, B, C, D, . . . , J) from among all available mixes and then randomly assign five to drying method 1 and five to drying method 2 (see Table 8.5 on page 382). The strength measurements obtained after conducting a series of strength tests would represent independent random samples of strengths attained by concrete specimens dried by the two different methods. The difference between the mean strength measurements, $(\mu_1 - \mu_2)$, could be estimated using the confidence interval procedure described in Section 8.6.

TABLE 8.5 Independent Random Samples of Cement Mixes Assigned to Each Method

Method 1	Method 2
Mix A	Mix B
Mix E	Mix C
Mix F	Mix D
Mix H	Mix G
Mix J	Mix I

A better method of sampling would be to match the concrete specimens in pairs according to type of mix. From each mix pair, one specimen would be randomly selected to be dried by method 1; the other specimen would be assigned to be dried by method 2, as shown in Table 8.6. Then the differences between **matched pairs** of strength measurements should provide a clearer picture of the difference in strengths for the two drying methods because the matching would tend to cancel the effects of the factors that formed the basis of the matching (i.e., the effects of the different cement mixes).

TABLE 8.6 Set-up of the Matched-Pairs Design for Comparing Two Methods of Drying Concrete

Type of Mix	Method 1	Method 2
A	Specimen 2	Specimen 1
B	Specimen 2	Specimen 1
C	Specimen 1	Specimen 2
D	Specimen 2	Specimen 1
E	Specimen 1	Specimen 2

In a matched-pairs experiment, the symbol μ_d is commonly used to denote the mean difference between matched pairs of measurements, where $\mu_d = (\mu_1 - \mu_2)$. Once the differences in the sample are calculated, a confidence interval for μ_d is identical to the confidence interval for the mean of a single population given in Section 8.5.

The procedure for estimating the difference between two population means based on matched-pairs data for both large and small samples is given in the box.

$(1 - \alpha)100\%$ Confidence Interval for $\mu_d = (\mu_1 - \mu_2)$: Matched Pairs

Let $d_1, d_2, \ldots, d_n$ represent the differences between the pairwise observations in a *random sample* of n matched pairs, $\bar{d}$ = mean of the n sample differences, and s_d = standard deviation of the n sample differences.

Large Sample

$$\bar{d} \pm z_{\alpha/2}\left(\frac{\sigma_d}{\sqrt{n}}\right)$$

where σ_d is the population deviation of differences.

Assumption: $n \geq 30$

[*Note:* When σ_d is unknown (as is usually the case), use s_d to approximate σ_d.]

Small Sample

$$\bar{d} \pm t_{\alpha/2}\left(\frac{s_d}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ is based on $(n - 1)$ degrees of freedom.

Assumption: The population of paired differences is normally distributed.

EXAMPLE 8.12

One desirable characteristic of water pipes is that the quality of water they deliver be equal to or near the quality of water entering the system at the water treatment plant. A type of ductile iron pipe has provided an excellent water delivery system for the St. Louis County Water Company. The chlorine levels of water emerging from the South water treatment plant and at the Fire Station (Fenton Zone 13) were measured over a 12-month period, with the results shown in Table 8.7. Find a 95% confidence interval for the mean difference in monthly chlorine content between the two locations.

TABLE 8.7 Chlorine Content Data for Example 8.12

		Month											
		Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Location	South Plant	2.0	2.0	2.1	1.9	1.7	1.8	1.7	1.9	2.0	2.0	2.1	2.1
	Fire Station	2.2	2.2	2.1	2.0	1.9	1.9	1.8	1.7	1.9	1.9	1.8	2.0
Difference		−.2	−.2	0	−.1	−.2	−.1	−.1	.2	.1	.1	.3	.1

Source: "St. Louis County Standardizes Pipe and Procedures for Reliability." Staff Report, Water and Sewage Works, Dec. 1980.

Solution

Since the chlorine levels at the two plants were recorded over the same 12 months, the data are collected as matched pairs. We want to estimate $\mu_d = (\mu_1 - \mu_2)$, where

μ_1 = Mean monthly chlorine level at the South Plant

μ_2 = Mean monthly chlorine level at the Fire Station

The differences between pairs of monthly chlorine levels are computed as

$$d = (\text{South Plant level}) - (\text{Fire Station level})$$

and are shown in the last row of Table 8.7.

Since the number of differences, $n = 12$, is small, we must assume that these differences are from an approximately normal distribution in order to proceed. The mean and standard deviation of these sample differences are shown (shaded) on the SAS printout, Figure 8.12. From the printout, $\bar{d} = -.0083$ and $s_d = .1676$.

FIGURE 8.12 ▶ SAS descriptive statistics for matched pairs, Example 8.12

Analysis Variable : DIFF (Plant minus Station)

N Obs	N	Minimum	Maximum	Mean	Std Dev
12	12	-0.2000000	0.3000000	-0.0083333	0.1676486

The value of $t_{.025}$, based on $(n - 1) = (12 - 1) = 11$ degrees of freedom, is given in Table 7 of Appendix II as $t_{.025} = 2.201$. Substituting these values into the formula for the small-sample confidence interval, we obtain

$$\bar{d} \pm t_{.025}\left(\frac{s_d}{\sqrt{n}}\right)$$

$$= -.0083 \pm 2.201\left(\frac{.1676}{\sqrt{12}}\right)$$

$$= -.0083 \pm .1065$$

or $(-.1148, .0982)$.

We estimate, with 95% confidence, that the difference between the mean monthly chlorine levels of water at the two St. Louis locations falls within the interval from $-.1148$ to $.0982$. Since 0 is within the interval, there is insufficient evidence to conclude there is a difference between the two means.

In an analysis of matched-pair observations, it is important to stress that the pairing of the experimental units (the objects upon which the measurements are taken) must be performed *before* the data are collected. Recall that the objective is to compare two methods of "treating" the experimental units. By using the matched pairs of units that have similar characteristics, we are able to cancel out the effects of the variables used to match the pairs.

EXERCISES

8.39 Pesticides applied to an extensively grown crop can result in inadvertent area-wide air contamination. *Environmental Science & Technology* (Oct. 1993) reported on air deposition residues of the insecticide diazinon used on dormant orchards in the San Joaquin Valley, California. Ambient air samples were collected and analyzed at an orchard site for each of 11 days during the most intensive period of spraying. The levels of diazinon residue (in ng/m^3) during the day and at night are recorded in the table. The researchers want to know whether the mean diazinon residue levels differ from day to night.

	Diazinon Residue	
Date	*Day*	*Night*
Jan. 11	5.4	24.3
12	2.7	16.5
13	34.2	47.2
14	19.9	12.4
15	2.4	24.0
16	7.0	21.6
17	6.1	104.3
18	7.7	96.9
19	18.4	105.3
20	27.1	78.7
21	16.9	44.6

Source: Selber, J. N., et al. "Air and fog deposition residues for organophosphate insecticides used on dormant orchards in the San Joaquin Valley, California." *Environmental Science & Technology*, Vol. 27, No. 10, Oct. 1993, p. 2240 (Table IV).

a. Analyze the data using a 90% confidence interval.
b. What assumptions are necessary for the validity of the interval estimation procedure of part **a**?
c. Use the interval, part **a**, to answer the researchers' question.

8.40 The *Journal of Environmental Engineering* (Feb. 1986) reported on a heat transfer model designed to predict winter heat loss in wastewater treatment clarifiers. The analysis involved a comparison of clear-sky solar irradiation for horizontal surfaces at different sites in the midwest. The day-long solar irradiation levels (in BTU/sq. ft.) at two midwestern locations of different latitudes (St. Joseph, Missouri, and Iowa Great Lakes) were recorded on each of seven clear-sky winter days. The data are given in the table. Find a 95% confidence interval for the mean difference between the day-long clear-sky solar irradiation levels at the two sites. Interpret the results.

Date	*St. Joseph, Mo.*	*Iowa Great Lakes*
December 21	782	593
January 6	965	672
January 21	948	750
February 6	1,181	988
February 21	1,414	1,226
March 7	1,633	1,462
March 21	1,852	1,698

Source: Wall, D. J., and Peterson, G. "Model for winter heat loss in uncovered clarifiers." *Journal of Environmental Engineering*, Vol. 112, No. 1, Feb. 1986, p. 128.

8.41 A federal traffic safety researcher was hired to ascertain the effect of wearing safety devices (shoulder harnesses, seat belts) on reaction times to peripheral stimuli. To investigate this question, he randomly selected 15 subjects from the students enrolled in a driver education program. Each subject performed a simulated driving task that allowed reaction times to be recorded under two conditions, wearing a safety device (restrained condition) and no safety device (unrestrained condition). Thus, each subject received two reaction-time scores, one for the restrained condition and one for the unrestrained condition. The data (in hundredths of a second) are shown in the accompanying table, followed by a MINITAB printout of the analysis.

Driver	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Restrained	36.7	37.5	39.3	44.0	38.4	43.1	36.2	40.6	34.9	31.7	37.5	42.8	32.6	36.8	38.0
Unrestrained	36.1	35.8	38.4	41.7	38.3	42.6	33.6	40.9	32.5	30.7	37.4	40.2	33.1	33.6	37.5

```
               N        MEAN      STDEV   SE MEAN     90.0 PERCENT C.I.
RminusU       15       1.180      1.191     0.307   (   0.638,    1.722)
```

a. Find a 90% confidence interval for the difference between mean reaction-time scores for the restrained and unrestrained drivers.
b. What assumptions are necessary for the validity of the interval estimation procedure of part **a**?
c. Based on the interval of part **a**, what would you infer about the mean reaction times for the driving conditions?

8.42 Medical researchers believe that exposure to dust from cotton bract induces respiratory disease in susceptible field workers. An experiment was conducted to determine the effect of air-dried green cotton bract extract (GBE) on the cells of mill workers not exposed to dust (*Environmental Research*, Feb. 1986). Blood samples taken on eight workers were incubated with varying concentrations of GBE. After a short period of time, the cyclic AMP level (a measure of cell activity expressed in picomoles per million cells) of each blood sample was measured. The data for two GBE concentrations, 0 mg/ml (salt buffer, control solution) and .2 mg/ml, are reproduced in the table. [Note that one blood sample was taken from each worker, with one aliquot exposed to the salt buffer solution and the other to the GBE.]

Worker	*GBE Concentration, mg/ml*	
	0	*.2*
A	8.8	4.4
B	13.0	5.7
C	9.2	4.4
D	6.5	4.1
F	9.1	4.4
H	17.0	7.9

Source: Butcher, B. T., Reed, M. A., and O'Neil, C. E. "Biochemical and immunologic characterization of cotton bract extract and its effect on *in vitro* cyclic AMP production." *Environmental Research*, Vol. 39, No. 1, Feb. 1986, p. 119.

a. Find a 95% confidence interval for the mean difference between the cyclic AMP levels of blood samples exposed to the two concentrations of GBE.
b. Based on the interval obtained in part **a**, is there evidence that exposure to GBE blocks cell activity?

8.43 Many Vietnam veterans have dangerously high levels of the dioxin 2,3,7,8-TCDD in blood and fat tissue as a result of their exposure to the defoliant Agent Orange. A study published in *Chemosphere* (Vol. 20, 1990) reported on the TCDD levels of 20 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The amounts of TCDD (measured in parts per trillion) in blood plasma and fat tissue drawn from each veteran are shown in the table followed by a SAS printout giving summary statistics. Use the information on the printout to construct a confidence interval that will allow you to compare the mean TCDD level in plasma to the mean TCDD level in fat tissue for Vietnam veterans exposed to Agent Orange. Interpret the result.

Veteran	*TCDD Levels in Plasma*	*TCDD Levels in Fat Tissue*	*Veteran*	*TCDD Levels in Plasma*	*TCDD Levels in Fat Tissue*
1	2.5	4.9	11	6.9	7.0
2	3.1	5.9	12	3.3	2.9
3	2.1	4.4	13	4.6	4.6
4	3.5	6.9	14	1.6	1.4
5	3.1	7.0	15	7.2	7.7
6	1.8	4.2	16	1.8	1.1
7	6.0	10.0	17	20.0	11.0
8	3.0	5.5	18	2.0	2.5
9	36.0	41.0	19	2.5	2.3
10	4.7	4.4	20	4.1	2.5

Source: Schecter, A., et. al. "Partitioning of 2,3,7,8-chlorinated dibenzo-*p*-dioxins and dibenzofurans between adipose tissue and plasma lipid of 20 Massachusetts Vietnam veterans." *Chemosphere*, Vol. 20, Nos. 7–9, 1990, pp. 954–955 (Table I & II).

N Obs	Variable	N	Minimum	Maximum	Mean	Std Dev
20	PLASMA	20	1.6000000	36.0000000	5.9900000	8.1279829
	FAT	20	1.1000000	41.0000000	6.8600000	8.4656209
	DIFF	20	-5.0000000	9.0000000	-0.8700000	2.9773001

8.8 Estimation of a Population Proportion

We will now consider the method for estimating the binomial proportion p of successes—that is, the proportion of elements in a population that have a certain characteristic. For example, a quality control inspector may be interested in the proportion of defective items produced on an assembly line; or a supplier of heating oil may be interested in the proportion of homes in its service area that are heated by natural gas.

A logical candidate for a point estimate of the population proportion p is the sample proportion $\hat{p} = y/n$, where y is the number of observations in a sample of size n that have the characteristic of interest (i.e., y is the number of "successes"). In Example 7.7, we showed that for large n, $\hat{p}$ is approximately normal with mean

$$E(\hat{p}) = p$$

and variance

$$V(\hat{p}) = \frac{pq}{n}$$

Therefore, $\hat{p}$ is an unbiased estimator of p and (proof omitted) has the smallest variance among all unbiased estimators; that is, $\hat{p}$ is the MVUE for p. Since $\hat{p}$ is approximately normal, we can use it as a pivotal statistic and apply Theorem 8.2 to derive the formula for a large-sample confidence interval for p shown in the box.

Large-Sample $(1 - \alpha)100\%$ Confidence Interval for a Population Proportion, p

$$\hat{p} \pm z_{\alpha/2}\sigma_{\hat{p}} \approx \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $\hat{p}$ is the sample proportion of observations with the characteristic of interest, and $\hat{q} = 1 - \hat{p}$.

[*Note:* The interval is approximate since we must substitute the sample $\hat{p}$ and $\hat{q}$ for the corresponding population values for $\sigma_{\hat{p}}$.]

Assumption: The sample size n is sufficiently large so that the approximation is valid. As a rule of thumb, the condition of a "sufficiently large" sample size will be satisfied if $n\hat{p} \geq 4$ and $n\hat{q} \geq 4$.

Note that we must substitute $\hat{p}$ and $\hat{q}$ into the formula for $\sigma_{\hat{p}} = \sqrt{pq/n}$ to construct the interval. This approximation will be valid as long as the sample size n is sufficiently large. Many researchers adopt the rule of thumb that n is "sufficiently large" if the interval $\hat{p} \pm 2\sqrt{\hat{p}\hat{q}/n}$ does not contain 0 or 1. Recall (Section 7.6) that this rule is satisfied if $n\hat{p} \geq 4$ and $n\hat{q} \geq 4$.

EXAMPLE 8.13

Stainless steels are frequently used in chemical plants to handle corrosive fluids. However, these steels are especially susceptible to stress corrosion cracking in certain environments. In a sample of 295 steel alloy failures that occurred in oil refineries and petrochemical plants in Japan over the last 10 years, 118 were caused by stress corrosion cracking and corrosion fatigue (*Materials Performance*, June 1981). Construct a 95% confidence interval for the true proportion of alloy failures caused by stress corrosion cracking.

Solution The sample proportion of alloy failures caused by corrosion is

$$\hat{p} = \frac{\text{Number of alloy failures in sample caused by corrosion}}{\text{Number of alloy failures in sample}} = \frac{118}{295} = .4$$

Thus, $\hat{q} = 1 - .4 = .6$. The approximate 95% confidence interval is then

$$\hat{p} \pm z_{.025}\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 \pm 1.96\sqrt{\frac{(.4)(.6)}{295}} = .4 \pm .056$$

or (.344, .456). [Note that the approximation is valid since $n\hat{p} = 118$ and $n\hat{q} = 177$ both exceed 4.]

We are 95% confident that the interval from .344 to .456 encloses the true proportion of alloy failures that were caused by corrosion. If we repeatedly selected random samples of $n = 295$ alloy failures and constructed a 95% confidence interval based on each sample, then we would expect 95% of the confidence intervals constructed to contain p.

Small-sample procedures are available for the estimation of a population proportion p. These techniques are similar to those small-sample procedures for estimating a population mean μ. (Recall that $\hat{p} = y/n$ can be thought of as a mean of a sample of 0–1 Bernoulli outcomes.) The details are not included in our discussion, however, because most surveys in actual practice use samples that are large enough to employ the procedure of this section.

EXERCISES

8.44 An American Housing Survey (AHS) conducted by the U.S. Department of Commerce revealed that 705 of 1,500 sampled homeowners are "do-it-yourselfers"—they did most the work themselves on at least one of their home improvements or repairs (Bureau of the Census, *Statistical Brief*, May 1992). Estimate the true proportion of American homeowners who do most of the home improvement or repair work themselves using a 95% confidence interval. Interpret the result.

8.45 The "Black Hole" survey, sponsored by the Professional Employment Research Council, reports on the toughest jobs to fill on recruiters lists. In the most recent survey, 95 of 285 recruiters listed engineering positions as the "toughest to fill" (*Industrial Engineering*, Aug. 1990). Estimate the true percentage of recruiters who find it toughest to fill engineering positions. Use a 99% confidence interval.

8.46 Refer to the *Journal of the Medical Association* (Apr. 21, 1993) report on the prevalence of cigarette smoking among U.S. adults, Exercise 8.29. Of the 43,732 survey respondents, 11,239 indicated that they were current smokers and 10,539 indicated they were former smokers.

a. Construct and interpret a 90% confidence interval for the percentage of U.S. adults who currently smoke cigarettes.

b. Construct and interpret a 90% confidence interval for the percentage of U.S. adults who are former cigarette smokers.

8.47 According to a study conducted by the California Division of Labor Research and Statistics (*Engineering News Record*, Mar. 10, 1983), roofing is one of the most hazardous occupations. Of 2,514 worker injuries that caused absences for a full workday or shift after the injury, 23% were attributable to falls from high elevations on level surfaces, 21% to falling hand tools or other materials, 19% to overexertion, and 20% to burns or scalds. Assume that the 2,514 injuries can be regarded as a random sample from the population of all roofing injuries in California.

a. Construct a 95% confidence interval for the proportion of all injuries that are due to falls.

b. Construct a 95% confidence interval for the proportion of all injuries that are due to burns or scalds.

8.48 As part of a cooperative research agreement between the United States and Japan, a full-scale reinforced concrete building was designed and tested under simulated earthquake loading conditions in Japan (*Journal of Structural Engineering*, Jan. 1986). For one part of the study, several U.S. design engineers were asked to evaluate the new design. Of the 48 engineers surveyed, 36 believed the shear wall of the structure to be too lightly reinforced. Find a 95% confidence interval for the true proportion of U.S. design engineers who consider the shear wall of the building too lightly reinforced.

8.49 Astronauts often report episodes of disorientation as they move around the zero-gravity spacecraft. To compensate, crew members rely heavily on visual information to establish a top-down orientation. An empirical study was conducted to assess the potential of using color brightness as a body orientation cue (*Human Factors*, Dec. 1988). Ninety college students, reclining on their backs in the dark, were disoriented when positioned on a rotating platform under a slowly rotating disk that filled their entire field of vision. Half the disk was painted with a brighter level of color than the other half. The students were asked to say "stop" when they believed they were right-side up, and the brightness level of the disk was recorded. Of the 90 students, 58 selected the brighter color level.

a. Use this information to estimate the true proportion of subjects who use the bright color level as a cue to being right-side up. Construct a 95% confidence interval for the true proportion.

b. Can you infer from the result, part **a**, that a majority of subjects would select bright color levels over dark color levels as a cue to being right-side up? Explain.

8.50 The U.S. Food and Drug Administration (FDA) recently approved the marketing of a new chemical solution, Caridex, which dissolves cavities. In a study conducted by dental researchers at Northwestern University, 21 of 35 patients with cavities preferred treatment with Caridex to drilling (*Gainesville Sun*, Feb. 11, 1988). Estimate the true proportion of dental patients who prefer having their cavities dissolved with Caridex rather than drilled. Use a 99% confidence interval and interpret the result.

8.9 Estimation of the Difference Between Two Population Proportions

This section extends the method of Section 8.8 to the case in which we want to estimate the difference between two binomial proportions. For example, we may be interested in comparing the proportion p_1 of defective items produced by machine 1 to the proportion p_2 of defective items produced by machine 2.

Let y_1 and y_2 represent the numbers of successes in two independent binomial experiments with samples of size n_1 and n_2, respectively. To estimate the difference

$(p_1 - p_2)$, where p_1 and p_2 are binomial parameters—i.e., the probabilities of success in the two independent binomial experiments—consider the proportion of successes in each of the samples:

$$\hat{p}_1 = \frac{y_1}{n_1} \qquad \text{and} \qquad \hat{p}_2 = \frac{y_2}{n_2}$$

Intuitively, we would expect $(\hat{p}_1 - \hat{p}_2)$ to provide a reasonable estimate of $(p_1 - p_2)$. Since $(\hat{p}_1 - \hat{p}_2)$ is a linear function of the binomial random variables y_1 and y_2, where $E(y_i) = n_i p_i$ and $V(y_i) = n_i p_i q_i$, we have

$$\begin{aligned} E(\hat{p}_1 - \hat{p}_2) &= E(\hat{p}_1) - E(\hat{p}_2) = E\left(\frac{y_1}{n_1}\right) - E\left(\frac{y_2}{n_2}\right) \\ &= \frac{1}{n_1}E(y_1) - \frac{1}{n_2}E(y_2) = \frac{1}{n_1}(n_1 p_1) - \frac{1}{n_2}(n_2 p_2) \\ &= p_1 - p_2 \end{aligned}$$

and

$$\begin{aligned} V(\hat{p}_1 - \hat{p}_2) &= V(\hat{p}_1) + V(\hat{p}_2) - 2\,\mathrm{Cov}(\hat{p}_1, \hat{p}_2) \\ &= V\left(\frac{y_1}{n_1}\right) + V\left(\frac{y_2}{n_2}\right) - 0 \qquad \text{since } y_1 \text{ and } y_2 \text{ are independent} \\ &= \frac{1}{n_1^2}V(y_1) + \frac{1}{n_2^2}V(y_2) \\ &= \frac{1}{n_1^2}(n_1 p_1 q_1) + \frac{1}{n_2^2}(n_2 p_2 q_2) \\ &= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \end{aligned}$$

Thus, $(\hat{p}_1 - \hat{p}_2)$ is an unbiased estimator of $(p_1 - p_2)$ and, in addition, it has minimum variance (proof omitted).

The central limit theorem also guarantees that, for sufficiently large sample sizes n_1 and n_2, the sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ will be approximately normal. It follows (Theorem 8.2) that a large-sample confidence interval for $(p_1 - p_2)$ may be obtained as shown in the box on page 392.

Note that we must substitute the values of $\hat{p}_1$ and $\hat{p}_2$ for p_1 and p_2, respectively, to obtain an estimate of $\sigma_{(\hat{p}_1 - \hat{p}_2)}$. As in the one-sample case, this approximation is reasonably accurate when both n_1 and n_2 are sufficiently large, i.e., if the intervals

$$\hat{p}_1 \pm 2\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1}} \quad \text{and} \quad \hat{p}_2 \pm 2\sqrt{\frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

do not contain 0 or the sample size (n_1 or n_2). This will be true if $n_1\hat{p}_1$, $n_2\hat{p}_2$, $n_1\hat{q}_1$, and $n_2\hat{q}_2$ are all greater than or equal to 4.

Large-Sample $(1-\alpha)100\%$ Confidence Interval for $(p_1 - p_2)$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sigma_{(\hat{p}_1-\hat{p}_2)} \approx (\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

where $\hat{p}_1$ and $\hat{p}_2$ are the sample proportions of observations with the characteristic of interest.

[*Note:* We have followed the usual procedure of substituting the sample values $\hat{p}_1$, $\hat{q}_1$, $\hat{p}_2$, and $\hat{q}_2$ for the corresponding population values required for $\sigma_{(\hat{p}_1-\hat{p}_2)}$.

Assumption: The samples are sufficiently large that the approximation is valid. As a general rule of thumb, we will require that $n_1\hat{p}_1 \geq 4$, $n_1\hat{q}_1 \geq 4$, $n_2\hat{p}_2 \geq 4$, and $n_2\hat{q}_2 \geq 4$.

EXAMPLE 8.14

A traffic engineer conducted a study of vehicular speeds on a segment of street that had the posted speed limit changed several times. When the posted speed limit on the street was 30 miles per hour, the engineer monitored the speeds of 100 randomly selected vehicles traversing the street and observed 49 violations of the speed limit. After the speed limit was raised to 35 miles per hour, the engineer again monitored the speeds of 100 randomly selected vehicles and observed 19 vehicles in violation of the speed limit. Find a 99% confidence interval for $(p_1 - p_2)$, where p_1 is the true proportion of vehicles that (under similar driving conditions) exceed the lower speed limit (30 miles per hour) and p_2 is the true proportion of vehicles that (under similar driving conditions) exceed the higher speed limit (35 miles per hour). Interpret the interval.

Solution

In this example,

$$\hat{p}_1 = \frac{49}{100} = .49 \quad \text{and} \quad \hat{p}_2 = \frac{19}{100} = .19$$

Note that

$$n_1\hat{p}_1 = 49 \qquad n_1\hat{q}_1 = 51$$
$$n_2\hat{p}_2 = 19 \qquad n_2\hat{q}_2 = 81$$

all exceed 4. Thus, we can apply the approximation for a large-sample confidence interval for $(p_1 - p_2)$.

For a confidence interval of $(1 - \alpha) = .99$, we have $\alpha = .01$ and $z_{\alpha/2} = z_{.005} = 2.58$ (from Table 4 of Appendix II). Substitution into the confidence interval formula yields:

$$\begin{aligned}(\hat{p}_1 - \hat{p}_2) &\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \\ &= (.49 - .19) \pm 2.58\sqrt{\frac{(.49)(.51)}{100} + \frac{(.19)(.81)}{100}} \\ &= .30 \pm .164\end{aligned}$$

Our interpretation is that the true difference, $(p_1 - p_2)$, falls between .136 and .464 with 99% confidence. Since the lower bound on our estimate is positive (.136), we are fairly confident that the proportion of all vehicles in violation of the lower speed limit (30 miles per hour) exceeds the corresponding proportion in violation of the higher speed limit (35 miles per hour) by at least .136.

Small-sample estimation procedures for $(p_1 - p_2)$ will not be discussed here for the reasons outlined at the end of Section 8.8.

EXERCISES

8.51 Geneticists at Duke University Medical Center have identified the E2F1 transcription factor as an important component of cell proliferation control (*Nature*, Sept. 23, 1993). The researchers induced DNA synthesis in two batches of serum-starved cells. Each cell in one batch was micro-injected with the E2F1 gene, whereas the cells in the second batch (the controls) were not exposed to E2F1. After 30 hours, the number of cells in each batch that exhibited altered growth was determined. The results of the experiment are summarized in the table.

	Control	*E2F1 Treated Cells*
Total Number of Cells	158	92
Number of Growth-Altered Cells	15	41

Source: Johnson, D. G., et al. "Expression of transcription factor E2F1 induces quiescent to enter S phase." *Nature*, Vol. 365, No. 6444, Sept. 23, 1993, p. 351 (Table 1).

a. Compare the percentages of cells exhibiting altered growth in the two batches with a 90% confidence interval.

b. Use the interval, part **a**, to make an inference about the ability of the E2F1 transcription factor to induce cell growth.

8.52 The nuclear mishap at Three Mile Island near Harrisburg, Pennsylvania, on March 28, 1979, forced many local residents to evacuate their homes—some temporarily, others permanently. To assess the impact of the accident on the area population, a questionnaire was designed and mailed to a sample of 150 households within 2 weeks after the accident occurred. Residents were asked how they felt both before and after the accident about having some of their electricity generated from nuclear power. The summary results are provided in the table on page 394.

	Attitude Toward Nuclear Power			*Totals*
	Favor	*Oppose*	*Indifferent*	
Before Accident	62	35	53	150
After Accident	52	72	26	150

Source: Brown, A., et al. Final Report on a Survey of Three Mile Island Area Residents. Department of Geography, Michigan State University, Aug. 1979.

a. Construct a 99% confidence interval for the difference in the true proportions of Three Mile Island residents who favor nuclear power before and after the accident.

b. Construct a 99% confidence interval for the difference in the true proportions of Three Mile Island residents who oppose nuclear power before and after the accident.

8.53 The *Journal of Fish Biology* (Aug. 1990) reported on a study comparing the prevalence of parasites (tapeworms) found in species of Mediterranean and Atlantic fish. In the Mediterranean Sea, 588 brill were captured and dissected, and 211 were found to be infected by the parasite. In the Atlantic Ocean, 123 brill were captured and dissected, and 26 were found to be infected. Compare the proportions of infected brill at the two capture sites using a 90% confidence interval. Interpret the interval.

8.54 The Egyptian National Scientific and Technical Information Network (ENSTINET) operates an on-line database search service of existing U.S. databases. A database "search" occurs when a specific request is executed by ENSTINET during a single session. In situations when the search produces irrelevant or no output, the search is "rerun." According to *Information Processing & Management* (Vol. 22, No. 3, 1986), ENSTINET performed 342 database searches in 1982, of which 40 were rerun. In 1985, 83 of 2,117 searches required reruns. Assuming that the two samples of database searches are independent and random, construct a 95% confidence interval for the difference between the proportions of database search reruns performed by ENSTINET in 1982 and 1985. Interpret the interval.

8.55 Refer to the marketing research study of consulting engineering services to industrial firms in the Midwest, Exercise 2.49. Forty of the firms surveyed (20 large and 20 small firms) indicated they have no need for outside consulting engineering services (*Journal of the Boston Society of Civil Engineers*, Vol. 71, 1985). The primary reason cited by the "nonneeders" was that they obtained consulting assistance from corporate headquarters whenever necessary. However, twice as many large firms (12) as small firms (6) cited this reason. Establish a 90% confidence interval for the difference in the percentages of large and small industrial firms that cite assistance from corporate headquarters as the primary reason why they have no need for outside consulting engineering services.

8.10 Estimation of a Population Variance

In the previous sections, we considered interval estimates for population means and proportions. In this section, we discuss confidence intervals for a population variance σ^2, and, in Section 8.11, confidence intervals for the ratio of two variances, σ_1^2/σ_2^2. Unlike means and proportions, the pivotal statistics for variances do not possess a normal (z) distribution or a t distribution. In addition, certain assumptions are required regardless of the sample size.

Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. From Theorem 7.4, we know that

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

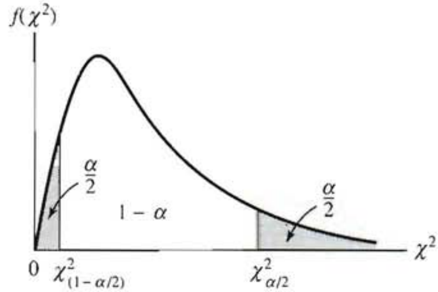
possesses a chi-square distribution with $(n-1)$ degrees of freedom. Confidence intervals for σ^2 are based on the pivotal statistic, χ^2.

Recall that upper-tail areas of the chi-square distribution have been tabulated and are given in Table 8 of Appendix II. Unlike the z and t distributions, the chi-square distribution is not symmetric about 0. To find values of χ^2 that locate an area a in the lower tail of the distribution, we must find χ^2_{1-a}, where $P(\chi^2 > \chi^2_{1-a}) = 1 - a$. For example, the value of χ^2 that places an area $a = .05$ in the lower tail of the distribution when df $= 9$ is $\chi^2_{1-a} = \chi^2_{.95} = 3.32511$ (see Table 8 of Appendix II). We use this fact to write a probability statement for the pivotal statistic χ^2:

$$P(\chi^2_{1-\alpha/2} \le \chi^2 \le \chi^2_{\alpha/2}) = 1 - \alpha$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{(1-\alpha/2)}$ are tabulated values of χ^2 that place a probability of $\alpha/2$ in each tail of the chi-square distribution (see Figure 8.13).

FIGURE 8.13 ▶
The location of $\chi^2_{(1-\alpha/2)}$ and $\chi^2_{\alpha/2}$ for a chi-square distribution



Substituting $[(n-1)s^2]/\sigma^2$ for χ^2 in the probability statement and performing some simple algebraic manipulations, we obtain

$$P\left(\chi^2_{(1-\alpha/2)} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2}\right)$$
$$= P\left(\frac{\chi^2_{(1-\alpha/2)}}{(n-1)s^2} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{\alpha/2}}{(n-1)s^2}\right)$$
$$= P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}\right) = 1 - \alpha$$

Thus, a $(1-\alpha)100\%$ confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

A $(1 - \alpha)100\%$ Confidence Interval for a Population Variance, σ^2

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{(1-\alpha/2)}$ are values of χ^2 that locate an area of $\alpha/2$ to the right and $\alpha/2$ to the left, respectively, of a chi-square distribution based on $(n - 1)$ degrees of freedom.

Assumption: The population from which the sample is selected has an approximate normal distribution.

Note that the estimation technique applies to either large or small n, and that the assumption of normality is required in either case.

EXAMPLE 8.15

A quality control supervisor in a cannery knows that the exact amount each can contains will vary, since there are certain uncontrollable factors that affect the amount of fill. The mean fill per can is important, but equally important is the variation σ^2 of the amount of fill. If σ^2 is large, some cans will contain too little and others too much. To estimate the variation of fill at the cannery, the supervisor randomly selects 10 cans and weighs the contents of each. The following weights (in ounces) are obtained:

7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02

Constuct a 90% confidence interval for the true variation in fill of cans at the cannery.

Solution

The supervisor wishes to estimate σ^2, the population variance of the amount of fill. A $(1 - \alpha)100\%$ confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

For the confidence interval to be valid, we must assume that the sample of observations (amounts of fill) is selected from a normal population.

To compute the interval, we need to calculate either the sample variance s^2 or the sample standard deviation s. Descriptive statistics for the sample data are provided in the SAS printout shown in Figure 8.14. The value of s, shaded in Figure 8.14, is $s = .043$.

Now, $(1 - \alpha) = .90$ and $\alpha/2 = .10/2 = .05$. Therefore, the tabulated values $\chi^2_{.05}$ and $\chi^2_{.95}$ for $(n - 1) = 9$ df (obtained from Table 8, Appendix II) are

$$\chi^2_{.05} = 16.9190 \quad \text{and} \quad \chi^2_{.95} = 3.32511$$

FIGURE 8.14 ▶
SAS descriptive statistics for Example 8.15

```
Variable=FILL

                    Moments

N                 10   Sum Wgts          10
Mean           7.989   Sum            79.89
Std Dev     0.043063   Variance    0.001854
Skewness     -0.8538   Kurtosis    0.479371
USS         638.2579   CSS          0.01669
CV          0.539032   Std Mean    0.013618
T:Mean=0    586.6587   Prob>|T|      0.0001
Sgn Rank        27.5   Prob>|S|      0.0020
Num ^= 0          10

               Quantiles(Def=5)

  100% Max      8.04         99%       8.04
   75% Q3       8.02         95%       8.04
   50% Med     7.995         90%      8.035
   25% Q1       7.96         10%       7.93
    0% Min       7.9          5%        7.9
                              1%        7.9
  Range         0.14
  Q3-Q1         0.06
  Mode          7.96
```

Substituting these values into the formula, we obtain

$$\frac{(10 - 1)(.043)^2}{16.9190} \leq \sigma^2 \leq \frac{(10 - 1)(.043)^2}{3.32511}$$

$$.00098 \leq \sigma^2 \leq .00500$$

We are 90% confident that the true variance in amount of fill of cans at the cannery falls between .00098 and .00500. The quality control supervisor could use this interval to check whether the variation of fill at the cannery is too large and in violation of government regulatory specifications.

EXAMPLE 8.16

Refer to Example 8.15. Find a 90% confidence interval for σ, the true standard deviation of the can weights.

Solution

A confidence interval for σ is obtained by taking the square roots of the lower and upper endpoints of a confidence interval for σ^2. Thus, the 90% confidence interval is

$$\sqrt{.00098} \leq \sigma \leq \sqrt{.00500}$$

$$.031 \leq \sigma \leq .071$$

We are 90% confident that the true standard deviation of can weights is between .031 and .071 ounce.

EXERCISES

8.56 For each of the following combinations of a and degrees of freedom (df), find the value of chi-square, χ^2_a, that places an area a in the upper tail of the chi-square distribution:
a. $a = .05$, df = 7 b. $a = .10$, df = 16 c. $a = .01$, df = 10
d. $a = .025$, df = 8 e. $a = .005$, df = 5

8.57 *Jitter* is a term used to describe the variation in conduction time of a modular pulsed-water power system. Low throughput jitter is critical to successful waterline technology. An investigation of throughput jitter in the plasma opening switch of a prototype system (*Journal of Applied Physics*, Sept. 1993) yielded the following descriptive statistics on conduction time for $n = 18$ trials:

$\bar{y} = 334.8$ nanoseconds $s = 6.3$ nanoseconds

(Conduction time is defined as the length of time required for the downstream current to equal 10% of the upstream current.)

a. Construct a 95% confidence interval for the true standard deviation of conduction times of the prototype system.
b. A system is considered to have low throughput jitter if the true conduction time standard deviation is less than 7 nanoseconds. Does the prototype system satisfy this requirement? Explain.

8.58 Refer to the *IEEE Transactions* (June 1990) study of a new hybrid algorithm for solving polynomial 0–1 mathematical programs, Exercise 8.31. A SAS printout giving descriptive statistics for the sample of 52 solution times is reproduced here. Use this information to compute an approximate 95% confidence interval for the variance of the solution times. Interpret the result.

Analysis Variable : CPU

N Obs	N	Mean	Variance	Std Dev
52	52	0.8121923	2.2643035	1.5047603

8.59 An interlaboratory study was conducted to determine the variation in the measured level of polychlorinated biphenyls (PCBs) in environmentally contaminated sediments (*Analytical Chemistry*, Nov. 1985). Samples of sediment from New Bedford Harbor (Massachusetts) known to be contaminated with PCBs were collected and aliquot solutions prepared. For one part of the study, the PCB concentration in each of a random sample of five aliquots was determined by a single laboratory using the Webb–McCall procedure. The analysis yielded a mean PCB concentration of 56 mg/kg and a standard deviation of .45 mg/kg. Find a 90% confidence interval for the variance in the PCB levels of contaminated sediment, determined using the Webb–McCall procedure.

8.60 An experiment was conducted to investigate the precision of measurements of a saturated solution of iodine after an extended period of continuous stirring. The data shown in the table represent $n = 10$ iodine concentration measurements on the same solution. The population variance σ^2 measures the variability—i.e., the precision—of a measurement. Use the information in the accompanying MINITAB printout to find a 95% confidence interval for σ^2. Interpret the result.

Run	*Concentration*	*Run*	*Concentration*
1	5.507	6	5.527
2	5.506	7	5.504
3	5.500	8	5.490
4	5.497	9	5.500
5	5.506	10	5.497

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
conctrat	10	5.5034	5.5020	5.5021	0.0098	0.0031

	MIN	MAX	Q1	Q3
conctrat	5.4900	5.5270	5.4970	5.5062

8.61 Geologists analyze fluid inclusions in rock to infer the compositions of fluids present when the rocks crystallized. A new technique, called laser Raman microprobe (LRM) spectroscopy, has been developed for this purpose. An experiment was conducted to estimate the precision of the LRM technique (*Applied Spectroscopy*, Feb. 1986). A chip of natural Brazilian quartz with several artificially produced fluid inclusions was subjected to LRM spectroscopy. The amount of liquid carbon dioxide (CO_2) present in the inclusion was recorded for the same inclusion on four different days. The data (in mole percentage) follow:

86.6 84.6 85.5 85.9

a. Obtain an estimate of the precision of the LRM technique by constructing a 99% confidence interval for the variation in the CO_2 concentration measurements.
b. What assumption is required for the interval estimate to be valid?

8.11 Estimation of the Ratio of Two Population Variances

The common statistical procedure for comparing two population variances, σ_1^2 and σ_2^2, makes an inference about the ratio σ_1^2/σ_2^2. This is because the sampling distribution of the estimator of σ_1^2/σ_2^2 is well known when the samples are randomly and independently selected from two normal populations. Under these assumptions, a confidence interval for σ_1^2/σ_2^2 is based on the pivotal statistic

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2}$$

where χ_1^2 and χ_2^2 are chi-square random variables with $\nu_1 = (n_1 - 1)$ and $\nu_2 = (n_2 - 1)$ degrees of freedom, respectively. Substituting $(n - 1)s^2/\sigma^2$ for χ^2 (see

Theorem 7.4), we may write

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_2} = \frac{\dfrac{(n_1-1)s_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)s_2^2}{\sigma_2^2}\Big/(n_2-1)}$$

$$= \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

$$= \left(\frac{s_1^2}{s_2^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$$

From Definition 7.2 we know that F has an F distribution with $\nu_1 = (n_1 - 1)$ numerator degrees of freedom and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom. An F distribution can be symmetric about its mean, skewed to the left, or skewed to the right; its exact shape depends on the degrees of freedom associated with s_1^2 and s_2^2, i.e., $(n_1 - 1)$ and $(n_2 - 1)$.

To establish lower and upper confidence limits for σ_1^2/σ_2^2, we need to be able to find tabulated F values corresponding to the tail areas of the distribution. The *upper-tail* F values can be found in Tables 9, 10, 11, and 12 of Appendix II for $a = .10$, $.05$, $.025$, and $.01$, respectively. Table 10 of Appendix II is partially reproduced in Table 8.8. The columns of Tables 9–12 of Appendix II correspond to various degrees of freedom for the numerator sample variance, s_1^2, in the pivotal statistic, whearas the rows correspond to the degrees of freedom for the denominator sample variance, s_2^2. For example, with numerator degrees of freedom $\nu_1 = 7$ and denominator degrees of freedom $\nu_2 = 9$, we have $F_{.05} = 3.29$ (shaded in Table 8.8). Thus, $a = .05$ is the tail area to the right of 3.29 in the F distribution with 7 numerator df and 9 denominator df, i.e., $P(F > F_{.05}) = .05$.

Lower-tail values of the F distribution are not given in Tables 9–12 of Appendix II. However, it can be shown (proof omitted) that

$$F_{1-a(\nu_1,\nu_2)} = \frac{1}{F_{a(\nu_2,\nu_1)}}$$

where $F_{1-a(\nu_1,\nu_2)}$ is the F value that cuts off an area a in the *lower* tail of an F distribution based on ν_1 numerator and ν_2 denominator degrees of freedom, and $F_{a(\nu_2,\nu_1)}$ is the F value that cuts off an area a in the *upper* tail of an F distribution based on ν_2 numerator and ν_1 denominator degrees of freedom. For example, suppose we want to find the value that locates an area $a = .05$ in the *lower* tail of an F distribution with $\nu_1 = 7$ and $\nu_2 = 9$. That is, we want to find $F_{1-a(\nu_1,\nu_2)} = F_{.95(7,9)}$. First, we find the upper-tail values, $F_{.05(9,7)} = 3.68$, from Table 8.8. (Note that we must switch the numerator and denominator degrees of freedom to obtain this value.) Then, we calculate

$$F_{.95(7,9)} = \frac{1}{F_{.05(9,7)}} = \frac{1}{3.68} = .272$$

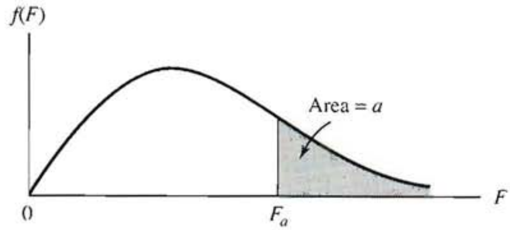


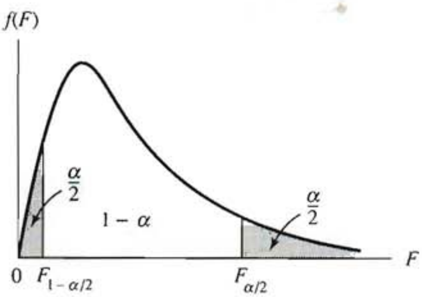
TABLE 8.8 Abbreviated Version of Table 10 of Appendix II: Tabulated Values of the F Distribution, $\alpha = .05$

ν_2 \ ν_1	Numerator Degrees of Freedom								
Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65

Using the notation established previously, we can write a probability statement for the pivotal statistic F (see Figure 8.15):

$$P(F_{1-\alpha/2(\nu_1,\nu_2)} \leq F \leq F_{\alpha/2(\nu_1,\nu_2)}) = 1 - \alpha$$

FIGURE 8.15 ▶
F distribution with $\nu_1 = (n_1 - 1)$ and $\nu_2 = (n_2 - 1)$



Letting $F_L = F_{1-\alpha/2}$ and $F_U = F_{\alpha/2}$, and substituting $(s_1^2/s_2^2)(\sigma_2^2/\sigma_1^2)$ for F, we obtain:

$$P(F_L \le F \le F_U) = P\left[F_L \le \left(\frac{s_1^2}{s_2^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right) \le F_U\right]$$
$$= P\left(\frac{s_2^2}{s_1^2}F_L \le \frac{\sigma_2^2}{\sigma_1^2} \le \frac{s_2^2}{s_1^2}F_U\right)$$
$$= P\left(\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_U} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}\cdot\frac{1}{F_L}\right) = 1 - \alpha$$

or

$$P\left(\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2(\nu_1,\nu_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{1-\alpha/2(\nu_1,\nu_2)}}\right) = 1 - \alpha$$

Replacing $F_{1-\alpha/2(\nu_1,\nu_2)}$ with $1/F_{\alpha/2(\nu_2,\nu_1)}$, we obtain the final form of the confidence interval:

$$P\left(\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2(\nu_1,\nu_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}\cdot F_{\alpha/2(\nu_2,\nu_1)}\right) = 1 - \alpha$$

A $(1 - \alpha)100\%$ Confidence Interval for the Ratio of Two Population Variances, σ_1^2/σ_2^2

$$\frac{s_1^2}{s_2^2}\cdot\frac{1}{F_{\alpha/2(\nu_1,\nu_2)}} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{s_1^2}{s_2^2}F_{\alpha/2(\nu_2,\nu_1)}$$

where $F_{\alpha/2(\nu_1,\nu_2)}$ is the value of F that locates an area $\alpha/2$ in the upper tail of the F distribution with $\nu_1 = (n_1 - 1)$ numerator and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom, and $F_{\alpha/2(\nu_2,\nu_1)}$ is the value of F that locates an area $\alpha/2$ in the upper tail of the F distribution with $\nu_2 = (n_2 - 1)$ numerator and $\nu_1 = (n_1 - 1)$ denominator degrees of freedom.

Assumptions:
1. Both of the populations from which the samples are selected have relative frequency distributions that are approximately normal.
2. The random samples are selected in an independent manner from the two populations.

As in the one-sample case, normal populations must be assumed regardless of the sizes of the two samples.

EXAMPLE 8.17

A firm has been experimenting with two different physical arrangements of its assembly line. It has been determined that both arrangements yield approximately the same average number of finished units per day. To obtain an arrangement that produces

TABLE 8.9 Summary Statistics for Example 8.17

Assembly Line 1	Assembly Line 2
$n_1 = 21$ days	$n_2 = 25$ days
$s_1^2 = 1{,}432$	$s_2^2 = 3{,}761$

greater process control, you suggest that the arrangement with the smaller variance in the number of finished units produced per day be permanently adopted. Two independent random samples yield the results shown in Table 8.9. Construct a 95% confidence interval for σ_1^2/σ_2^2, the ratio of the variances of the number of finished units for the two assembly line arrangements. Based on the result, which of the two arrangements would you recommend?

Solution

First, we must assume that the distributions of the numbers of finished units for the two assembly lines are both approximately normal. Since we want a 95% confidence interval, the value of $\alpha/2$ is .025, and we need to find $F_{.025(\nu_1,\nu_2)}$ and $F_{.025(\nu_2,\nu_1)}$. The sample sizes are $n_1 = 21$ and $n_2 = 25$; thus, $F_{.025(\nu_1,\nu_2)}$ is based on $\nu_1 = (n_1 - 1) = 20$ numerator df and $\nu_2 = (n_2 - 1) = 24$ denominator df. Consulting Table 11 of Appendix II, we obtain $F_{.025(20,24)} = 2.33$. In contrast, $F_{.025(\nu_1,\nu_2)}$ is based on $\nu_2 = (n_2 - 1) = 24$ numerator df and $\nu_1 = (n_1 - 1) = 20$ denominator df; hence (from Table 11 of Appendix II), $F_{.025(24,20)} = 2.41$. Substituting the values for s_1^2, s_2^2, $F_{.025(\nu_1,\nu_2)}$ and $F_{.025(\nu_2,\nu_1)}$ into the confidence interval formula, we have

$$\frac{1{,}432}{3{,}761}\left(\frac{1}{2.33}\right) \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1{,}432}{3{,}761}(2.41)$$

$$.163 \le \frac{\sigma_1^2}{\sigma_2^2} \le .918$$

We estimate with 95% confidence that the ratio σ_1^2/σ_2^2 of the true population variances will fall between .163 and .918. Since all the values within the interval (.163,.918) are less than 1.0, we can be confident that the variance in the number of units finished on line 1 (as measured by σ_1^2) is less than the corresponding variance for line 2 (as measured by σ_2^2).

EXERCISES

8.62 Find F_a for an F distribution with 15 numerator df and 12 denominator df for the following values of a:
a. $a = .025$ **b.** $a = .05$ **c.** $a = .10$

8.63 Find $F_{.05}$ for an F distribution with:
a. Numerator df = 7, denominator df = 25 **b.** Numerator df = 10, denominator df = 8
c. Numerator df = 30, denominator df = 60 **d.** Numerator df = 15, denominator df = 4

8.64 In *Environmental Science & Technology* (Oct. 1993), scientists reported on a study of the transport and transformation of PCDD, a pollutant emitted from solid waste incineration, motor vehicles, steel mills, and metal production. Ambient air specimens were collected over several different days at two locations in Swden: Rörvik (11 days) and Gothenburg (3 days). The level of PCDD (measured in pg/m^3) detected in each specimen is recorded here. Use interval estimation to compare the variation in PCDD levels at the two locations. Draw an inference from the analysis.

Rörvik				*Gothenburg*		
2.38	3.03	1.44	.47	.50	.61	.90
.50	.22	.26	.31			
.46	1.09	2.14				

Source: Tysklind, M., et al. "Atmospheric transport and transformation of polychlorinated dibenzo-p-dioxins and dibenzofurans." *Environmental Science & Technology*, Vol. 27, No. 10, Oct. 1993, p. 2193 (Table III).

8.65 Refer to the *Journal of Structural Engineering* (Feb. 1986) experiment with epoxy-repaired truss joints, Exercise 8.35. The data are reproduced here for convenience. Construct a 90% confidence interval for the ratio of the shear stress variances of epoxy-repaired truss joints for the two species of wood. Based on this interval, is there evidence to indicate that the two shear stress variances differ? Explain.

	Southern Pine	*Ponderosa Pine*
Sample Size	100	47
Mean Shear Stress, psi	1,312	1,352
Standard Deviation	422	271

Source: Avent, R. R. "Design criteria for epoxy repair of timber structures." *Journal of Structural Engineering*, Vol. 112, No. 2, Feb. 1986, pp. 232.

8.66 Refer to the strength and capacity guidelines for manual materials handling activities, Exercise 2.52. The guidelines were established by observing the maximum weight that random samples of men and women can safely lift from the floor to knuckle height (*Human Factors*, June 1980). When lifting at the rate of 1 lift per minute, males lifted a mean maximum weight of 30.25 kilograms (kg) with a standard deviation of 8.56 kg, whereas the mean and standard deviation for females was 19.79 kg and 3.11 kg, respectively.

a. Assuming the sample consisted of 60 males and 60 females, construct a 90% confidence interval for the ratio of the variances of the maximum weights that can safely be lifted by males and females.

b. What assumptions must be satisfied to ensure the validity of the interval estimate of part **a**?

8.67 Refer to the cancer death rate increases for fluoridated and nonfluoridated cities given in Exercise 8.36. The data are reproduced here for convenience. Find a 95% confidence interval for the ratio of the variances of the cancer death rate increases for the two groups of cities. Based on the interval, does it appear that the assumption of equal variances required to conduct the analysis of Exercise 8.36 is satisfied?

Fluoridated City	Annual Increase in Cancer Death Rate	*Nonfluoridated* City	Annual Increase in Cancer Death Rate
Chicago	1.0640	Los Angeles	.8875
Philadelphia	1.4118	Boston	1.7358
Baltimore	2.1115	New Orleans	1.0165
Cleveland	1.9401	Seattle	.4923
Washington	3.8772	Cincinnati	4.0155
Milwaukee	−.4561	Atlanta	−1.1744
St. Louis	4.8359	Kansas City	2.8132
San Francisco	1.8875	Columbus	1.7451
Pittsburgh	4.4964	Newark	−.5676
Buffalo	1.4045	Portland	2.4471

Source: Maritz, J. S., and Jarrett, R. C. "The use of statistics to examine the association between fluoride in drinking water and cancer death rates." *Applied Statistics*, Vol. 32, No. 2, 1983, pp. 97–101.

8.68 Refer to the PCB study described in Exercise 8.59. Recall that level of PCB was measured in each of a sample of five aliquots using the Webb–McCall procedure. Another sample of five aliquots of sediment was measured for PCBs using a diffferent procedure, called the Aroclor Standard comparison. Summary statistics on PCB concentration for the two samples are given in the table.

	Webb–McCall	*Aroclor Standard*
Sample Size	5	5
Mean PCB Concentration, mg/kg	56	60
Standard Deviation	.45	.89

Source: Alford-Stevens, A. L., Budde, W. L., and Bellar, T. A. "Interlaboratory study on determination of polychlorinated biphenyls in environmentally contaminated sediments." *Analytical Chemistry*, Vol. 57, No. 13, Nov. 1985, p. 2454. Reprinted with permission from *Analytical Chemistry*. Copyright 1985 American Chemical Society.

a. Construct a 90% confidence interval for the ratio of the variances in the PCB levels measured by the two techniques.

b. What assumptions are required for the interval estimate to be valid?

8.12 Choosing the Sample Size

One of the first problems encountered when applying statistics in a practical situation is to decide on the number of measurements to include in the sample(s). The solution to this problem depends on the answers to the following questions: Approximately how wide do you want your confidence interval to be? What confidence coefficient do you require?

You have probably noticed that the half-widths of many of the confidence intervals presented in Sections 8.5–8.11 are functions of the sample size and the estimated standard error of the point estimator involved. For example, the half-width H of the small-sample confidence interval for μ is

$$H = t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

where $t_{\alpha/2}$ depends on the sample size n, and s is a statistic computed from the sample data. Since we will not know s before selecting the sample and we have no control over its value, the easiest way to decrease the width of the confidence interval is to increase the sample size n. Generally speaking, the larger the sample size, the more information you will acquire and the smaller will be the width of the confidence interval. We illustrate the procedure for selecting the sample size in Examples 8.18 and 8.19.

EXAMPLE 8.18

As part of a Department of Energy (DOE) survey, American families will be randomly selected and questioned about the amount of money they spent last year on home heating oil or gas. Of particular interest to the DOE is the average amount μ spent last year on heating fuel. If the DOE wants the estimate of μ to be correct to within \$10 with a confidence coefficient of .95, how many families should be included in the sample?

Solution

The DOE wants to obtain an interval estimate of μ, with confidence coefficient equal to $(1 - \alpha) = .95$ and half-width of the interval equal to 10. The half-width of a large-sample confidence interval for μ is

$$H = z_{\alpha/2}\sigma_{\bar{y}} = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

In this example, we have $H = 10$ and $z_{\alpha/2} = z_{.025} = 1.96$. To solve the equation for n, we need to know σ. But, as will usually be the case in practice, σ is unknown. Suppose, however, that the DOE knows from past records that the yearly amounts spent on heating fuel have a range of approximately \$520. Then we could approximate σ by letting the range equal 4σ.* Thus,

$$4\sigma \approx 520 \quad \text{or} \quad \sigma \approx 130$$

Solving for n, we have

$$H = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \quad \text{or} \quad 10 = 1.96\left(\frac{130}{\sqrt{n}}\right)$$

*From the Empirical Rule, we expect about 95% of the observations to fall between $\mu - 2\sigma$ and $\mu + 2\sigma$. Thus,

$$\text{Range} \approx (\mu + 2\sigma) - (\mu - 2\sigma) = 4\sigma$$

or

$$n = \frac{(1.96)^2(130)^2}{(10)^2} \approx 650$$

Consequently, the DOE will need to elicit responses from 650 American families to estimate the mean amount spent on home heating fuel last year to within \$10 with 95% confidence. Since this would require an extensive and costly survey, the DOE might decide to allow a larger half-width (say, $H = 15$ or $H = 20$) to reduce the sample size, or the DOE might decrease the desired confidence coefficient. The important point is that the experimenter can obtain an idea of the sampling effort necessary to achieve a specified precision in the final estimate by determining the approximate sample size *before* the experiment is begun.

EXAMPLE 8.19

A production supervisor suspects a difference exists between the proportions p_1 and p_2 of defective items produced by two different machines. Experience has shown that the proportion defective for each of the two machines is in the neighborhood of .03. If the supervisor wants to estimate the difference in the proportions correct to within .005 with probability .95, how many items must be randomly sampled from the production of each machine? (Assume that you want $n_1 = n_2 = n$.)

Solution

Since we want to estimate $(p_1 - p_2)$ with a 95% confidence interval, we will use $z_{\alpha/2} = z_{.025} = 1.96$. For the estimate to be correct to within .005, the half-width of the confidence interval must equal .005. Then, letting $p_1 = p_2 = .03$ and $n_1 = n_2 = n$, we find the required sample size per machine by solving the following equation for n:

$$H = z_{\alpha/2}\sigma_{(\hat{p}_1-\hat{p}_2)} \quad \text{or} \quad H = z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

$$.005 = 1.96\sqrt{\frac{(.03)(.97)}{n} + \frac{(.03)(.97)}{n}}$$

$$.005 = 1.96\sqrt{\frac{2(.03)(.97)}{n}}$$

$$n = \frac{(1.96)^2(2)(.03)(.97)}{(.005)^2} \approx 8{,}944$$

You can see that this may be a tedious sampling procedure. If the supervisor insists on estimating $(p_1 - p_2)$ correct to within .005 with probability equal to .95, approximately 9,000 items will have to be inspected for each machine.

You can see from the calculations in Example 8.19 that $\sigma_{(\hat{p}_1-\hat{p}_2)}$ (and hence the solution, $n_1 = n_2 = n$) depends on the actual (but unknown) values of p_1 and p_2. In fact, the required sample size $n_1 = n_2 = n$ is largest when $p_1 = p_2 = .5$. Therefore, if you have no prior information on the approximate values of p_1 and p_2, use $p_1 = p_2 = .5$ in the formula for $\sigma_{(\hat{p}_1 - \hat{p}_2)}$. If p_1 and p_2 are in fact close to .5, then the resulting values of n_1 and n_2 will be correct. If p_1 and p_2 differ substantially from .5, then your solutions for n_1 and n_2 will be larger than needed. Consequently, using $p_1 = p_2 = .5$ when solving for n_1 and n_2 is a conservative procedure because the sample sizes n_1 and n_2 will be at least as large as (and probably larger than) needed.

The formulas for calculating the sample size(s) required for estimating the parameters μ, $(\mu_1 - \mu_2)$, p, and $(p_1 - p_2)$ are summarized in the following boxes. Sample size calculations for variances require more sophisticated techniques and are beyond the scope of this text.

Choosing the Sample Size for Estimating a Population Mean μ to Within H Units with Probability $(1 - \alpha)$

$$n = \left(\frac{z_{\alpha/2}\sigma}{H}\right)^2$$

[*Note:* The population standard deviation σ will usually have to be approximated.]

Choosing the Sample Sizes for Estimating the Difference $(\mu_1 - \mu_2)$ Between a Pair of Population Means Correct to Within H Units with Probability $(1 - \alpha)$

$$n_1 = n_2 = \left(\frac{z_{\alpha/2}}{H}\right)^2(\sigma_1^2 + \sigma_2^2)$$

where n_1 and n_2 are the numbers of observations sampled from each of the two populations, and σ_1^2 and σ_2^2 are the variances of the two populations.

Choosing the Sample Size for Estimating a Population Proportion p to Within H Units with Probability $(1 - \alpha)$

$$n = \left(\frac{z_{\alpha/2}}{H}\right)^2 pq$$

where p is the value of the population proportion that you are attempting to estimate and $q = 1 - p$.

[*Note:* This technique requires previous estimates of p and q. If none are available, use $p = q = .5$ for a conservative choice of n.]

Choosing the Sample Sizes for Estimating the Difference $(p_1 - p_2)$ Between Two Population Proportions to Within H Units with Probability $(1 - \alpha)$

$$n_1 = n_2 = \left(\frac{z_{\alpha/2}}{H}\right)^2 (p_1q_1 + p_2q_2)$$

where p_1 and p_2 are the proportions for populations 1 and 2, respectively, and n_1 and n_2 are the numbers of observations to be sampled from each population.

EXERCISES

8.69 *Cost Engineering* (Oct. 1988) reports on a study of the percentage difference between the low bid and the engineer's estimate of the cost for building contracts (see Exercise 7.19). For contracts with four bidders, the mean percentage error is $\mu = -7.02$ and the standard deviation is $\sigma = 24.66$. Suppose you want to estimate the mean percentage error for building contracts with five bidders. How many five-bidder contracts must be sampled to estimate with 90% confidence the mean to witihin 5 percentage points of its true value? Assume that the standard deviation for five-bidder contracts is approximately equal to the standard deviation for four-bidder contracts.

8.70 Refer to the *Human Factors* study on the use of color brightness as a body orientation cue, Exercise 8.49. How many subjects are required for a similar experiment to estimate the true proportion who use a bright color level as a cue to being right-side up to within .05 with 95% confidence? Use the sample proportion calculated in Exercise 8.49 as an estimate of p.

8.71 The federal government requires states to certify that they are enforcing the 55-miles-per-hour speed limit and that motorists are driving at that speed. A state is in jeopardy of losing millions of dollars in federal road funds if more than 60% of its vehicles on 55-miles-per-hour highways are exceeding the speed limit. The state highway patrol conducts 70 radar surveys each year at a total of 50 sites to estimate the proportion p of vehicles exceeding 55 miles per hour. Each sample survey involves at least 400 vehicles.

a. How large a sample should be selected at a particular site to estimate p to within 3% with 90% confidence? Last year approximately 60% of all vehicles exceeded 55 miles per hour.

b. The highway patrol also estimates μ, the average speed of vehicles on state highways. Accordingly, it wants to know whether the sample size determined in part a is large enough to also estimate μ to within .25 mile per hour with 90% confidence. Assume that the standard deviation of vehicle speeds is approximately 2 miles per hour. How large a sample should be taken at a particular site to estimate μ with the desired reliability?

8.72 A consumer protection agency wants to compare the work of two electrical contractors to evaluate their safety records. The agency plans to inspect residences in which each of these contractors has done the wiring to estimate the difference in the proportions of residences that are electrically deficient. Suppose the proportions of deficient work are expected to be about .10 for both contractors. How many homes should be inspected to estimate the difference in proportions to within .05 with 90% confidence?

8.73 A large steel corporation conducted an experiment to compare the average iron contents of two consignments of lumpy iron ore. In accordance with industrial standards, n increments of iron ore were randomly selected from each consignment and measured for iron content. From previous experiments, it is known that iron contents vary over a range of roughly 3%. How large should n be if the steel company wants to estimate the difference in mean iron contents of the two consignments correct to within .05% with 95% confidence? [*Hint:* To obtain an approximate value for σ_1 and σ_2, set $\sigma_1 = \sigma_2 = \sigma$ and set Range $= 4\sigma$. Then $3 \approx 4\sigma$ and $\sigma \approx \frac{3}{4}$.]

8.74 *Materials requirements planning (MRP) systems* are computerized planning and control systems for manufacturing operations. Since their introduction in the mid-1960s, MRP systems have been used to manage raw materials and work-in-process inventories while improving customer service. Suppose you want to estimate the proportion p of manufacturing firms that use MRP systems. Approximately how large a sample would be required to estimate p to within .02 with a confidence coefficient of .95? (Use a conservative estimate of $p \approx .5$ in your calculations.)

OPTIONAL EXERCISE

8.75 When determining the sample size required to estimate p, show that the sample size n is largest when $p = .5$.

8.13 Summary

Estimation is a procedure for inferring the value(s) of one (or more) population parameters. An **estimator**, a rule that tells how to calculate a particular **estimate** of a parameter based on information contained in a sample, can be one of two types. A **point estimator** uses the sample data to calculate a single number that serves as an estimate of a population parameter. An **interval estimator** uses the sample data to calculate two numbers that define an interval that is intended to enclose the estimated parameter with some predetermined probability.

Point and interval estimators can be acquired intuitively; it seems reasonable to use sample statistics to estimate the corresponding population parameters (the **method of moments**). In addition, point estimators can be acquired using the **method of maximum likelihood** (Section 8.3) or the **method of least squares** (Chapter 11); interval estimators can be constructed using **pivotal statistics** and the procedure illustrated in Section 8.4. In general, we prefer point estimators that are **unbiased** and possess **minimum variance**, i.e., **minimum variance unbiased estimators** (**MVUE**). For a given **confidence coefficient**, we prefer interval estimators with a mean interval width that is small and subject to a small amount of variation.

We presented a number of point and interval estimators and demonstrated how they can be applied in practical situations. (These results are summarized in Tables 8.10a and 8.10b.) By reviewing the examples, you can see that estimation as a method of inference attempts to answer the question, "What is the value of the parameter θ?" We will approach inference-making from a different point of view in Chapter 9.

TABLE 8.10a Summary of Estimation Procedures: One-Sample Case

Parameter θ	Estimator $\hat{\theta}$	$E(\hat{\theta})$	$\sigma_{\hat{\theta}}$	Approximation to $\sigma_{\hat{\theta}}$	$(1-\alpha)100\%$ Confidence Interval	Sample Size	Additional Assumptions
Mean μ	$\bar{y}$	μ	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$	$\bar{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$	$n \geq 30$	None
					$\bar{y} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ where $t_{\alpha/2}$ is based on $(n-1)$ df	$n < 30$	Normal population
Binomial proportion p	$\hat{p} = \frac{y}{n}$	p	$\sqrt{\frac{pq}{n}}$	$\sqrt{\frac{\hat{p}\hat{q}}{n}}$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$	n large enough so that the interval $\hat{p} \pm 2\sqrt{\frac{\hat{p}\hat{q}}{n}}$ does not contain 0 or 1	None
Variance σ^2	s^2	σ^2	Not needed	Not needed	$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$ where $\chi^2_{\alpha/2}$ and $\chi^2_{(1-\alpha/2)}$ are the tabulated values of χ^2, given in Table 8 of Appendix II, that locate $\alpha/2$ in each tail of the chi-square distribution with $(n-1)$ df, i.e., $P(\chi^2 \geq \chi^2_{\alpha/2}) = \alpha/2$ and $P(\chi^2 \geq \chi^2_{(1-\alpha/2)}) = 1 - \alpha/2$	All n	Normal population

TABLE 8.10b Summary of Estimation Procedures: Two-Sample Case

Parameter θ	Estimator $\hat{\theta}$	$E(\hat{\theta})$	$\sigma_{\hat{\theta}}$	Approximation to $\sigma_{\hat{\theta}}$	$(1-\alpha)100\%$ Confidence Interval	Sample Sizes	Additional Assumptions
$(\mu_1 - \mu_2)$ Difference between population means: Independent samples	$(\bar{y}_1 - \bar{y}_2)$ Difference between sample means	$(\mu_1 - \mu_2)$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$(\bar{y}_1 - \bar{y}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	$n_1 \geq 30$, $n_2 \geq 30$	None
			$\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$	$\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $t_{\alpha/2}$ is based on $(n_1 + n_2 - 2)$ df	Either $n_1 < 30$ or $n_2 < 30$, or both	Both populations normal with equal variances $(\sigma_1^2 = \sigma_2^2)$
$\mu_d = (\mu_1 - \mu_2)$ Difference between population means: Matched pairs	$\bar{d} = \Sigma d_i/n$ Mean of sample differences	μ_d	$\frac{\sigma_d}{\sqrt{n_d}}$	$\frac{s_d}{\sqrt{n_d}}$ where s_d is the standard deviation of the sample of differences	$\bar{d} \pm z_{\alpha/2}\left(\frac{s_d}{\sqrt{n_d}}\right)$	$n_d \geq 30$	None
					$\bar{d} \pm t_{\alpha/2}\left(\frac{s_d}{\sqrt{n_d}}\right)$ where $t_{\alpha/2}$ is based on $(n_d - 1)$ df	$n_d < 30$	Population of differences d_i is normal
$(p_1 - p_2)$ Difference between two binomial parameters	$(\hat{p}_1 - \hat{p}_2)$ Difference between the sample proportions $\hat{p}_1 = y_1/n_1$ and $\hat{p}_2 = y_2/n_2$	$(p_1 - p_2)$	$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$	$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$	n_1 and n_2 large enough so that the intervals $\hat{p}_1 \pm 2\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}}$ and $\hat{p}_2 \pm 2\sqrt{\frac{\hat{p}_2\hat{q}_2}{n_2}}$ do not contain 0 or 1	Independent samples
σ_1^2/σ_2^2 Ratio of population variances	s_1^2/s_2^2 Ratio of sample variances	σ_1^2/σ_2^2	Not needed	Not needed	$\left(\frac{s_1^2}{s_2^2}\right)\frac{1}{F_{\alpha/2(\nu_1,\nu_2)}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \left(\frac{s_1^2}{s_2^2}\right)F_{\alpha/2(\nu_2,\nu_1)}$ where $F_{\alpha/2(\nu_1,\nu_2)}$ and $F_{\alpha/2(\nu_2,\nu_1)}$ are the tabulated values of F (Tables 9, 10, 11, and 12 of Appendix II) that place an area equal to $\alpha/2$ in the upper tail of the F distribution, where $F_{\alpha/2(\nu_1,\nu_2)}$ is based on $\nu_1 = (n_1 - 1)$ numerator and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom, and $F_{\alpha/2(\nu_2,\nu_1)}$ is based on $\nu_2 = (n_2 - 1)$ numerator and $\nu_1 = (n_1 - 1)$ denominator degrees of freedom	All n_1 and n_2	Independent samples from two normal populations

SUPPLEMENTARY EXERCISES

8.76 What do college recruiters think are the most important topics to be covered in a job interview? To answer this and other questions, Taylor and Sniezek elicited the opinions of recruiters interviewing at a small midwestern college and a large midwestern university (*Journal of Occupational Psychology*, 1984). Recruiters were asked to rate on a 105-point scale the importance of each in a list of 25 interview topics [where 0 = least important (can sometimes be omitted without hurting the interview), 52.5 = average importance (can sometimes be omitted without hurting the interview), and 105 = most important (can never be omitted without hurting the interview)]. The topic concerning "applicant's skill in communicating ideas to others" received the highest ratings of the $n = 58$ college recruiters who returned the questionnaire. The sample mean rating and sample standard deviation for this topic were $\bar{y} = 84.84$ and $s = 15.67$, respectively.

a. Give a point estimate for the true mean rating of the importance of "applicant's skill in communicating ideas to others" by all college recruiters.

b. Use the sample information to construct a 95% confidence interval for the true mean rating.

c. What is the confidence coefficient for the interval of part **b**? Interpret this value.

8.77 When new instruments are developed to perform chemical analyses of products (food, medicine, etc.), they are usually evaluated with respect to two criteria: accuracy and precision. *Accuracy* refers to the ability of the instrument to identify correctly the nature and amounts of a product's components. *Precision* refers to the consistency with which the instrument will identify the components of the same material. Thus, a large variability in the identification of a single sample of a product indicates a lack of precision. Suppose a pharmaceutical firm is considering two brands of an instrument designed to identify the components of certain drugs. As part of a comparison of precision, ten test-tube samples of a well-mixed batch of a drug are selected and then five are analyzed by instrument A and five by instrument B. The data shown in the table are the percentages of the primary component of the drug given by the instruments. A SAS printout giving descriptive statistics follows.

Instrument A	43	48	37	52	45
Instrument B	46	49	43	41	48

```
Analysis Variable : READING

---------------------------- INSTRMNT=A -------------------------

N Obs    N          Minimum        Maximum            Mean         Std Dev
--------------------------------------------------------------------------
    5    5     37.0000000     52.0000000      45.0000000       5.6124861
--------------------------------------------------------------------------


---------------------------- INSTRMNT=B -------------------------

N Obs    N          Minimum        Maximum            Mean         Std Dev
--------------------------------------------------------------------------
    5    5     41.0000000     49.0000000      45.4000000       3.3615473
--------------------------------------------------------------------------
```

a. Construct a 90% confidence interval to compare the precision of the two instruments.

b. Based on the interval estimate of part **a**, what would you infer about the precision of the two instruments?
c. What assumptions must be satisfied to ensure the validity of any inferences derived from the estimate?

8.78 A regional computer center wants to evaluate the performance of its disk memory system. One measure of performance is the average time between failures of a disk drive. Since the computer center operates two disk drives, it wants to compare the mean times between failures of the two disk drives. Independent random samples of $n_1 = 10$ and $n_2 = 15$ failures produced the following statistics:

Disk Drive 1	*Disk Drive 2*
$\bar{y}_1 = 92$ hours	$\bar{y}_2 = 108$ hours
$s_1 = 16$ hours	$s_2 = 12$ hours

Which of the two disk drives appears to give better performance?

8.79 According to a report by the U.S. surgeon general, electrical engineers have the lowest smoking rate among all workers surveyed (*IEEE Spectrum*, Apr. 1986). Only 16% of the male electrical engineers in the sample smoke cigarettes regularly. How many male electrical engineers must be sampled to estimate the proportion of all male electrical engineers who smoke regularly to within 3% of its true value with 95% confidence?

8.80 The pesticide Temik is used for controlling insects that feed on potatoes, oranges, and other crops. According to federal standards, drinking water wells with levels of Temik above 1 part per billion are considered contaminated. The accompanying table lists the results of tests for Temik contamination conducted in five states over the past few years. For each state, construct a 95% confidence interval for the true proportion of wells contaminated with Temik. Interpret the results.

State	*Number of Wells Tested*	*Number of Contaminated Wells*
New York	10,500	2,750
Wisconsin	700	105
Maine	124	82
Florida	825	4
Virginia	76	17

Source: *Orlando Sentinel*, July 4, 1983.

8.81 A machine used to fill beer cans must operate so that the amount of beer actually dispensed varies very little. If too much beer is released, the cans will overflow, causing waste. If too little beer is released, the cans will not contain enough beer, causing complaints from customers. A random sample of the fills for 20 cans yielded a standard deviation of .07 ounce. Estimate the true variance of the fills using a 95% confidence interval.

8.82 Refer to the LRM spectroscopy experiment described in Exercise 8.61. The amount of liquid CO_2 present in each of two different fluid inclusions (named FREO and FRITZ) was measured on each of four randomly selected days. The data are reproduced in the table. Use interval estimation to compare the mean difference between the CO_2 concentrations (in mole percentage) of the two fluid inclusions.

Day	*Inclusion FREO*	*Inclusion FRITZ*
1	86.6	83.8
2	84.6	85.3
3	85.5	84.6
4	85.9	83.4

Source: Wopenka, B., and Pasteris, J. D. "Limitations to quantitative analysis of fluid inclusions in geological samples by laser Raman microprobe spectroscopy." *Applied Spectroscopy*, Vol. 40, No. 2, Feb. 1986, p. 149.

8.83 Some power plants are located near rivers or oceans so that the available water can be used for cooling the condensers. As part of an environmental impact study, suppose a power company wants to estimate the difference in mean water temperature between the discharge of its plant and the offshore waters. How many sample measurements must be taken at each site to estimate the true difference between means to within .2°C with 95% confidence? Assume the range in readings will be about 4°C at each site and the same number of readings will be taken at each site.

8.84 A study was conducted to compare the attitudes of American and Soviet teenagers on nuclear war (*New England Journal of Medicine*, Aug. 18, 1988). A team of American and Soviet researchers surveyed 3,370 public school students in Maryland and 2,148 students in central Russia. One question asked whether the students believe a nuclear war will occur in their lifetime. Forty-two percent of the Maryland students and 9% of the Russian students responded affirmatively.

a. Calculate a 99% confidence interval for the difference between the proportions of Maryland and Russian students who believe that a nuclear war will occur in their lifetime. Interpret the interval.

b. How could the width of the interval of part **a** be reduced?

c. Although Maryland students were recruited randomly for the study, there is speculation that the Soviet students were selected much more carefully. How could the nonrandom Soviet sample bias the results obtained in part **a**?

8.85 Two alloys, A and B, are used in the manufacture of steel bars. Suppose a steel producer wants to compare the two alloys on the basis of average load capacity, where the load capacity of a steel bar is defined as the maximum load (weight in tons) it can support without breaking. Steel bars containing alloy A and steel bars containing alloy B were randomly selected and tested for load capacity. The results are summarized in the accompanying table.

Alloy A	*Alloy B*
$n_1 = 11$	$n_2 = 17$
$\bar{y}_1 = 43.7$	$\bar{y}_2 = 48.5$
$s_1^2 = 24.4$	$s_2^2 = 19.9$

a. Find a 99% confidence interval for the difference between the true average load capacities for the two alloys.

b. For the interval of part **a** to be valid, what assumptions must be satisfied?

c. Interpret the interval of part **a**. Can you conclude that the average load capacities for the two alloys are different?

d. How many steel bars of each type should be sampled to estimate the true difference in average load capacities to within 2 tons with 99% confidence? (Assume $n_1 = n_2 = n$.)

OPTIONAL SUPPLEMENTARY EXERCISES

8.86 Let $\bar{y}_1$ be the mean of a random sample of n_1 observations from a Poisson distribution with mean λ_1, and let $\bar{y}_2$ be the mean of a random sample of n_2 observations from a Poisson distribution with mean λ_2. Assume the samples are independent.

a. Show that $(\bar{y}_1 - \bar{y}_2)$ is an unbiased estimator of $(\lambda_1 - \lambda_2)$.

b. Find $V(\bar{y}_1 - \bar{y}_2)$. How could you estimate this variance?

c. Construct a large-sample $(1 - \alpha)100\%$ confidence interval for $(\lambda_1 - \lambda_2)$. [*Hint:* Consider

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - (\lambda_1 - \lambda_2)}{\sqrt{\dfrac{\bar{y}_1}{n_1} + \dfrac{\bar{y}_2}{n_2}}}$$

as a pivotal statistic.]

8.87 Let $y_1, y_2, \ldots, y_n$ denote a random sample from a uniform distribution with probability density

$$f(y) = \begin{cases} 1 & \text{if } \theta \leq y \leq \theta + 1 \\ 0 & \text{elsewhere} \end{cases}$$

a. Show that $\bar{y}$ is a biased estimator of θ, and compute the bias.

b. Find $V(\bar{y})$.

c. What function of $\bar{y}$ is an unbiased estimator of θ?

8.88 Suppose y is a random sample of size $n = 1$ from a normal distribution with mean 0 and unknown variance σ^2.

a. Show that y^2/σ^2 has a chi-square distribution with 1 degree of freedom. [*Hint:* The result follows directly from Theorem 7.4.]

b. Derive a 95% confidence interval for σ^2 using y^2/σ^2 as a pivotal statistic.

8.89 Suppose y is a random sample of size $n = 1$ from a gamma distribution with parameters $\alpha = 1$ and arbitrary β.

a. Show that $2y/\beta$ has a gamma distribution with parameters $\alpha = 1$ and $\beta = 2$. [*Hint:* Use the distribution function approach of Section 7.2.]

b. Use the result of part **a** to show that $2y/\beta$ has a chi-square distribution with 2 degrees of freedom. [*Hint:* The result follows directly from Section 5.7.]

c. Derive a 95% confidence interval for β using $2y/\beta$ as a pivotal statistic.

8.90 Suppose y is a single observation from a normal distribution with mean μ and variance 1. Use y to find a 95% confidence interval for μ. [*Hint:* Start with the pivotal statistic $z = (y - \mu)$. Since z is a standard normal random variable,

$$P(-z_{.025} \leq y - \mu \leq z_{.025}) = .95$$

Follow the method of Example 8.6.]

8.91 A confidence interval for θ is said to be *unbiased* if the expected value of the interval midpoint is equal to θ.

a. Show that the small-sample confidence interval for μ,

$$\bar{y} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) \leq \mu \leq \bar{y} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

is unbiased.

b. Show that the confidence interval for σ^2,

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$$

is biased.

8.92 Suppose y is a single observation from a uniform distribution defined on the interval from 0 to θ. Find a 95% confidence limit LCL for θ such that $P(\text{LCL} < \theta < \infty) = .95$. [*Hint:* Start with the pivotal statistic y/θ and show (using the method of Chapter 7) that y/θ is uniformly distributed on the interval from 0 to 1. Then observe that

$$P\left(0 < \frac{y}{\theta} < .95\right) = \int_0^{.95} (1)dy = .95$$

and proceed to obtain LCL.]

COMPUTER LAB: Confidence Intervals for Means

Most commercial statistical software packages (e.g., SAS) do not have modules for computing confidence intervals for the parameters discussed in this chapter. Those that do (e.g., MINITAB) are limited in scope. For example, MINITAB will produce confidence intervals for means but not for variances or proportions. The MINITAB programs presented here give the confidence interval commands for estimating the parameters μ, $\mu_1 - \mu_2$, and μ_d. The outputs of the programs are shown in Figures 8.16–8.18, on pages 418–419, respectively.

MINITAB

a. Confidence Interval for μ—Data from Example 8.9

Command line		
1	`SET SILICON PPM IN C1`	Data entry instruction
2	`229 255 280 203 229`	Input data (5 observations per line)
3	`NAME C1 = 'PPM'`	
4	`TINTERVAL 99 C1`	99% confidence interval

COMMAND 4 The TINTERVAL command produces a confidence interval for the mean of the data stored in C1. The confidence interval (in this case, 99%) is specified following TINTERVAL. (The default is a 95% confidence interval.)

NOTE When σ is unknown, as is usually the case, TINTERVAL uses the appropriate value from the t distribution to calculate the interval regardless of the size of the sample. For large samples, recall that $t_{\alpha/2} \approx z_{\alpha/2}$.

FIGURE 8.16 ▶
Output for MINITAB program a.

```
                N      MEAN     STDEV  SE MEAN    99.0 PERCENT C.I.
ppm             4     241.8      33.2     16.6  (    144.8,    338.7)
```

b. Confidence Interval for $\mu_1 - \mu_2$, Independent Samples—Data from Example 8.11

Command line

```
1   READ THREE IN C1, SEVEN IN C2    Data entry command
2   1189 853  )
3    840 900  |  Input data values
4   1020 733  |  (1 observation per line)
5    980 785  )
6   NAME C1 = '3%ASPH' C2 = '7%ASPH'
7   TWOSAMPLE 95 C1 C2;  }  95% confidence interval
8   POOLED.              }
```

COMMAND 7 TWOSAMPLE produces a confidence interval on the difference between the mean of the data in C1 and the mean of the data in C2. By default, a 95% confidence interval is computed. To change the confidence level, specify 99, 90, etc., following the TWOSAMPLE command.

COMMAND 8 The POOLED subcommand instructs MINITAB to use s_p^2 in the calculation of a small-sample confidence interval. Omit the POOLED subcommand if you want MINITAB to compute a large sample confidence interval for $\mu_1 - \mu_2$.

NOTE TWOSAMPLE uses the appropriate value from the t distribution to compute the confidence interval regardless of the sample size. For large samples, recall that $t_{\alpha/2} \approx z_{\alpha/2}$.

FIGURE 8.17 ▶
Output for MINITAB program b.

```
TWOSAMPLE T FOR 3%asph VS 7%asph
         N      MEAN      STDEV   SE MEAN
3%asph   4      1007        144        72
7%asph   4     817.8       73.6        37

95 PCT CI FOR MU 3%asph - MU 7%asph: (-8, 387)

TTEST MU 3%asph = MU 7%asph (VS NE): T= 2.35  P=0.057  DF=  6

POOLED STDEV =          114
```

c. Confidence Interval for $\mu_d = (\mu_1 - \mu_2)$, Paired Samples—Data from Example 8.12

Command line

```
1   READ PLANT DATA IN C1, STATION DATA IN C2    Data entry instruction
2   2.0 2.2  )
.       .    |
.       .    }  input data (1 observation per line)
.       .    |
13  2.1 2.0  )
14  SUBTRACT C2 FROM C1, PUT IN C3
15  NAME C3 = "DIFF"
16  TINTERVAL 95 C3    95% confidence interval
```

COMMAND 14 Use the SUBTRACT command to calculate the differences for the paired observations in C1 and C2.

COMMAND 16 Use the TINTERVAL command to compute a 95% confidence interval for the mean of the differences in C3.

FIGURE 8.18 ►
Output for MINITAB program c.

	N	MEAN	STDEV	SE MEAN	95.0 PERCENT C.I.
diff	12	-0.0083	0.1676	0.0484	(-0.1149, 0.0982)

References

Devore, J. *Probability and Statistics for Engineering and the Sciences*, 2nd ed. Monterey, California: Brooks/Cole, 1987.

Freedman, D., Pisani, R., and Purves, R. *Statistics*. New York: W. W. Norton and Co., 1978.

Hoel, P. G. *Introduction to Mathematical Statistics*, 4th ed. New York: Wiley, 1971.

Hogg, R. V., and Craig, A. T. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.

Mendenhall, W. *Introduction to Probability and Statistics*, 8th ed. Boston: Duxbury, 1990.

Mendenhall, W., Wackerly, D. D., and Scheaffer, R. L. *Mathematical Statistics with Applications*, 3rd ed. Boston: Duxbury, 1989.

Mood, A. M., Graybill, F. A., and Boes, D. *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill, 1974.

Mosteller, F., and Tukey, J. W. *Data Analysis and Regression*. Reading, Massachusetts: Addison-Wesley, 1977. Chapters 8 and 10.

Snedecor, G. W., and Cochran, W. G. *Statistical Methods*, 7th ed. Ames, Iowa: Iowa State University Press, 1980.

Tukey, J. W. "Bias and Confidence in Not-Quite Large Samples." *Annals of Mathematical Statistics*, Vol. 29, 1958, p. 614.

CHAPTER NINE

Tests of Hypotheses

Objective

To introduce the basic concepts of a statistical test of a hypothesis; to present statistical tests for several common population parameters and to illustrate their use in practical sampling situations

Contents

9.1 The Relationship Between Statistical Tests of Hypotheses and Confidence Intervals

As stated in Chapter 8, there are two general methods available for making inferences about population parameters. We can estimate their values using confidence intervals (the subject of Chapter 8) or we can make decisions about them. Making decisions about specific values of the population parameters—**testing hypotheses** about these values—is the topic of this chapter.

Confidence intervals and hypothesis tests are related and can be used to make decisions about parameters. For example, suppose an investigator for the Environmental Protection Agency (EPA) wants to determine whether the mean level μ of a certain type of pollutant released into the atmosphere by a chemical company meets the EPA guidelines. If 3 parts per million is the upper limit allowed by the EPA, the investigator would want to use sample data (daily pollution measurements) to decide whether the company is violating the law, i.e., to decide whether $\mu > 3$. If, say, a 99% confidence interval for μ contained only numbers greater than 3, then the EPA would be confident that the mean exceeds the established limit.

As a second example, consider a manufacturer that purchases terminal fuses in lots of 10,000, and suppose that the supplier of the fuses guarantees that no more than 1% of the fuses in any given lot are defective. Since the manufacturer cannot test each of the 10,000 fuses in a lot, he must decide whether to accept or reject a lot based on an examination of a sample of fuses selected from the lot. If the number y of defective fuses in a sample of, say, $n = 100$, is large, he will reject the lot and send it back to the supplier. Thus, he wants to decide whether the proportion p of defectives in the lot exceeds .01, based on information contained in a sample. If a confidence interval for p falls below .01, then the manufacturer will accept the lot and be confident that the proportion of defectives is less than 1%; otherwise, he will reject it.

The examples in the preceding paragraphs illustrate how a confidence interval can be used to make a decision about a parameter. Note that both applications are one-directional; the EPA wants to determine whether $\mu > 3$ and the manufacturer wants to know if $p > .01$. (In contrast, if the manufacturer is interested in determining whether $p > .01$ or $p < .01$, the inference would be two-directional.)

Recall, from Chapter 8, that to find the value of z (or t) used in a $(1 - \alpha)100\%$ confidence interval, the value of α is divided in half and $\alpha/2$ is placed in both the upper and lower tails of the z (or t) distribution. Consequently, confidence intervals are designed to be two-directional. Use of a two-directional technique in a situation where a one-directional method is desired will lead the researcher (e.g., the EPA or the manufacturer) to understate the level of confidence associated with the method. As we will explain in this chapter, hypothesis tests are appropriate for either one- or two-directional decisions about a population parameter.

.2 Elements of a Statistical Test

We now return to the EPA example to introduce the concepts involved in a test of a hypothesis. We will use a method analogous to proof by contradiction. The theory the EPA wants to support, called the **alternative** (or **research**) **hypothesis**, is that $\mu > 3$, where μ is the true mean level of pollution in parts per million. The alternative hypothesis is denoted by the symbol H_a. The theory contradictory to the alternative hypothesis, that μ is at most equal to 3, say, $\mu = 3$, is called the **null hypothesis** and is denoted by the symbol H_0. Thus, the EPA hopes to show support for the alternative hypothesis, $\mu > 3$, by obtaining sample evidence indicating that the null hypothesis, $\mu = 3$, is false. That is, the EPA wants to test

$$H_0: \quad \mu = 3$$
$$H_a: \quad \mu > 3$$

The decision whether to reject the null hypothesis is based on a statistic, called a **test statistic**, computed from sample data. For example, suppose the EPA plans to base its decision on a sample of $n = 30$ daily pollution readings. If the sample mean $\bar{y}$ of the 30 pollution measurements is much larger than 3, the EPA would tend to reject the null hypothesis and conclude that $\mu > 3$.However, if $\bar{y}$ is smaller than 3, say, $\bar{y} = 2.8$ parts per million, there is insufficient evidence to refute the null hypothesis. Thus, the sample mean $\bar{y}$ serves as a test statistic.

The values that the test statistic $\bar{y}$ can assume will be divided into two sets. Those larger than some specified value, say, $\bar{y} \geq 3.1$, will imply rejection of the null hypothesis and acceptance of the alternative hypothesis. This set of values of the test statistic is known as the **rejection region** for the test. A test of the null hypothesis, H_0: $\mu = 3$, against the alternative hypothesis, H_a: $\mu > 3$, employing the sample mean $\bar{y}$ as a test statistic and $\bar{y} \geq 3.1$ as a rejection region, represents one particular test that possesses specific properties. If we change the rejection region to $\bar{y} \geq 3.2$, we obtain a different test with different properties.

The preceding discussion indicates that a statistical test consists of the four elements summarized in the box.

Elements of a Statistical Test

1. **Null hypothesis**, H_0, about one or more population parameters
2. **Alternative hypothesis**, H_a, that we will accept if we decide to reject the null hypothesis
3. **Test statistic**, computed from sample data
4. **Rejection region**, indicating the values of the test statistic that will imply rejection of the null hypothesis

In Section 9.3, we will show how to evaluate the reliability of a statistical test, how to compare one test with another, and how to evaluate the reliability of a particular test decision. We will apply the results to several practical examples.

9.3 Evaluating the Properties of a Statistical Test

Since a statistical test can result in one of only two outcomes—rejecting or not rejecting the null hypothesis—the test conclusion is subject to only two types of error. To illustrate, consider the EPA example of Section 9.2. Recall that the investigator wants to test H_0: $\mu = 3$ against H_a: $\mu > 3$, where μ = mean level of pollutant released into the atmosphere by a chemical company. If the investigator concludes that H_a is true (i.e., if he rejects H_0), then the EPA will charge the company with violating its pollution standards. The two errors that the EPA can make are shown in Table 9.1.

TABLE 9.1 Conclusions and Consequences for the EPA's Test of Hypothesis

	True State of Nature	
EPA Decision	Company Not in Violation (H_0 true)	Company in Violation (H_a true)
Company in Violation (Reject H_0)	Type I error	Correct decision
Company Not in Violation (Accept H_0)	Correct decision	Type II error

The EPA might reject the null hypothesis if, in fact, it is true. That is, the EPA might charge the company with violating its standards, when, in fact, the company is innocent (Type I error). Or the EPA might decide to accept the null hypothesis if, in fact, it is false. That is, the EPA may conclude that the company is not in violation of the pollution standards when, in fact, the company is in violation (Type II error). The probabilities of making these two types of errors measure the risks of making incorrect decisions when we perform a test of hypothesis and, consequently, provide measures of the goodness of this inferential decision-making procedure.

Definition 9.1

Rejecting the null hypothesis if it is true is a **Type I error**. The probability of making a Type I error is denoted by the symbol α.

Definition 9.2

Accepting the null hypothesis if it is false is a **Type II error**. The probability of making a Type II error is denoted by the symbol β.

Which of the two errors, Type I or Type II, is more serious? From the EPA's perspective, the Type I error is the more serious error. If the EPA falsely accuses the company of violating the pollution limits, a costly lawsuit will likely occur. On the other hand, the residents who live near the chemical company would probably view the Type II error as more serious; if this error occurs, the EPA is failing to charge the company when it is, in fact, polluting the surrounding air. In either case, it is important to compute the probabilities, α and β, to assess the reliability of inferences derived from the hypothesis test. The next four examples illustrate how to compute these probabilities.

EXAMPLE 9.1

A manufacturer of minicomputers believes that it can sell a particular software package to more than 20% of the buyers of its computers. Ten prospective purchasers of the computer were randomly selected and questioned about their interest in the software package. Of these, four indicated that they planned to buy the package. Does this sample provide sufficient evidence to indicate that more than 20% of the computer purchasers will buy the software package?

Solution

Let p be the true proportion of all prospective computer buyers who will purchase the software package. Since we want to show that $p > .2$, we choose H_a: $p > .2$ for the alternative hypothesis and H_0: $p = .2$ for the null hypothesis. We will use the binomial random variable y, the number of prospective purchasers in the sample who plan to buy the software, as the test statistic and will reject H_0: $p = .2$ if y is large. A graph of $p(y)$ for $n = 10$ and $p = .2$ is shown in Figure 9.1.

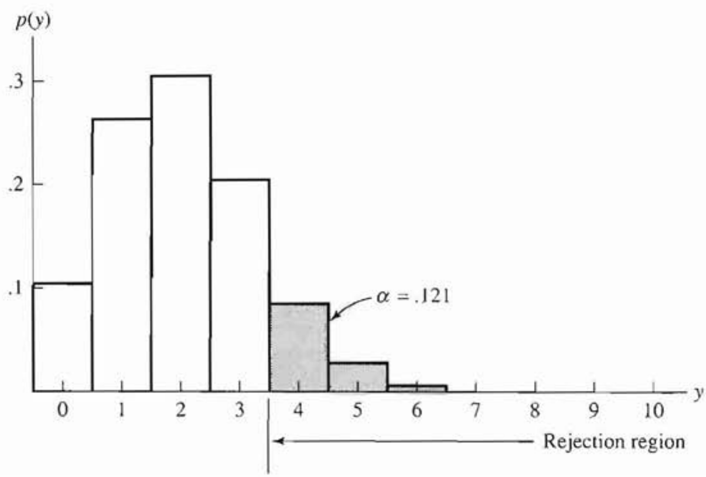


FIGURE 9.1 ▶ Graph of $p(y)$ for $n = 10$ and $p = .2$, i.e., if the null hypothesis is true

Large values of y will support the alternative hypothesis, H_a: $p > .2$, but what values of y should we include in the rejection region? Suppose that we select values of $y \geq 4$ as the rejection region. Then the elements of the test are:

H_0: $p = .2$

H_a: $p > .2$

Test statistic: y

Rejection region: $y \geq 4$

To conduct the test, we note that the observed value of y, $y = 4$, falls in the rejection region. Thus, for this test procedure, we reject the null hypothesis, H_0: $p = .2$, and conclude that the manufacturer is correct, i.e., $p > .2$.

EXAMPLE 9.2

What is the probability that the statistical test procedure of Example 9.1 would lead us to an incorrect decision if, in fact, the null hypothesis is true?

Solution

We will calculate the probability α that the test procedure would lead us to make a Type I error, i.e., to reject H_0 if, in fact, H_0 is true. This is the probability that y falls in the rejection region if in fact $p = .2$:

$$\alpha = P(y \geq 4 \text{ if in fact } p = .2) = 1 - \sum_{y=0}^{3} p(y)$$

The partial sum $\Sigma_{y=0}^{3} p(y)$ for a binomial random variable with $n = 10$ and $p = .2$ is given in Table 1 of Appendix II as .879. Therefore,

$$\alpha = 1 - \sum_{y=0}^{3} p(y) = 1 - .879 = .121$$

The probability that the test procedure would lead us to conclude that $p > .2$, if in fact it is not, is .121. This probability corresponds to the area of the shaded region in Figure 9.1.

In Example 9.1, we computed the probability α of committing a Type I error. *The probability β of making a Type II error, i.e., failing to detect a value of p greater than .2, depends on the value of p.* For example, if $p = .20001$, it will be very difficult to detect this small deviation from the null hypothesized value of $p = .2$. In contrast, if $p = 1.0$, then *every* prospective purchaser of the minicomputer will want to buy the software package, and in such a case it will be very evident from the sample information that $p > .2$. We will illustrate the procedure for calculating β in Example 9.3.

EXAMPLE 9.3

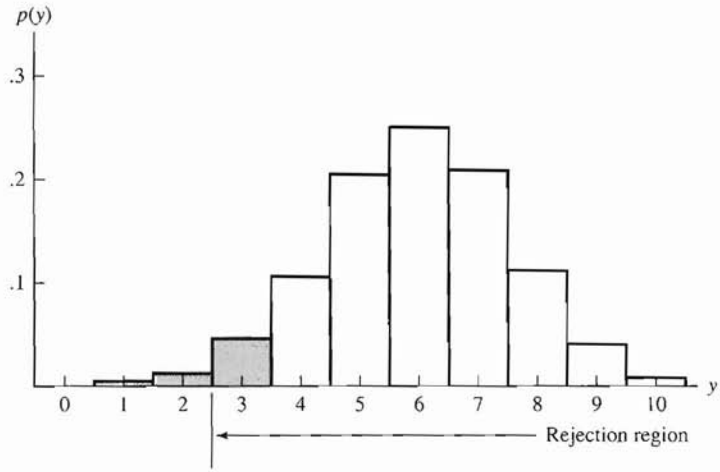
Refer to Example 9.2 and suppose that p is actually equal to .60. What is the probability β that the test procedure will fail to reject H_0: $p = .2$ if, in fact, $p = .6$?

Solution

The binomial probability distribution $p(y)$ for $n = 10$ and $p = .6$ is shown in Figure 9.2. The probability that we will fail to reject H_0 is equal to the probability that $y = 0, 1, 2$, or 3, i.e., the probability that y does not fall in the rejection region. This probability, β, corresponds to the shaded area under the probability histogram in the figure. Therefore,

$$\beta = P(y \leq 3 \text{ if in fact } p = .6) = \sum_{y=0}^{3} p(y) \quad \text{for } n = 10 \text{ and } p = .6$$

FIGURE 9.2 ▶ Graph of $p(y)$ for $n = 10$ and $p = .6$, i.e., if the alternative hypothesis is true



This partial sum, given in Table 1 of Appendix II for a binomial random variable with $n = 10$ and $p = .6$, is .055. Therefore, the probability that we will fail to reject H_0: $p = .2$ if p is as large as .6 is $\beta = .055$.

Another important property of a statistical test is its ability to detect departures from the null hypothesis when they exist. This is measured by the probability of rejecting H_0 when, in fact, H_0 is false. Note that this probability is simply $(1 - \beta)$:

$$\begin{aligned} P(\text{Reject } H_0 \text{ when } H_0 \text{ is false}) &= 1 - P(\text{Accept } H_0 \text{ when } H_0 \text{ is false}) \\ &= 1 - P(\text{Type II error}) \\ &= 1 - \beta \end{aligned}$$

The probability $(1 - \beta)$ is called the **power of the test**. The higher the power, the greater the probability of detecting departures from H_0 when they exist.

Definition 9.3

The **power** of a statistical test, $(1 - \beta)$, is the probability of rejecting the null hypothesis H_0 when, in fact, H_0 is false.

EXAMPLE 9.4

Refer to the test of hypothesis in Example 9.1. Find the power of the test if in fact $p = .3$.

Solution

From Definition 9.3, the power of the test is the probability $(1 - \beta)$. The probability of making a Type II error, i.e., failing to reject H_0: $p = .2$, if in fact $p = .3$, will be larger than the value of β calculated in Example 9.3 because $p = .3$ is much closer to the hypothesized value of $p = .2$. Thus,

$$\beta = P(y \leq 3 \text{ if in fact } p = .3) = \sum_{y=0}^{3} p(y) \quad \text{for } n = 10 \text{ and } p = .3$$

The value of this partial sum, given in Table 1 of Appendix II for a binomial random variable with $n = 10$ and $p = .3$, is .650. Therefore, the probability that we will fail to reject H_0: $p = .2$ if in fact $p = .3$ is $\beta = .650$ and the power of the test is $(1 - \beta) = (1 - .650) = .350$. You can see that the closer the actual value of p is to the hypothesized null value, the more unlikely it is that we will reject H_0: $p = .2$.

The preceding examples indicate how we can calculate α and β for a simple statistical test and thereby measure the risks of making Type I and Type II errors. These probabilities describe the properties of this inferential decision-making procedure and enable us to compare one test with another. For two tests, each with a rejection region selected so that α is equal to some specified value, say, .10, we would select the test that, for a specified alternative, has the smaller risk of making a Type II error, i.e., one that has the smaller value of β. This is equivalent to choosing the test with the higher power.

We will present a number of statistical tests in the following sections. In each case, the probability α of making a Type I error is known, i.e., α is selected by the experimenter and the rejection region is determined accordingly. In contrast, the value of β for a specific alternative is often difficult to calculate. This explains why we attempt to show that H_a is true by showing that the data do not support H_0. We hope that the sample evidence will support the alternative (or research) hypothesis. If it does, we will be concerned only about making a Type I error, i.e., rejecting H_0 if it is true. The probability α of committing such an error will be known.

EXERCISES

9.1 Define α and β for a statistical test of hypothesis.

9.2 Explain why each of the following statements is incorrect:
- **a.** The probability that the null hypothesis is correct is equal to α.
- **b.** If the null hypothesis is rejected, then the test proves that the alternative hypothesis is correct.
- **c.** In all statistical tests of hypothesis, $\alpha + \beta = 1$.

9.3 Pharmaceutical companies are continually searching for new drugs. Testing the thousands of compounds for the few that might be effective is known in the pharmaceutical industry as *drug screening*. Dunnett (1978) views the drug-screening procedure in its preliminary stage in terms of a statistical decision problem: "In drug screening, two actions are possible: (1) to 'reject' the drug, meaning to conclude that the tested drug has little or no effect, in which case it will be set aside and a new drug selected for screening; and (2) to 'accept' the drug provisionally, in which case it will be subjected to further, more refined experimentation."* Since it is the goal of the researcher to find a drug that effects a cure, the null and alternative hypotheses in a statistical test would take the following form:

H_0: Drug is ineffective in treating a particular disease

H_a: Drug is effective in treating a particular disease

Dunnett comments on the possible errors associated with the drug-screening procedure: "To abandon a drug when in fact it is a useful one (a *false negative*) is clearly undesirable, yet there is always some risk in that. On the other hand, to go ahead with further, more expensive testing of a drug that is in fact useless (a *false positive*) wastes time and money that could have been spent on testing other compounds."

a. A false negative corresponds to which type of error, Type I or Type II?
b. A false positive corresponds to which type of error, Type I or Type II?
c. Which of the two errors is more serious? Explain.

9.4 Pascal is a high-level programming language used frequently in minicomputers and microprocessors. An experiment was conducted to investigate the proportion of Pascal variables that are *array* variables (in contrast to *scalar* variables, which are less efficient in terms of execution time). Twenty variables are randomly selected from a set of Pascal programs and y, the number of array variables, is recorded. Suppose we want to test the hypothesis that Pascal is a more efficient language than Algol, in which 20% of the variables are array variables. That is, we will test H_0: $p = .20$ against H_a: $p > .20$, where p is the probability of observing an array variable on each trial. (Assume that the 20 trials are independent.)

a. Find α for the rejection region $y \geq 8$.
b. Find α for the rejection region $y \geq 5$.
c. Find β for the rejection region $y \geq 8$ if $p = .5$. [*Note:* Past experience has shown that approximately half the variables in most Pascal programs are array variables.]
d. Find β for the rejection region $y \geq 5$ if $p = .5$.
e. Which of the rejection regions, $y \geq 8$ or $y \geq 5$, is more desirable if you want to minimize the probability of a Type I error? Type II error?
f. Find the rejection region of the form $y \geq a$ so that α is approximately equal to .01.
g. For the rejection region determined in part f, find the power of the test, if in fact $p = .4$.
h. For the rejection region determined in part f, find the power of the test, if in fact $p = .7$.

9.5 A manufacturer of power meters, which are used to regulate energy thresholds of a data-communications system, claims that when its production process is operating correctly, only 10% of the power meters will be defective. A vendor has just received a shipment of 25 power meters from the manufacturer. Suppose the vendor wants to test H_0: $p = .10$ against H_a: $p > .10$, where p is the true proportion of power meters that are defective. Use $y \geq 6$ as the rejection region.

a. Determine the value of α for this test procedure.

*From Tanur, J. M., et al., eds. *Statistics: A Guide to the Unknown*. San Francisco: Holden-Day, 1978.

b. Find β if in fact $p = .2$. What is the power of the test for this value of p?
c. Find β if in fact $p = .4$. What is the power of the test for this value of p?

OPTIONAL EXERCISE

9.6 Show that for a fixed sample size n, α increases as β decreases, and vice versa.

9.4 Finding Statistical Tests: An Example of a Large-Sample Test

To find a statistical test about one or more population parameters, we must (1) find a suitable test statistic and (2) specify a rejection region. One method for finding a reasonable test statistic for testing a hypothesis was proposed by R. A. Fisher. For example, suppose we want to test a hypothesis about the sole parameter θ of a probability function $p(y)$ or density function $f(y)$, and let L represent the likelihood of the sample. Then to test the null hypothesis, H_0: $\theta = \theta_0$, Fisher's **likelihood ratio test statistic** is

$$\lambda = \frac{\text{Likelihood assuming } \theta = \theta_0}{\text{Likelihood assuming } \theta = \hat{\theta}} = \frac{L(\theta_0)}{L(\hat{\theta})}$$

where $\hat{\theta}$ is the maximum likelihood estimator of θ. Fisher reasoned that if θ differs from θ_0, then the value of the likelihood L when $\theta = \hat{\theta}$ will be larger than when $\theta = \theta_0$. Thus, the rejection region for the test contains values of λ that are small—say, smaller than some value λ_R.

If you are interested in learning more about Fisher's likelihood ratio test, consult the references at the end of this chapter. Fortunately, most of the statistics that we would choose intuitively for test statistics are functions of the corresponding likelihood ratio statistic λ. These are the pivotal statistics used to construct confidence intervals in Chapter 8.

Recall that most of the pivotal statistics in Chapter 8 have approximately normal sampling distributions for large samples. This fact allows us to easily derive a large-sample statistical test of hypothesis. To illustrate, suppose that we want to test a hypothesis, H_0: $\theta = \theta_0$, about a parameter θ and that the estimator $\hat{\theta}$ possesses a normal sampling distribution with mean θ and standard deviation $\sigma_{\hat{\theta}}$. We will further assume that $\sigma_{\hat{\theta}}$ is known or that we can obtain a good approximation for it when the sample size(s) is (are) large. It can be shown (proof omitted) that the likelihood ratio test statistic λ reduces to the standard normal variable z:

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

The location of the rejection region for this test can be deduced by examining the formula for the test statistic z. The farther $\hat{\theta}$ departs from θ_0, i.e., the larger the absolute value of the deviation $|\hat{\theta} - \theta_0|$, the greater will be the weight of evidence to indicate that θ is not equal to θ_0. If we want to detect values of θ larger than θ_0, i.e., H_a: $\theta > \theta_0$, we locate the rejection region in the upper tail of the sampling distribution of the standard normal z test statistic (see Figure 9.3a). If we want to detect only values of θ less than θ_0, i.e., H_a: $\theta < \theta_0$, we locate the rejection region in the lower tail of the z distribution (see Figure 9.3b). These two tests are called **one-tailed statistical tests** because the entire rejection region is located in only one tail of the z distribution. However, if we want to detect *either* a value of θ larger than θ_0 or a value smaller than θ_0, i.e., H_a: $\theta \neq \theta_0$, we locate the rejection region in both the upper and the lower tails of the z distribution (see Figure 9.3c). This is called a **two-tailed statistical test**.

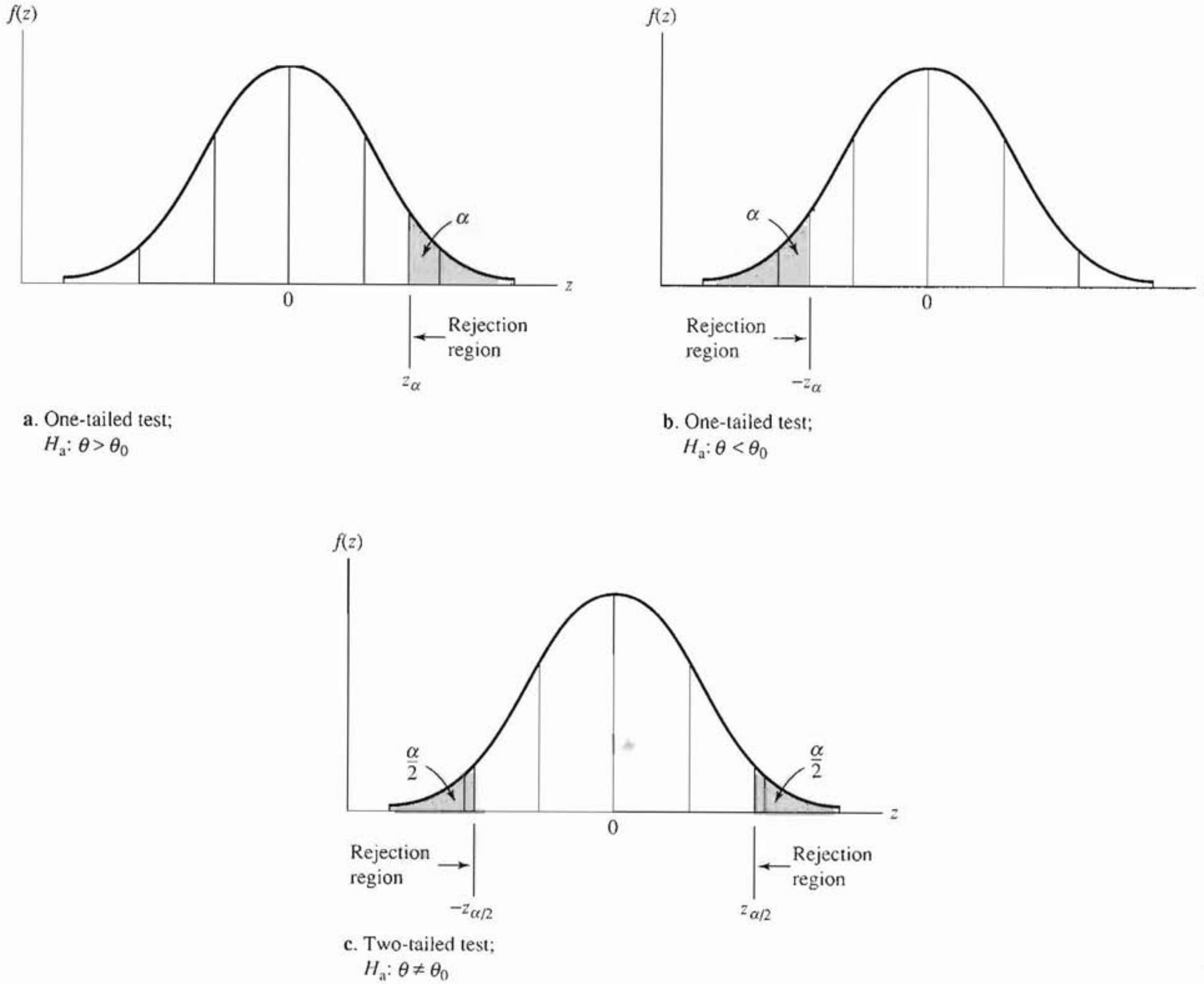


FIGURE 9.3 ▲
Rejection regions for one- and two-tailed tests

The large-sample statistical test that we have described is summarized in the box on page 432. Many of the population parameters and test statistics discussed in the remaining sections of Chapter 9 satisfy the assumptions of this test. We will illustrate the use of the test with a practical example on the population mean μ.

A Large-Sample Test Based on the Standard Normal z Test Statistic

One-Tailed Test

H_0: $\theta = \theta_0$

H_a: $\theta > \theta_0$
(or H_a: $\theta < \theta_0$)

Test statistic: $z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

Rejection region: $z > z_\alpha$
(or $z < -z_\alpha$)

where $P(z > z_\alpha) = \alpha$

Two-Tailed Test

H_0: $\theta = \theta_0$

H_a: $\theta \neq \theta_0$

Test statistic: $z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

Rejection region: $|z| > z_{\alpha/2}$

where $P(z > z_{\alpha/2}) = \alpha/2$

EXAMPLE 9.5

The Department of Highway Improvements, responsible for repairing a 25-mile stretch of interstate highway, wants to design a surface that will be structurally efficient. One important consideration is the volume of heavy freight traffic on the interstate. State weigh stations report that the average number of heavy-duty trailers traveling on a 25-mile segment of the interstate is 72 per hour. However, the section of highway to be repaired is located in an urban area and the department engineers believe that the volume of heavy freight traffic for this particular sector is greater than the average reported for the entire interstate. To validate this theory, the department monitors the highway for 50 1-hour periods randomly selected throughout the month. Suppose the sample mean and standard deviation of the heavy freight traffic for the 50 sampled hours are

$$\bar{y} = 74.1 \qquad s = 13.3$$

Do the data support the department's theory? Use $\alpha = .10$.

Solution

For this example, the parameter of interest is μ, the average number of heavy-duty trailers traveling on the 25-mile stretch of interstate highway. Recall that the sample mean $\bar{y}$ is used to estimate μ, and that for large n, $\bar{y}$ has an approximately normal sampling distribution. Thus, we can apply the large-sample test outlined in the box.

The elements of the test are

H_0: $\mu = 72$

H_a: $\mu > 72$

Test statistic: $z = \dfrac{\bar{y} - 72}{\sigma_{\bar{y}}} = \dfrac{\bar{y} - 72}{\sigma/\sqrt{n}} \approx \dfrac{\bar{y} - 72}{s/\sqrt{n}}$

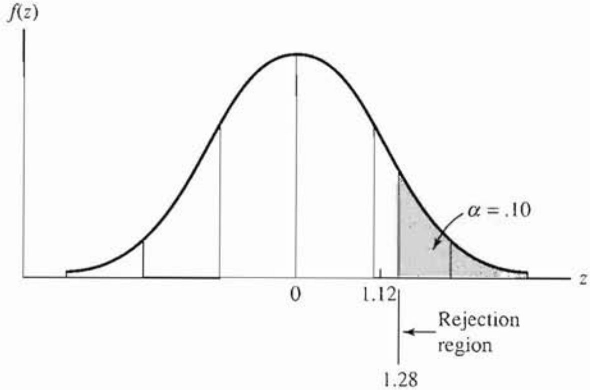
Rejection region: $z > 1.28$
(since $z_{.10} = 1.28$, from Table 4 of Appendix II)

We now substitute the sample statistics into the test statistic to obtain

$$z \approx \frac{74.1 - 72}{13.3/\sqrt{50}} = 1.12$$

Thus, although the average number of heavy freight trucks per hour in the sample exceeds the state's average by more than 2, the z value of 1.12 does not fall in the rejection region (see Figure 9.4). Therefore, this sample does not provide sufficient evidence at $\alpha = .10$ to support the Department of Highway Improvements' theory.

FIGURE 9.4 ▶
Location of the test statistic for Example 9.5



What is the risk of making an incorrect decision in Example 9.5? If we reject the null hypothesis then we know that the probability of making a Type I error (rejecting H_0 if it is true) is $\alpha = .10$. However, we failed to reject the null hypotheses in Example 9.5 and, consequently, we must be concerned about the possibility of making a Type II error (accepting H_0 if in fact it is false). We will evaluate the risk of making a Type II error in Example 9.6.

EXAMPLE 9.6

Refer to the one-tailed test for μ, Example 9.5. If the mean number μ of heavy freight trucks traveling a particular 25-mile stretch of interstate highway is in fact 78 per hour, what is the probability that the test procedure of Example 9.5 would fail to detect it? That is, what is the probability β that we would fail to reject H_0: $\mu = 72$ in this one-tailed test if μ is actually equal to 78?

Solution

To calculate β for the large-sample z test, we need to specify the rejection region in terms of the point estimator $\hat{\theta}$, where, for this example, $\hat{\theta} = \bar{y}$. From Figure 9.4, you can see that the rejection region consists of values of $z \geq 1.28$. To determine the value of $\bar{y}$ corresponding to $z = 1.28$, we substitute into the equation

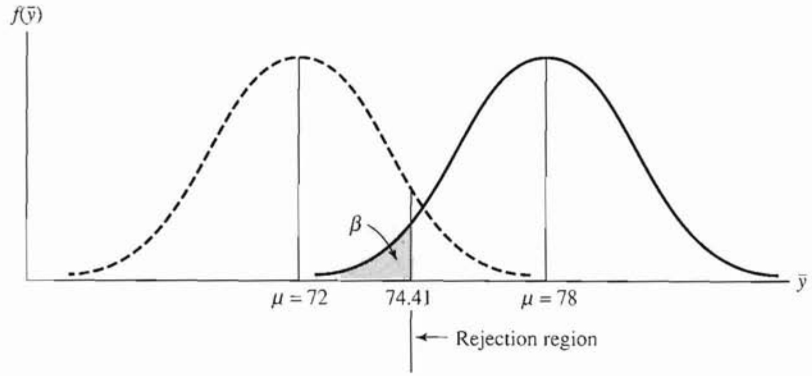
$$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}} \quad \text{or} \quad 1.28 = \frac{\bar{y} - 72}{13.3/\sqrt{50}}$$

Solving for $\bar{y}$, we obtain $\bar{y} = 74.41$. Therefore, the rejection region for the test is $z \geq 1.28$ or, equivalently, $\bar{y} \geq 74.41$.

The dotted curve in Figure 9.5 is the sampling distribution for $\bar{y}$ if H_0: $\mu = 72$ is true. This curve was used to locate the rejection region for $\bar{y}$ (and, equivalently, z), i.e., values of $\bar{y}$ contradictory to H_0: $\mu = 72$. The solid curve is the sampling distribution for $\bar{y}$ if $\mu = 78$. Since we want to find β if H_0 is in fact false and $\mu = 78$, we want to find the probability that $\bar{y}$ does not fall in the rejection region if $\mu = 78$. This probability corresponds to the shaded area under the solid curve for values of $\bar{y} <$ 74.41. To find this area under the normal curve, we need to find the area A corresponding to

$$z = \frac{\bar{y} - 78}{\sigma/\sqrt{n}} \approx \frac{74.41 - 78}{13.3/\sqrt{50}} = -1.91$$

FIGURE 9.5 ▶ The probability β of making a Type II error if $\mu = 78$ in Example 9.6



The value of A, given in Table 4 of Appendix II, is .4719. Then from Figure 9.5, it can be seen that

$$\beta = .5 - A = .5 - .4719 = .0281$$

Therefore, the probability of failing to reject H_0: $\mu = 72$ if μ is, in fact, as large as $\mu = 78$, is only .0281.

Example 9.6 illustrates that it is not too difficult to calculate β for various alternatives for the large-sample z test (see box). However, it may be extremely difficult to calculate β for other tests. Although sophisticated techniques are available for evaluating the risk of making a Type II error when the exact value of β is unavailable or is difficult to calculate, they are beyond the scope of this text. Consult the references at the end of this chapter if you are interested in learning about these methods.

Calculating β for a Large-Sample z Test

Consider a large-sample test of H_0: $\theta = \theta_0$ at significance level α. The value of β for a specific value of the alternative $\theta = \theta_a$ is calculated as follows:

Upper-tailed test: $\beta = P\left(z < \dfrac{\hat{\theta}_0 - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_0 = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$ is the value of the estimator corresponding to the border of the rejection region

Lower-tailed test: $\beta = P\left(z > \dfrac{\hat{\theta}_0 - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_0 = \theta_0 - z_\alpha \sigma_{\hat{\theta}}$ is the value of the estimator corresponding to the border of the rejection region

Two-tailed test: $\beta = P\left(\dfrac{\hat{\theta}_{0,\text{L}} - \theta_a}{\sigma_{\hat{\theta}}} < z < \dfrac{\hat{\theta}_{0,\text{U}} - \theta_a}{\sigma_{\hat{\theta}}}\right)$

where $\hat{\theta}_{0,\text{U}} = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$ and $\hat{\theta}_{0,\text{L}} = \theta_0 - z_\alpha \sigma_{\hat{\theta}}$ are the values of the estimator corresponding to the borders of the rejection region

EXERCISES

OPTIONAL EXERCISES

9.7 Suppose $y_1, y_2, \ldots, y_n$ is a random sample from a normal distribution with unknown mean μ and variance $\sigma^2 = 1$, i.e.,

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2}$$

Show that the likelihood L of the sample is

$$L(\mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\Sigma_{i=1}^n (y_i-\mu)^2/2}$$

9.8 Refer to Optional Exercise 9.7. Suppose we want to test H_0: $\mu = 0$ against the alternative H_a: $\mu > 0$. Since the estimator of μ is $\hat{\mu} = \bar{y}$, the likelihood ratio test statistic is

$$\lambda = \frac{L(\mu_0)}{L(\hat{\mu})} = \frac{L(0)}{L(\bar{y})}$$

Show that

$$\lambda = e^{-n(\bar{y})^2/2}$$

[*Hint:* Use the fact that $\Sigma_{i=1}^n (y_i - \bar{y})^2 = \Sigma_{i=1}^n y_i^2 - n\bar{y}^2$.]

9.9 Refer to Optional Exercises 9.7 and 9.8. Show that the rejection region $\lambda \le \lambda_\alpha$ is equivalent to the rejection region $\bar{y} \ge \bar{y}_\alpha$, where $P(\lambda \le \lambda_\alpha) = \alpha$ and $P(\bar{y} \ge \bar{y}_\alpha) = \alpha$. [*Hint:* Use the fact that $e^{-a^2} \to 0$ as $a \to \infty$.]

9.5 Choosing the Null and Alternative Hypotheses

Now that you have conducted a large-sample statistical test of hypothesis and have seen how to calculate the value of β—the probability of failing to reject H_0: $\theta = \theta_0$ if θ is in fact equal to some alternative value, $\theta = \theta_a$—the logic for choosing the null and alternative hypotheses may make more sense to you. The theory that we want to support (or detect if true) is usually chosen as the alternative hypothesis because, if the data support H_a (i.e., if we reject H_0), we immediately know the value of α, the probability of incorrectly rejecting H_0 if it is true. For example, in Example 9.5, the Department of Highway Improvements theorized that the mean number of heavy-duty vehicles traveling a certain segment of interstate exceeds 72 per hour. Consequently, the department set up the alternative hypothesis as H_a: $\mu > 72$. In contrast, if we choose the null hypothesis as the theory that we want to support, and if the data support this theory, i.e., the test leads to nonrejection of H_0, then we would have to investigate the values of β for some specific alternatives. Clearly, we want to avoid this tedious and sometimes extremely difficult task, if possible.

Another issue that arises in a practical situation is whether to conduct a one- or a two-tailed test. The decision depends on what you want to detect. For example, suppose you operate a chemical plant that produces a variable amount y of product per day and that if μ, the mean value of y, is less than 100 tons per day, you will eventually be bankrupt. If μ exceeds 100 tons per day, you are financially safe. To determine whether your process is leading to financial disaster, you will want to detect whether $\mu < 100$ tons, and you will conduct a one-tailed test of H_0: $\mu = 100$ versus H_a: $\mu < 100$. If you were to conduct a two-tailed test for this situation, you would reduce your chance of detecting values of μ less than 100 tons, i.e., you would increase the values of β for alternative values of $\mu < 100$ tons.

As a different example, suppose you have designed a new drug so that its mean potency is some specific level, say, 10%. As the mean potency tends to exceed 10%, you lose money. If it is less than 10% by some specified amount, the drug becomes ineffective as a pharmaceutical (and you lose money). To conduct a test of the mean potency μ for this situation, you would want to detect values of μ either larger than or smaller than $\mu = 10$. Consequently, you would select H_a: $\mu \ne 10$ and conduct a two-tailed statistical test (or alternatively, construct a confidence interval).

These examples demonstrate that a statistical test is an attempt to detect departures from H_0; the key to the test is to define the specific *alternatives* that you want to detect. We must stress, however, that H_0 *and* H_a *should be constructed prior to obtaining and observing the sample data*. If you use information in the sample data to aid in selecting H_0 and H_a, the prior information gained from the sample biases the test results—specifically, the true probability of a Type I error will be larger than the preselected value of α.

EXAMPLE 9.7

A metal lathe is checked periodically by quality control inspectors to determine whether it is producing machine bearings with a mean diameter of .5 inch. If the mean diameter of the bearings is larger or smaller than .5 inch, then the process is out of control and needs to be adjusted. Formulate the null and alternative hypotheses that could be used to test whether the bearing production process is out of control.

Solution

The hypotheses must be stated in terms of a population parameter. Thus, we define

μ = True mean diameter (in inches) of all bearings produced by the lathe

If either $\mu > .5$ or $\mu < .5$, then the metal lathe's production process is out of control. Since we wish to be able to detect either possibility, the null and alternative hypotheses would be

H_0: $\mu = .5$ (i.e., the process is in control)

H_a: $\mu \neq .5$ (i.e., the process is out of control)

In Sections 9.6–9.13, we will present applications of the hypothesis-testing logic developed in this chapter. The cases to be considered are those for which we developed estimation procedures in Chapter 8. Since the theory and reasoning involved are based on the developments of Chapter 8 and Sections 9.1–9.5, we will present only a summary of the hypothesis-testing procedure for one-tailed and two-tailed tests in each situation.

EXERCISES

In Exercises 9.10–9.15, formulate the appropriate null and alternative hypotheses.

9.10 A herpetologist wants to determine whether the egg-hatching rate for a certain species of frog exceeds .5 when the eggs are exposed to ultraviolet radiation.

9.11 A manufacturer of fishing line wants to show that the mean breaking strength of a competitor's 22-pound line is really less than 22 pounds.

9.12 A craps player who has experienced a long run of bad luck at the craps table wants to test whether the casino dice are "loaded," i.e., whether the proportion of "sevens" occurring in many tosses of the two dice is different from $\frac{1}{6}$ (if the dice are fair, the probability of tossing a "seven" is $\frac{1}{6}$).

9.13 Each year, *Computerworld* magazine reports the Datapro ratings of all computer software vendors. Vendors are rated on a scale from 1 to 4 (1 = poor, 4 = excellent) in such areas as reliability, efficiency, ease of installation, and ease of use by a random sample of software users. A software vendor wants to determine whether its product has a higher mean Datapro rating than a rival vendor's product.

9.14 The Environmental Protection Agency wishes to test whether the mean amount of radium-226 in soil in a Florida county exceeds the maxium allowable amount, 4 pCi/L.

9.15 Industrial engineers want to compare two methods of real-time scheduling in a manufacturing operation. Specifically, they want to determine whether the mean number of items produced differs for the two methods.

9.6 Testing a Population Mean

In Example 9.5, we developed a large-sample test for a population mean based on the standard normal z statistic. The elements of this test are summarized in the box.

Large-Sample ($n \geq 30$) Test of Hypothesis About a Population Mean μ

One-Tailed Test

H_0: $\mu = \mu_0$

H_a: $\mu > \mu_0$

(or H_a: $\mu < \mu_0$)

Test statistic:

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

Rejection region:

$z > z_\alpha$ (or $z < -z_\alpha$)

Two-Tailed Test

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

Test statistic:

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

Rejection region: $|z| > z_{\alpha/2}$

where z_α is the z value such that $P(z > z_\alpha) = \alpha$; and $z_{\alpha/2}$ is the z value such that $P(z > z_{\alpha/2}) = \alpha/2$. [*Note:* μ_0 is our symbol for the particular numerical value specified for μ in the null hypothesis.]

Assumptions: None (since the central limit theorem guarantees that $\bar{y}$ is approximately normal regardless of the distribution of the sampled population)

EXAMPLE 9.8

Humerus bones from the same species of animal tend to have approximately the same length-to-width ratios. When fossils of humerus bones are discovered, archeologists can often determine the species of animal by examining the length-to-width ratios of the bones. It is known that species A has a mean ratio of 8.5. Suppose 41 fossils of humerus bones were unearthed at an archeological site in East Africa, where species A is believed to have inhabited. (Assume that the unearthed bones are all from the same unknown species.) The length-to-width ratios of the bones were measured and are listed in Table 9.2.

TABLE 9.2 Length-to-Width Ratios of a Sample of Humerus Bones

10.73	8.89	9.07	9.20	10.33	9.98	9.84	9.59
8.48	8.71	9.57	9.29	9.94	8.07	8.37	6.85
8.52	8.87	6.23	9.41	6.66	9.35	8.86	9.93
8.91	11.77	10.48	10.39	9.39	9.17	9.89	8.17
8.93	8.80	10.02	8.38	11.67	8.30	9.17	12.00
9.38							

We wish to test the hypothesis that μ, the population mean ratio of all bones of this particular species, is equal to 8.5 against the alternative that it is different from 8.5, i.e., we wish to test whether the unearthed bones are from species A.

a. Suppose we want a very small chance of rejecting H_0, if, in fact, μ is equal to 8.5. That is, it is important that we avoid making a Type I error. Select an appropriate value of the significance level, α.

b. Test whether μ, the population mean length-to-width ratio, is different from 8.5, using the significance level selected in part a.

Solution

a. The hypothesis-testing procedure that we have developed gives us the advantage of being able to choose any significance level that we desire. Since the significance level, α, is also the probability of a Type I error, we will choose α to be very small. In general, researchers who consider a Type I error to have very serious practical consequences should perform the test at a very low α value—say, $\alpha = .01$. Other researchers may be willing to tolerate an α value as high as .10 if a Type I error is not deemed a serious error to make in practice. For this example, we will test at $\alpha = .01$.

b. We formulate the following hypotheses:

$$H_0: \quad \mu = 8.5$$
$$H_a: \quad \mu \neq 8.5$$

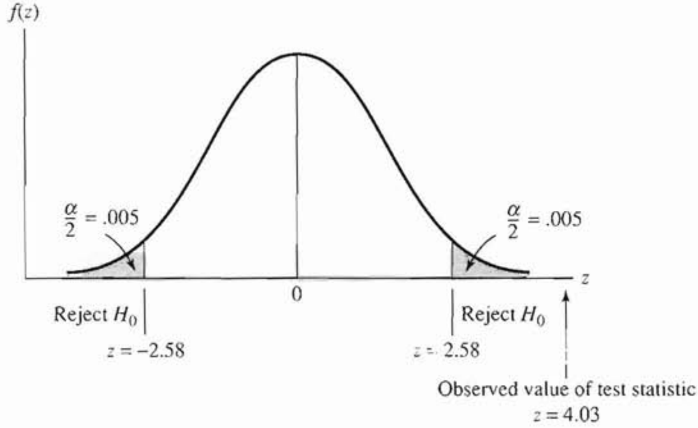
Note that this is a two-tailed test, since we want to detect departures from $\mu = 8.5$ in either direction. The sample size is large ($n = 41$); thus, we may proceed with the large-sample test about μ.

At significance level $\alpha = .01$, we will reject the null hypothesis for this two-tailed test if

$$|z| > z_{\alpha/2} = z_{.005}$$

i.e., if $z < -2.58$ or if $z > 2.58$. This rejection region is shown in Figure 9.6 on page 440.

FIGURE 9.6 ► Rejection region for Example 9.8



After entering the data of Table 9.2 into a computer, we obtained the summary statistics shown in the SAS printout, Figure 9.7. The values $\bar{y} = 9.257$ and $s = 1.203$ (shaded in the printout) are used to compute the test statistic

$$z \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{9.257 - 8.5}{1.203/\sqrt{41}} = 4.03$$

Since this value lies within the rejection region (see Figure 9.6), we reject H_0 and conclude that the mean length-to-width ratio of all humerus bones of this particular species is significantly different from 8.5. If the null hypothesis is in fact true (i.e., if $\mu = 8.5$), then the probability that we have incorrectly rejected it is equal to $\alpha = .01$.

FIGURE 9.7 ► SAS printout for Example 9.8

```
Analysis Variable : LWRATIO

N Obs        Minimum       Maximum           Mean        Std Dev
-----------------------------------------------------------------
   41      6.2300000    12.0000000      9.2575610      1.2035651
-----------------------------------------------------------------
```

The *practical* implications of the result obtained in Example 9.8 remain to be studied further. Perhaps the animal discovered at the archeological site is of some species other than A. Alternatively, the unearthed humerus bones may have larger than normal length-to-width ratios because of unusual feeding habits of species A. **It is not always the case that a statistically significant result implies a practically significant result.** The researcher must retain his or her objectivity and judge the practical significance using, among other criteria, his or her knowledge of the subject matter and the phenomenon under investigation.

A small-sample statistical test for making inferences about a population mean is (like its associated confidence interval of Section 8.5) based on the assumption that the sample data are independent observations on a normally distributed random variable. The test statistic is based on the t distribution given in Section 8.5.

The elements of the statistical test are listed in the accompanying box. As we suggested in Chapter 8, the small-sample test will possess the properties specified in the box even if the sampled population is moderately nonnormal. However, for data that departs greatly from normality (i.e., highly skewed data), we must resort to one of the nonparametric techniques discussed in Chapter 15.

Small-Sample Test of Hypothesis About a Population Mean μ

One-Tailed Test	*Two-Tailed Test*
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_a: $\mu > \mu_0$ (or H_a: $\mu < \mu_0$)	H_a: $\mu \neq \mu_0$

Test statistic: $$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

Rejection region: $t > t_\alpha$ (or $t < -t_\alpha$)	*Rejection region:* $\lvert t \rvert > t_{\alpha/2}$

where the distribution of t is based on $(n - 1)$ degrees of freedom; t_α is the t value such that $P(t > t_\alpha) = \alpha$; and $t_{\alpha/2}$ is the t value such that $P(t > t_{\alpha/2}) = \alpha/2$.

Assumption: The relative frequency distribution of the population from which the sample was selected is approximately normal.

Warning: If the data departs greatly from normality, this small-sample test may lead to erroneous inferences. In this case, use the nonparametric sign test that is discussed in Section 15.2.

EXAMPLE 9.9

Scientists have labeled benzene, a chemical solvent commonly used to synthesize plastics, as a possible cancer-causing agent. Studies have shown that people who work with benzene more than 5 years have 20 times the incidence of leukemia than the general population. As a result, the federal government has lowered the maximum allowable level of benzene in the workplace from 10 parts per million (ppm) to 1 ppm (reported in *Florida Times-Union*, Apr. 2, 1984). Suppose a steel manufacturing plant, which exposes its workers to benzene daily, is under investigation by the Occupational

Safety and Health Administration (OSHA). Twenty air samples, collected over a period of 1 month and examined for benzene content, yielded the following summary statistics:

$$\bar{y} = 2.1 \text{ ppm} \qquad s = 1.7 \text{ ppm}$$

Is the steel manufacturing plant in violation of the new government standards? Test the hypothesis that the mean level of benzene at the steel manufacturing plant is greater than 1 ppm, using $\alpha = .05$.

Solution

The OSHA wants to establish the research hypothesis that the mean level of benzene, μ, at the steel manufacturing plant exceeds 1 ppm. The elements of this small-sample one-tailed test are

H_0: $\mu = 1$

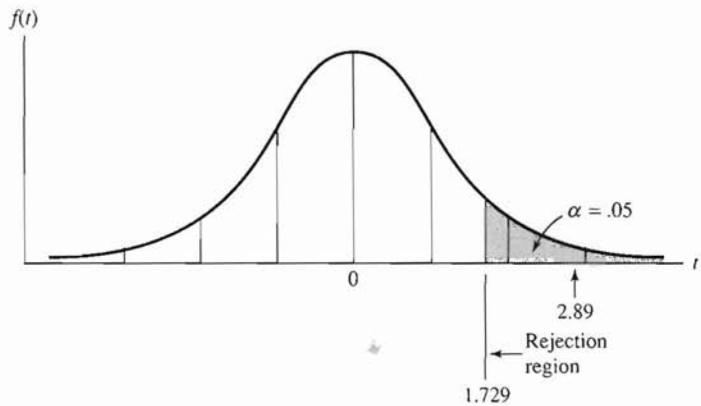
H_a: $\mu > 1$

Test statistic: $t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$

Assumption: The relative frequency distribution of the population of benzene levels for all air samples at the steel manufacturing plant is approximately normal.

Rejection region: For $\alpha = .05$ and df $= (n - 1) = 19$, reject H_0 if $t > t_{.05} = 1.729$ (see Figure 9.8)

FIGURE 9.8 ▶ Rejection region for Example 9.9



We now calculate the test statistic:

$$t = \frac{\bar{y} - 1}{s/\sqrt{n}} = \frac{2.1 - 1}{1.7/\sqrt{20}} = 2.89$$

Since the calculated t falls in the rejection region, the OSHA concludes that $\mu > 1$ part per million and the plant is in violation of the new government standards. The reliability associated with this inference is $\alpha = .05$. This implies that if the testing procedure was applied repeatedly to random samples of data collected at the plant, the OSHA would falsely reject H_0 for only 5% of the tests. Consequently, the OSHA is highly confident (95% confident) that the plant is violating the new standards.

EXERCISES

9.16 Radium-226 is a naturally occurring radioactive gas. Elevated levels of radium-226 in metropolitan Dade County (Florida) were recently investigated (*Florida Scientist*, Summer/Autumn 1991). The data in the table are radium-226 levels (measured in pCi/L) for 26 soil specimens collected in southern Dade County. The Environmental Protection Agency (EPA) has set maximum exposure levels of radium-226 at 4.0 pCi/L. Use the information in the accompanying MINITAB printout to determine whether the mean radium-226 level of soil specimens collected in southern Dade County is less than the EPA limit of 4.0 pCi/L. Use $\alpha = .10$.

1.46	.58	4.31	1.02	.17	2.92	.91	.43	.91
1.30	8.24	3.51	6.87	1.43	1.44	4.49	4.21	1.84
5.92	1.86	1.41	1.70	2.02	1.65	1.40	.75	

Source: Moore, H. E., and Gussow, D. G. "Radium and radon in Dade County ground water and soil samples." *Florida Scientist*, Vol. 54, No. 3/4, Summer/Autumn, 1991, p. 155 (portion of Table 3).

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
RadLevel	26	2.413	1.555	2.264	2.081	0.408
	MIN	**MAX**	**Q1**	**Q3**		
RadLevel	0.170	8.240	0.993	3.685		

9.17 The effect of machine breakdowns on the performance of a manufacturing system was investigated using computer simulation (*Industrial Engineering*, Aug. 1990). The simulation study focused on a single machine tool system with several characteristics, including a mean interarrival time of 1.25 minutes, a constant processing time of 1 minute, and a machine that breaks down 10% of the time. After $n = 5$ independent simulation runs of length 160 hours, the mean throughput per 40-hour week was $\bar{y} = 1{,}908.8$ parts. For a system with no breakdowns, the mean throughput for a 40-hour week will be equal to 1,920 parts. Assuming the standard deviation of the 5 sample runs was $s = 18$ parts per 40-hour week, test the hypothesis that the true mean throughput per 40-hour week for the system is less than 1,920 parts. Test using $\alpha = .05$.

9.18 Refer to the *Science* (Nov. 1988) study of inbreeding in tropical swarm-founding wasps, Exercise 8.30. A sample of 197 wasps, captured, frozen, and subjected to a series of genetic tests, yielded a sample mean inbreeding coefficient of $\bar{y} = .044$ with a standard deviation of $s = .884$. Recall that if the wasp has no tendency to inbreed, the true mean inbreeding coefficient μ for the species will equal 0.

a. Test the hypothesis that the true mean inbreeding coefficient μ for this species of wasp exceeds 0. Use $\alpha = .05$.

b. Compare the inference, part **a**, to the inference obtained in Exercise 8.28 using a confidence interval. Do the inferences agree? Explain.

9.19 Results of the second National Health and Nutrition Examination Survey indicate that the mean blood lead concentration of individuals between the ages of 6 months and 74 years is 14 μg/dl (*Analytical Chemistry*, Feb. 1986). However, the blood lead concentration in black children under the age of 5 years was found to be significantly higher than this figure. Suppose that in a random sample of 200 black children below the age of 5 years, the mean blood lead concentration is 21 μg/dl and the standard deviation is 10 μg/dl. Is there sufficient evidence to indicate that the true mean blood lead concentration in young black children is greater than 14 μg/dl? Test using $\alpha = .01$.

9.20 The EPA sets a limit of 5 parts per million on PCB (a dangerous substance) in water. A major manufacturing firm producing PCB for electrical insulation discharges small amounts from the plant. The company management, attempting to control the amount of PCB in its discharge, has given instructions to halt production if the mean amount of PCB in the effluent exceeds 3 parts per million. A random sampling of 50 water specimens produced the following statistics:

$\bar{y} = 3.1$ parts per million $\qquad s = .5$ part per million

a. Do these statistics provide sufficient evidence to halt the production process? Use $\alpha = .01$.
b. If you were the plant manager, would you want to use a large or a small value for α for the test in part **a**? Explain.

9.21 "Deep hole" drilling is a family of drilling processes used when the ratio of hole depth to hole diameter exceeds 10. Successful deep hole drilling depends on the satisfactory discharge of the drill chip. An experiment was conducted to investigate the performance of deep hole drilling when chip congestion exists (*Journal of Engineering for Industry*, May 1993). The length (in millimeters) of 50 drill chips resulted in the following summary statistics: $\bar{y} = 81.2$ mm, $s = 50.2$ mm. Conduct a test to determine whether the true mean drill chip length, μ, differs from 75 mm. Use a significance level of $\alpha = .01$.

9.22 *Environmental Science & Technology* (Oct. 1993) reported on a study of contaminated soil in The Netherlands. A total of 72 400-gram soil specimens were sampled, dried, and analyzed for the contaminant cyanide. The cyanide concentration (milligrams per kilogram of soil) of each soil specimen was determined using an infrared microscopic method. The sample resulted in a mean cyanide level of $\bar{y} = 84$ mg/kg and a standard deviation of $s = 80$ mg/kg. Use this information to test the hypothesis that the true mean cyanide level in soil in The Netherlands falls below 100 mg/kg. Test at $\alpha = .10$.

9.23 The building specifications in a certain city require that the sewer pipe used in residential areas have a mean breaking strength of more than 2,500 pounds per lineal foot. A manufacturer who would like to supply the city with sewer pipe has submitted a bid and provided the following additional information: An independent contractor randomly selected seven sections of the manufacturer's pipe and tested each for breaking strength. The results (pounds per lineal foot) follow:

2,610 2,750 2,420 2,510 2,540 2,490 2,680

Is there sufficient evidence to conclude that the manufacturer's sewer pipe meets the required specifications? Use a significance level of $\alpha = .10$.

9.24 Refer to Examples 9.5 and 9.6. Find the value of β for $\mu_a = 74$. What is the power of the test?

9.25 Refer to Example 9.9.
a. Find the value of β for $\mu_a = 1.015$.
b. Find the power of the test for $\mu_a = 1.045$.

9.26 Refer to Optional Exercises 9.7–9.9. Show that the rejection region for the likelihood ratio test is given by $z > z_\alpha$, where $P(z > z_\alpha) = \alpha$. [*Hint:* Under the assumption that H_0: $\mu = 0$ is true, show that $\sqrt{n}(\bar{y})$ is a standard normal random variable.]

9.7 The Observed Significance Level for a Test

According to the statistical test procedures described in the preceding sections, the rejection region and the corresponding value of α are selected prior to conducting the test and the conclusion is stated in terms of rejecting or not rejecting the null hypothesis. A second method of presenting the result of a statistical test is one that reports the extent to which the test statistic disagrees with the null hypothesis and leaves the reader the task of deciding whether to reject the null hypothesis. This measure of disagreement is called the **observed significance level** (or **p-value**) for the test.*

Definition 9.4

The **observed significance level**, or **p-value**, for a specific statistical test is the probability (assuming H_0 is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the one computed from the sample data.

When publishing the results of a statistical test of hypothesis in journals, case studies, reports, etc., many researchers make use of p-values. Instead of selecting α a priori and then conducting a test as outlined in this chapter, the researcher may compute and report the value of the appropriate test statistic and its associated p-value. It is left to the reader of the report to judge the significance of the result, i.e., the reader must determine whether to reject the null hypothesis in favor of the alternative hypothesis, based on the reported p-value. Usually, *the null hypothesis will be rejected only if the observed significance level is* **less** *than the fixed significance level α chosen by the reader.* There are two inherent advantages of reporting test results in this manner: (1) Readers are permitted to select the maximum value of α that they would be willing to tolerate if they actually carried out a standard test of hypothesis in the manner outlined in this chapter, and (2) it is an easy way to present the results of test calculations performed by a computer. Most statistical software packages perform the calculations for a test, give the observed value of the test statistic, and leave it to the reader to formulate a conclusion. Others give the observed significance level for the test, a procedure that makes it easy for the user to decide whether to reject the null hypothesis.

*The term *p-value* or *probability value* was coined by users of statistical methods. The p in the expression *p-value* should not be confused with the binomial parameter p.

EXAMPLE 9.10

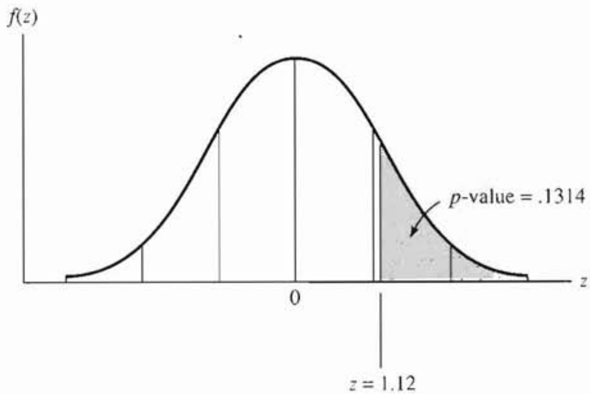
Find the observed significance level for the statistical test of Example 9.5 and interpret the result.

Solution

In Example 9.5, we tested a hypothesis about the mean μ of the number of heavy freight trucks per hour using a particular 25-mile stretch of interstate highway. Since we wanted to detect values of μ larger than $\mu_0 = 72$, we conducted a one-tailed test, rejecting H_0 for large values of $\bar{y}$, or equivalently, large values of z. The observed value of z, computed from the sample of $n = 50$ randomly selected 1-hour periods, was $z = 1.12$. Since any value of z larger than $z = 1.12$ would be even more contradictory to H_0, the observed significance level for the test is

$$p\text{-value} = P(z \geq 1.12)$$

FIGURE 9.9 ▶
Finding the p-value for an upper-tailed test when $z = 1.12$



This value corresponds to the shaded area in the upper tail of the z distribution shown in Figure 9.9. The area A corresponding to $z = 1.12$, given in Table 4 of Appendix II, is .3686. Therefore, the observed significance level is

$$p\text{-value} = P(z \geq 1.12) = .5 - \text{A} = .5 - .3686 = .1314$$

This result indicates that the probability of observing a z value at least as contradictory to H_0 as the one observed in this test (if H_0 is in fact true) is .1314. Therefore, we will reject H_0 only for preselected values of α greater than .1314. Recall that the Department of Highway Improvements selected a Type I error probability of $\alpha = .10$. Since $\alpha = .10$ is *less* than the p-value, the department has insufficient evidence to reject H_0. Note that this conclusion agrees with that of Example 9.5.

EXAMPLE 9.11

Suppose that the test of Example 9.5 had been a two-tailed test, i.e., suppose that the alternative of interest had been H_a: $\mu \neq 72$. Find the observed significance level for the test and interpret the result. Assume that $\alpha = .10$, as in Example 9.5.

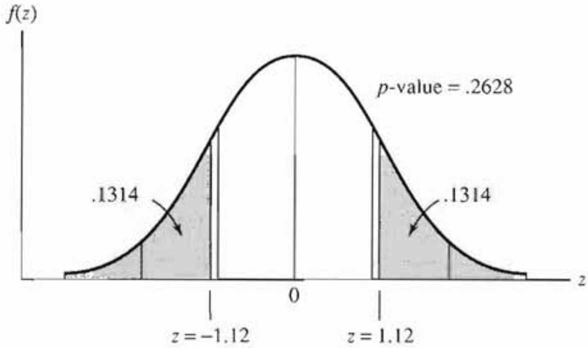
Solution

If the test were two-tailed, either very large or very small values of z would be contradictory to the null hypothesis H_0: $\mu = 72$. Consequently, values of $z \geq 1.12$ or $z \leq -1.12$ would be more contradictory to H_0 than the observed value of $z = 1.12$. Therefore, the observed significance level for the test (shaded in Figure 9.10) is

$$\begin{aligned} p\text{-value} &= P(z \geq 1.12) + P(z \leq -1.12) \\ &= 2(.1314) = .2628 \end{aligned}$$

Since we want to conduct the two-tailed test at $\alpha = .10$, and since the p-value exceeds α, we again have insufficient evidence to reject H_0.

FIGURE 9.10 ▶ Finding the p-value for a two-tailed test when $z = 1.12$



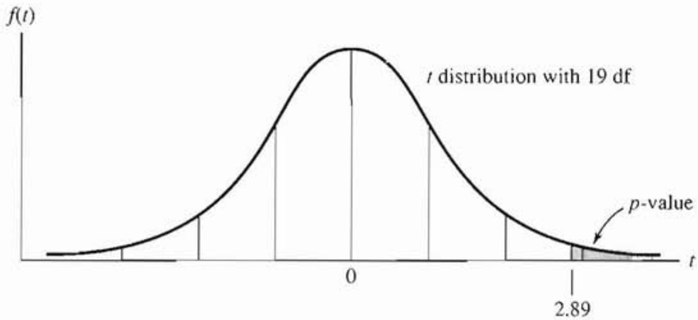
EXAMPLE 9.12

Find and interpret the observed significance level for the small-sample test described in Example 9.9. Recall that the test was conducted using $\alpha = .05$.

Solution

The test of Example 9.9 was a small-sample test of H_0: $\mu = 1$ versus H_a: $\mu > 1$. Since the value of t computed from the sample data was $t = 2.89$, the observed significance level (or p-value) for the test is equal to the probability that t would assume a value greater than or equal to 2.89, if in fact H_0 were true. This is equal to the area in the upper tail of the t distribution (shaded in Figure 9.11). To find this area, i.e., the p-value for the test, we consult the t table (Table 7 of Appendix II).

FIGURE 9.11 ▶ The observed significance level for the test of Example 9.12



Unlike the table of areas under the normal curve, Table 7 gives only the t values corresponding to the areas .100, .050, .025, .010, .005, .001, and .0005. Therefore, we can only approximate the p-value for the test. Since the observed t value was based on 19 degrees of freedom, we use the df = 19 row in Table 7 and move across the row until we reach the t values that are closest to the observed $t = 2.89$. The t values corresponding to p-values of .001 and .005 are 3.579 and 2.861, respectively. Since the observed t value falls between $t_{.001}$ and $t_{.005}$, the p-value for the test lies between .001 and .005. We could interpolate to more accurately locate the p-value for the test, but it is easier and adequate for our purposes to choose the larger area as the p-value and report it as .005. Thus, we would reject the null hypothesis, H_0: $\mu = 1$ part per million, for any value of α larger than .005. Since $\alpha = .05$ for this test, the correct conclusion is to reject H_0.

Calculating p-Values

Large-sample tests: $p\text{-value} = P(z \geq z_c)$ if upper-tailed

$p\text{-value} = P(z \leq z_c)$ if lower-tailed

$p\text{-value} = 2P(z \geq |z_c|)$ if two-tailed

where z_c is the computed value of the test statistic.

Small-sample tests: $p\text{-value} = P(t \geq t_c)$ if upper-tailed

$p\text{-value} = P(t \leq t_c)$ if lower-tailed

$p\text{-value} = 2P(t \geq |t_c|)$ if two-tailed

where t_c is the computed value of the test statistic.

[*Note:* $|z_c|$ and $|t_c|$ denote the **absolute values** of z_c and t_c and will always be positive.]

Interpreting p-Values

1. Choose the maximum value of α that you are willing to tolerate.
2. If the observed significance level (p-value) of the test is less than the maximum value α, then reject the null hypothesis.

You can see from Example 9.12 that calculating a p-value for a t test by hand will rarely lead to an exact value. If we desire an exact p-value, we need to resort to the use of a computer. The SAS printout for the t test of Examples 9.9 and 9.12 is shown in Figure 9.12. The p-value for a two-tailed test (shaded) is given under the heading PROB > $|T|$. The p-value for a one-tailed test is equal to the reported value divided by 2. Thus, the p-value for the one-tailed test H_0: $\mu = 1$ versus H_a: $\mu > 1$ is

$$p\text{-value} = \frac{.0088}{2} = .0044$$

FIGURE 9.12 ▶ SAS printout for t test of Example 9.12

```
Analysis Variable : BENZLEV

N Obs              T  Prob>|T|
------------------------------
   20      2.8937350    0.0088
------------------------------
```

EXERCISES

9.27 For a large-sample test of H_0: $\theta = \theta_0$ versus H_a: $\theta > \theta_0$, compute the p-value associated with each of the following test statistic values:
a. $z = 1.96$ b. $z = 1.645$ c. $z = 2.67$ d. $z = 1.25$

9.28 For a large-sample test of H_0: $\theta = \theta_0$ versus H_a: $\theta \neq \theta_0$, compute the p-value associated with each of the following test statistic values:
a. $z = -1.01$ b. $z = -2.37$ c. $z = 4.66$ d. $z = 1.45$

9.29 Compute and interpret the p-value for the test of Example 9.8, assuming the test is two-tailed.

9.30 Compute and interpret the p-values for the tests conducted in the following exercises.
a. Exercise 9.16 b. Exercise 9.17 c. Exercise 9.18 d. Exercise 9.19 e. Exercise 9.20

9.31 A SAS printout for the t test of Exercise 9.23 is shown here. Find and interpret the p-value of the test. Does the result agree with your inference in Exercise 9.23?

```
Analysis Variable : Y (strength - 2500)

N Obs              T  Prob>|T|
------------------------------
    7      1.6419203    0.1517
------------------------------
```

9.8 Testing the Difference Between Two Population Means: Independent Samples

Consider independent random samples from two populations with means μ_1 and μ_2, respectively. When the sample sizes are large (i.e., $n_1 \geq 30$ and $n_2 \geq 30$), a test of hypothesis for the difference between the population means $(\mu_1 - \mu_2)$ is based on the pivotal z statistic given in Section 8.6. A summary of the large-sample test is provided in the box.

Large-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Independent Samples

One-Tailed Test	*Two-Tailed Test*
H_0: $(\mu_1 - \mu_2) = D_0$	H_0: $(\mu_1 - \mu_2) = D_0$
H_a: $(\mu_1 - \mu_2) > D_0$ [or H_a: $(\mu_1 - \mu_2) < D_0$]	H_a: $(\mu_1 - \mu_2) \neq D_0$

Test statistic: $$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}} \approx \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Rejection region:	*Rejection region:*
$z > z_\alpha$ (or $z < -z_\alpha$)	$\lvert z\rvert > z_{\alpha/2}$

[*Note:* D_0 is our symbol for the particular numerical value specified for $(\mu_1 - \mu_2)$ in the null hypothesis. In many practical applications, we wish to hypothesize that there is no difference between the population means; in such cases, $D_0 = 0$.]

Assumptions:
1. The sample sizes n_1 and n_2 are sufficiently large—say, $n_1 \geq 30$ and $n_2 \geq 30$.
2. The two samples are selected randomly and independently from the target populations.

EXAMPLE 9.13

To reduce costs, a bakery has implemented a new leavening process for preparing commercial bread loaves. Loaves of bread were randomly sampled and analyzed for calorie content both before and after implementation of the new process. A summary of the results of the two samples is shown in the table. Do these samples provide sufficient evidence to conclude that the mean number of calories per loaf has decreased since the new leavening process was implemented? Test using $\alpha = .05$.

New Process	*Old Process*
$n_1 = 50$	$n_2 = 30$
$\bar{y}_1 = 1{,}255$ calories	$\bar{y}_2 = 1{,}330$ calories
$s_1 = 215$ calories	$s_2 = 238$ calories

Solution

We can best answer this question by performing a test of a hypothesis. Defining μ_1 as the mean calorie content per loaf manufactured by the new process and μ_2 as the mean calorie content per loaf manufactured by the old process, we will attempt to support the research (alternative) hypothesis that $\mu_2 > \mu_1$ [i.e., that $(\mu_1 - \mu_2) < 0$]. Thus, we will test the null hypothesis that $(\mu_1 - \mu_2) = 0$, rejecting this hypothesis if $(\bar{y}_1 - \bar{y}_2)$ equals a large negative value. The elements of the test are as follows:

H_0: $(\mu_1 - \mu_2) = 0$ (i.e., $D_0 = 0$)

H_a: $(\mu_1 - \mu_2) < 0$ (i.e., $\mu_1 < \mu_2$)

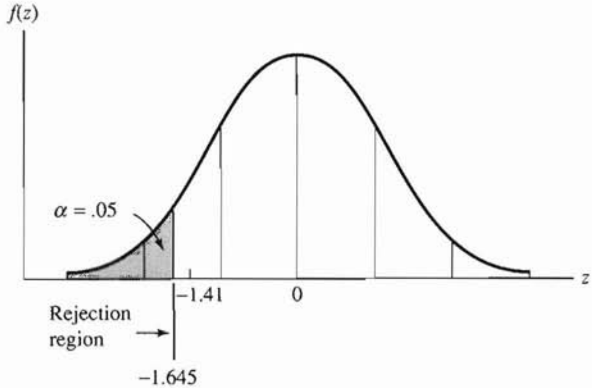
Test statistic: $z = \dfrac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}} = \dfrac{(\bar{y}_1 - \bar{y}_2) - 0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}}$

(since both n_1 and n_2 are greater than or equal to 30)

Rejection region: $z < -z_\alpha = -1.645$ (see Figure 9.13)

Assumptions: The two samples of bread loaves are independently selected.

FIGURE 9.13 ▶
Rejection region for Example 9.13



We now calculate

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sigma_{(\bar{y}_1 - \bar{y}_2)}} = \frac{(1{,}255 - 1{,}330)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$\approx \frac{-75}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{-75}{\sqrt{\dfrac{(215)^2}{50} + \dfrac{(238)^2}{30}}} = \frac{-75}{53.03} = -1.41$$

As you can see in Figure 9.13, the calculated z value does not fall in the rejection region. The samples do not provide sufficient evidence, with $\alpha = .05$, to conclude that the new process yields a loaf with fewer mean calories.

When the sample sizes n_1 and n_2 are inadequate to permit use of the large-sample procedure of Example 9.13, modifications may be made to perform a small-sample test of hypothesis about the difference between two population means. The test procedure is based on assumptions that are, again, more restrictive than in the large-sample case. The elements of the hypothesis test and the assumptions required are listed in the box. *Reminder:* When the assumption of normal population is grossly violated, the small-sample test outlined here will be invalid. In this case, we must resort to a nonparametric method.

Small-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Independent Samples

One-Tailed Test

H_0: $(\mu_1 - \mu_2) = D_0$

H_a: $(\mu_1 - \mu_2) > D_0$
[or H_a: $(\mu_1 - \mu_2) < D_0$]

Two-Tailed Test

H_0: $(\mu_1 - \mu_2) = D_0$

H_a: $(\mu_1 - \mu_2) \neq D_0$

Test statistic:
$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Rejection region: $t > t_\alpha$
[or $t < -t_\alpha$]

Rejection region: $|t| > t_{\alpha/2}$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

and the distribution of t is based on $n_1 + n_2 - 2$ df.

Assumptions:
1. The populations from which the samples are selected both have approximately normal relative frequency distributions.
2. The variances of the two populations are equal, i.e., $\sigma_1^2 = \sigma_2^2$.
3. The random samples are selected in an independent manner from the two populations.

Warning: When the assumption of normal populations is violated, the test may lead to erroneous inferences. In this case, use the nonparametric Wilcoxon test described in Section 15.3.

EXAMPLE 9.14

Computer response time is defined as the length of time a user has to wait for the computer to access information on the disk. Suppose a data center wants to compare the average response times of its two computer disk drives. If μ_1 is the mean response time of disk 1 and μ_2 is the mean response time of disk 2, we want to detect a difference between μ_1 and μ_2—if such a difference exists. Therefore, we want to test the null hypothesis

$$H_0: \ (\mu_1 - \mu_2) = 0$$

against the alternative hypothesis

$$H_a\colon \ (\mu_1 - \mu_2) \neq 0 \quad (\text{i.e., } \mu_1 > \mu_2 \text{ or } \mu_1 < \mu_2)$$

Independent random samples of 13 response times for disk 1 and 15 response times for disk 2 were selected. The data (recorded in milliseconds), as well as summary statistics, are given in Table 9.3. Is there sufficient evidence to indicate a difference between the mean response times of the two disk drives? Test using $\alpha = .05$.

TABLE 9.3 Response Times for Two Disk Drives

Disk 1 ($n_1 = 13$)				Disk 2 ($n_2 = 15$)			
59	73	74	61	71	63	40	34
92	60	84		38	48	60	75
54	73	47		47	41	44	86
102	75	33		53	68	39	
$\bar{y}_1 = 68.2$	$s_1 = 18.6$			$\bar{y}_2 = 53.8$	$s_2 = 15.8$		

Solution

We first calculate

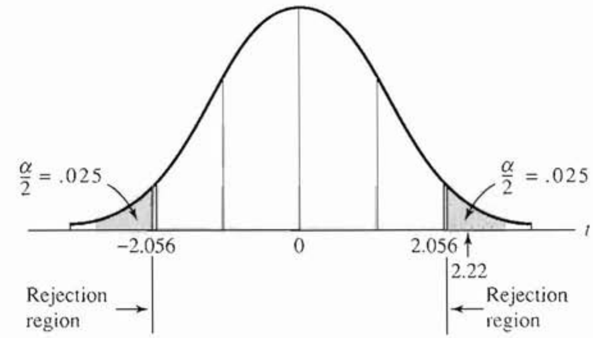
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(13 - 1)(18.6)^2 + (15 - 1)(15.8)^2}{13 + 15 - 2}$$
$$= \frac{7{,}646.48}{26} = 294.09$$

Then, if we can assume that the distributions of the response times for the two disk drives are both approximately normal with equal variances, the test statistic is

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(68.2 - 53.8) - 0}{\sqrt{294.09\left(\frac{1}{13} + \frac{1}{15}\right)}}$$
$$= \frac{14.4}{6.5} = 2.22$$

Since the observed value of t ($t = 2.22$) falls in the rejection region (see Figure 9.14 on page 454), the samples provide sufficient evidence to indicate that the mean response times differ for the two disk drives. Or, we say that the test results are statistically significant at the $\alpha = .05$ level of significance. Because the rejection was in the positive or upper tail of the t distribution, it appears that the mean response time for disk drive 1 exceeds that for disk drive 2.

FIGURE 9.14 ▶
Rejection region for Example 9.14



Refer to Example 9.14. The same conclusion can be reached using the p-value approach. The SAS printout for the t test of Example 9.14 is shown in Figure 9.15. The test statistic and p-value for the test are both shaded on the printout. Note that the two-tailed p-value (for the equal variances case), $p = .0356$, is less than $\alpha = .05$; thus, there is sufficient evidence to reject H_0.

FIGURE 9.15 ▶
SAS printout for Example 9.14

```
                          TTEST PROCEDURE

Variable: Y

DISK      N         Mean       Std Dev     Std Error      Minimum       Maximum
-------------------------------------------------------------------------------
  1      13  68.23076923   18.65991178    5.17532836  33.00000000   102.0000000
  2      15  53.80000000   15.80777386    4.08154966  34.00000000    86.0000000

Variances        T       DF     Prob>|T|
----------------------------------------
Unequal     2.1894     23.7       0.0387
Equal       2.2163     26.0       0.0356

For H0: Variances are equal, F' = 1.39    DF = (12,14)    Prob>F' = 0.5482
```

Recall from Section 8.6 that valid small-sample inferences about $(\mu_1 - \mu_2)$ can still be made when the assumption of equal variances is violated. We conclude this section by giving the modifications required to obtain approximate small-sample tests about $(\mu_1 - \mu_2)$ when $\sigma_1^2 \neq \sigma_2^2$ for the two cases described in Section 8.6: $n_1 = n_2$ and $n_1 \neq n_2$.

Modifications to Small-Sample Tests About $(\mu_1 - \mu_2)$ When $\sigma_1^2 \neq \sigma_2^2$: Independent Samples

$n_1 = n_2 = n$

Test statistic:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\dfrac{1}{n}(s_1^2 + s_2^2)}}$$

Degrees of freedom: $\nu = n_1 + n_2 - 2 = 2(n - 1)$

$n_1 \neq n_2$

Test statistic:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Degrees of freedom: $$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}\right]}$$

Note: The value of ν will generally not be an integer. Round down to the nearest integer to use the t table (Table 7 of Appendix II).

EXERCISES

9.32 Does competition between separate research and development (R&D) teams in the U.S. Department of Defense, working independently on the same project, improve performance? To answer this question, performance ratings were assigned to each of 58 multisource (competitive) and 63 sole source R&D contracts (*IEEE Transactions on Engineering Management*, Feb. 1990). With respect to quality of reports and products, the competitive contracts had a mean performance rating of 7.62, whereas the sole source contracts had a mean of 6.95.

a. Set up the null and alternative hypothesis for determining whether the mean quality performance rating of competitive R&D contracts exceeds the mean for sole source contracts.
b. Find the rejection region for the test using $\alpha = .05$.
c. The p-value for the test was reported to be between .02 and .03. What is the appropriate conclusion?

9.33 a. Use a random number table (Table 6 of Appendix II) to generate a random sample of $n = 40$ observations on DDT concentration in fish from the data of Appendix III. Compute $\bar{x}$ and s for the sample measurements.

b. The Food and Drug Administration (FDA) sets the limit for DDT content in individual fish at 5 parts per million (ppm). Does the sample of part a provide sufficient evidence to conclude that the average DDT content of individual fish inhabiting the Tennessee River and its creek tributaries exceeds 5 ppm? Test using a significance level of $\alpha = .01$.

c. Suppose the test of hypothesis, part b, was based on a random sample of only $n = 8$ fish. What are the disadvantages of conducting this small-sample test?

d. Repeat part **b** using only the information on the DDT contents of a sample of 8 fish (randomly selected from the 40 observations of part **a**). Compare the results of the large- and small-sample tests.

9.34 Many computer software packages utilize menu-driven user-interfaces to increase "user-friendliness." One feature that can be incorporated into the interface is a stacked menu display. Each time a menu item is selected, a submenu is displayed partially over the parent menu, thus creating a series of "stacked" menus. The *Special Interest Group on Computer Human Interaction Bulletin* (July 1993) reported on a study to determine the effects of the presence or absence of a stacked menu structure on search time. Twenty-two subjects were randomly placed into one of two groups, and each was asked to search a menu-driven software package for a particular item. In the experimental group ($n_1 = 11$), the stacked menu format was used; in the control group ($n_2 = 11$), only the current menu was displayed.

a. The researcher's initial hypothesis is that the mean time required to find a target item does not differ for the two menu displays. Describe the statistical method appropriate for testing this hypothesis.

b. What assumptions are required for inferences derived from the analysis to be valid?

c. The mean search times for the two groups were 11.02 seconds and 11.07 seconds, respectively. Is this enough information to conduct the test? Explain.

d. The observed significance level for the test, part **a**, exceeds .10. Interpret this result.

9.35 *Environmental Science & Technology* (Oct. 1993) reported on a study of insecticides used on dormant orchards in the San Joaquin Valley, California. Ambient air samples were collected and analyzed daily at an orchard site during the most intensive period of spraying. The thion and oxon levels (in ng/m^3) in the air samples are recorded in the table, as well as the oxon/thion ratios. Compare the mean oxon/thion ratios of foggy and clear/cloudy conditions at the orchard using a test of hypothesis. Use $\alpha = .05$.

Date	*Condition*	*Thion*	*Oxon*	*Oxon/Thion Ratio*
Jan. 15	Fog	38.2	10.3	.270
17	Fog	28.6	6.9	.241
18	Fog	30.2	6.2	.205
19	Fog	23.7	12.4	.523
20	Fog	62.3	(Air sample lost)	—
20	Clear	74.1	45.8	.618
21	Fog	88.2	9.9	.112
21	Clear	46.4	27.4	.591
22	Fog	135.9	44.8	.330
23	Fog	102.9	27.8	.270
23	Cloudy	28.9	6.5	.225
25	Fog	46.9	11.2	.239
25	Clear	44.3	16.6	.375

Source: Selber, J. N., et al. "Air and fog deposition residues of four organophosphate insecticides used on dormant orchards in the San Joaquin Valley, California." *Environmental Science & Technology*, Vol. 27, No. 10, Oct. 1993, p. 2240 (Table V).

9.36 Percentage of body fat can be a good indicator of an individual's energy metabolic status and general health. In an *American Journal of Physical Anthropology* (Jan. 1981) study of the percentage of body fat of college students in India, two groups of healthy male students, from urban and rural colleges in eastern India, were independently and randomly selected. The percentage of body fat in each was measured, with the results summarized in the table. Does the sample information provide sufficient evidence to conclude that the mean percentage of body fat in healthy male college students residing in urban areas of India differs from the corresponding mean for students residing in rural areas? Use a significance level of $\alpha = .05$.

Urban Students	*Rural Students*
$n_1 = 193$	$n_2 = 188$
$\bar{y}_1 = 12.07$	$\bar{y}_2 = 11.04$
$s_1 = 3.04$	$s_2 = 2.63$

Source: Bandyopadhyay, B., and Chattopadhyay, H. "Body fat in urban and rural male college students of eastern India." *American Journal of Physical Anthropology*, Jan. 1981, Vol. 54, pp. 119–122.

9.37 According to a popular model of managerial behavior, the current state of automation in a manufacturing firm influences managers' perceptions of problems of automation. To investigate this proposition, researchers at Concordia University (Montreal) surveyed managers at firms with a high level of automation and at firms with a low level of automation (*IEEE Transactions on Engineering Management*, Aug. 1990). Each manager was asked to give his/her perception of the problems of automation at the firm. Responses were measured on a 5-point scale (1: No problem, . . . 5: Major Problem). Summary statistics for the two groups of managers, provided in the table, were used to test the hypothesis of no difference in the mean perceptions of automation problems between managers of highly automated and less automated manufacturing firms.

	Sample Size	*Mean*	*Standard Deviation*
Low Level	17	3.274	.762
High Level	8	3.280	.721

Source: Farhoomand, A. F., Kira D., and Williams, J. "Managers' perceptions towards automation in manufacturing." *IEEE Transactions on Engineering Management*, Vol. 37, No. 3, Aug. 1990, p. 230.

a. Conduct the test for the researchers, assuming that the perception variances for the two groups of managers are equal. Use $\alpha = .01$.

b. Conduct the test for the researchers, if it is known that the perception variances differ for managers at low-level and high-level firms.

9.38 An industrial plant wants to determine which of two types of fuel—gas or electric—will produce more useful energy at the lower cost. One measure of economical energy production, called the *plant investment per delivered quad*, is calculated by taking the amount of money (in dollars) invested in the particular utility by the plant, and dividing by the delivered amount of energy (in quadrillion British thermal units). The smaller this ratio, the less an industrial plant pays for its delivered energy. Random samples of 11 plants using electrical utilities and 16 plants using gas utilities were taken, and the plant investment/quad wa calculated for each. The data are listed in the table, followed by a MINITAB printout of the analysis of th data.

Electric

204.15	.57	62.76	89.72
.35	85.46	.78	.65
44.38	9.28	78.60	

Gas

.78	16.66	74.94	.01
.54	23.59	88.79	.64
.82	91.84	7.20	66.64
.74	64.67	165.60	.36

```
TWOSAMPLE T FOR electric VS gas
           N      MEAN      STDEV    SE MEAN
electric  11      52.4       62.4         19
gas       16      37.7       49.0         12

95 PCT CI FOR MU electric - MU gas: (-30, 59)

TTEST MU electric = MU gas (VS NE): T= 0.68  P=0.50  DF=  25

POOLED STDEV =        54.8
```

a. Do these data provide sufficient evidence at the $\alpha = .05$ level of significance to indicate a difference in the average investment/quad between the plants using gas and those using electrical utilities?
b. What assumptions are required for the procedure to be valid?
c. Check whether the assumptions, part **b**, are reasonably satisfied. How does this impact on the validity of the result, part **a**?

9.39 A field experiment was conducted to ascertain the impact of desert granivores (seed-eaters) on the density and distribution of seeds in the soil (*Ecology*, Dec. 1979). Since some desert rodents are known to hoard seeds in surface caches, the study was specifically designed to determine whether these caches eventually produce more seedlings, on the average, than an adjacent control area. Forty small areas excavated by rodents were located and covered with plastic cages to prevent rodents from reusing the caches. A caged control area was set up adjacent to each of the caged caches. The numbers of seedlings germinating from the caches and from the control areas were then observed. A summary of the data is provided in the accompanying table. Is there sufficient evidence (at $\alpha = .05$) to indicate that the average number of seedlings germinating from the seed caches of desert rodents is significantly higher than the corresponding average for the control areas?

Caches	*Control Areas*
$n_1 = 40$	$n_2 = 40$
$\bar{y}_1 = 5.3$	$\bar{y}_2 = 2.7$
$s_1 = 1.3$	$s_2 = .7$

Source: Reichman, O. J. "Desert granivore foraging and its impact on seed densities and distributions." *Ecology*, Dec. 1979, Vol. 60, pp. 1085–1092. Copyright 1979, the Ecological Society of America. Reprinted by permission.

9.9 Testing the Difference Between Two Population Means: Matched Pairs

It may be possible to acquire more information on the difference between two population means by using data collected in matched pairs instead of independent samples. Consider, for example, an experiment to investigate the effectiveness of cloud seeding in the artificial production of rainfall. Two farming areas with similar past meteorological records were selected for the experiment. One is seeded regularly; the other is left unseeded. The monthly precipitation at the farms will be recorded for 6 randomly selected months. The resulting data, matched on months, can be used to test a hypothesis about the difference between the mean monthly precipitation in the seeded and unseeded areas. The appropriate procedures are summarized in the boxes.

Large-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Matched Pairs

One-Tailed Test

H_0: $(\mu_1 - \mu_2) = D_0$

H_a: $(\mu_1 - \mu_2) > D_0$

[or H_a: $(\mu_1 - \mu_2) < D_0$]

Two-Tailed Test

H_0: $(\mu_1 - \mu_2) = D_0$

H_a: $(\mu_1 - \mu_2) \neq D_0$

Test statistic: $$z = \frac{\bar{d} - D_0}{\sigma_d/\sqrt{n}} \approx \frac{\bar{d} - D_0}{s_d/\sqrt{n}}$$

where $\bar{d}$ and s_d represent the mean and standard deviation of the sample of differences.

Rejection region: $z > z_\alpha$ [or $z < -z_\alpha$]

Rejection region: $|z| > z_{\alpha/2}$

[*Note:* D_0 is our symbol for the particular numerical value specified for $(\mu_1 - \mu_2)$ in H_0. In many applications, we want to hypothesize that there is no difference between the population means; in such cases, $D_0 = 0$.]

Small-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Matched Pairs

One-Tailed Test	*Two-Tailed Test*
H_0: $(\mu_1 - \mu_2) = D_0$	H_0: $(\mu_1 - \mu_2) = D_0$
H_a: $(\mu_1 - \mu_2) > D_0$ [or H_a: $(\mu_1 - \mu_2) < D_0$]	H_a: $(\mu_1 - \mu_2) \neq D_0$

$$\textit{Test statistic:} \quad t = \frac{\bar{d} - D_0}{\sigma_d/\sqrt{n}} \approx \frac{\bar{d} - D_0}{s_d/\sqrt{n}}$$

where $\bar{d}$ and s_d represent the mean and standard deviation of the sample of differences.

Rejection region: $t > t_\alpha$ [or $t > -t_\alpha$]	*Rejection region:* $\lvert t \rvert > t_{\alpha/2}$

where the t-distribution is based on $(n - 1)$ degrees of freedom.

[*Note:* D_0 is our symbol for the particular numerical value specified for $(\mu_1 - \mu_2)$ in the null hypothesis. In many practical applications, we want to hypothesize that there is no difference between the population means; in such cases, $D_0 = 0$.]

Assumptions:
1. The relative frequency distribution of the population of differences is approximately normal.
2. The paired differences are randomly selected from the population of differences.

Warning: When the assumption of normality is grossly violated, the t test may lead to erroneous inferences. In this case, use the nonparametric Wilcoxon test described in Section 15.4.

EXAMPLE 9.15

Consider the cloud seeding experiment to compare monthly precipitation at the two farm areas. Do the data given in Table 9.4 provide sufficient evidence to indicate that the mean monthly precipitation at the seeded farm area exceeds the corresponding mean for the unseeded farm area? Test using $\alpha = .05$.

TABLE 9.4 Monthly Precipitation Data (in Inches) for Example 9.15

Farm Area	1	2	3	4	5	6
Seeded	1.75	2.12	1.53	1.10	1.70	2.42
Unseeded	1.62	1.83	1.40	.75	1.71	2.33
d	.13	.29	.13	.35	−.01	.09

Solution

Let μ_1 and μ_2 represent the mean monthly precipitation values for the seeded and unseeded farm areas, respectively. Since we want to be able to detect $\mu_1 > \mu_2$, we will conduct the one-tailed test:

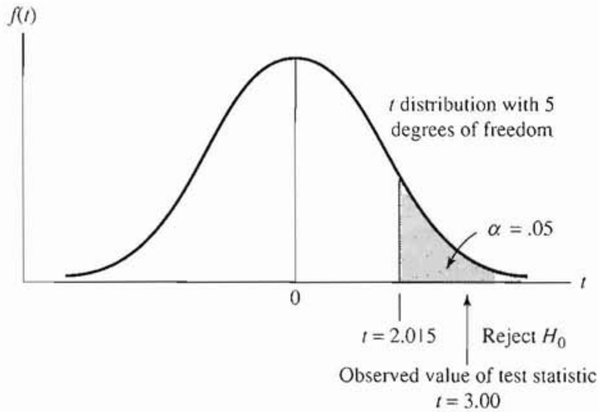
$$H_0\text{:}\quad (\mu_1 - \mu_2) = 0$$

$$H_a\text{:}\quad (\mu_1 - \mu_2) > 0$$

Assuming the differences in monthly precipitation values for the two areas are from an approximately normal distribution, the test statistic will have a t distribution based on $(n - 1) = (6 - 1) = 5$ degrees of freedom. We will reject the null hypothesis if

$$t > t_{.05} = 2.015 \quad \text{(see Figure 9.16)}$$

FIGURE 9.16 ▶
Rejection region for Example 9.15



To conduct the test by hand, we must first calculate the difference d in monthly precipitation at the two farm areas for each month. These differences (where the observations for the unseeded farm area is subtracted from the observation for the seeded area within each pair) are shown in the last row of Table 9.4. Next, we would calculate the mean $\bar{d}$ and standard deviation s_d for this sample of $n = 6$ differences to obtain the test statistic.

Rather than perform these calculations, we will rely on the output from a computer. The MINITAB printout for the analysis is shown in Figure 9.17. The test statistic, shaded in Figure 9.17, is $t = 3.01$.

FIGURE 9.17 ▶
Minitab printout for Example 9.17

```
TEST OF MU = 0.0000 VS MU N.E. 0.0000

              N      MEAN     STDEV   SE MEAN          T    P VALUE
diff          6    0.1633    0.1331    0.0543       3.01      0.030
```

Substituting the values $\bar{d} = .1633$ and $s_d = .1331$ into the formula for the test statistic, we have

$$t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{.1633 - 0}{.1331/\sqrt{6}} = 3.01$$

Since this value of the test statistic exceeds the critical value $t_{.05} = 2.015$, there is sufficient evidence (at $\alpha = .05$) to indicate that the mean monthly precipitation at the seeded farm area exceeds the mean for the unseeded farm area.

The same conclusion can be reached by examining the p-value of the test. The two-tailed p-value, shaded on the MINITAB printout, is .030. Consequently, the one-tailed p-value is $p = .030/2 = .015$. Since this value is less than the chosen α level (.05), we reject H_0.

In the experiment of Example 9.15, why did we collect the data in matched pairs rather than use independent random samples of months, with some assigned to only the seeded area and others to only the unseeded area? The answer is that we expected some months to have more rain than others. To cancel out this variation from month to month, the experiment was designed so that precipitation at both farm areas would be recorded during the same months. Then both farm areas would be subjected to the same weather pattern in a given month. By comparing precipitation *within* each month, we were able to obtain more information on the difference in mean monthly precipitation than we could have obtained by independent random sampling.

EXERCISES

9.40 Researchers at Purdue University compared human real-time scheduling in a processing environment to an automated approach that utilizes computerized robots and sensing devices (*IEEE Transactions*, Mar. 1993). The experiment consisted of eight simulated scheduling problems. Each task was performed by a human scheduler and by the automated system. Performance was measured by the *throughput rate*, defined as the number of good jobs produced weighted by product quality. The resulting throughput rates are shown in the accompanying table. Analyze the data using a test of hypothesis.

Task	*Human Scheduler*	*Automated Method*	*Task*	*Human Scheduler*	*Automated Method*
1	185.4	180.4	5	240.0	269.3
2	146.3	248.5	6	253.8	249.6
3	174.4	185.5	7	238.8	282.0
4	184.9	216.4	8	263.5	315.9

Source: Yih, Y., Liang, T., and Moskowitz, H. "Robot scheduling in a circuit board production line: A hybrid OR/ANN approach." *IEEE Transactions*, Vol. 25, No. 2, March 1993, p. 31 (Table 1).

9.41 For the perception of speech, profoundly deaf persons rely mainly on speechreading, i.e., they perceive spoken language by observing the articulatory movements, facial expressions, and gestures of the speaker. Can speech perception be improved by supplementing the speechreader with auditorily presented information about the prosody of the speech signal? To investigate this phenomenon, 10 normal-hearing subjects participated in an experiment in which they were asked to verbally reproduce sentences spoken but not heard on a video monitor (*Journal of the Acoustical Society of America*, Feb. 1986). The sentences were presented to the subjects under each of two conditions: (1) speechreading with information about the frequency and amplitude of the speech signal (denoted S + F + A), and (2) speechreading only (denoted S). For each of the 10 subjects, the difference between the percentage of correctly reproduced syllables under condition S + F + A and under condition S was calculated. The mean and standard deviation of the differences are as follows:

$\bar{d} = 20.4 \qquad s_d = 17.44$

Test the hypothesis that the mean percentage of correct syllables under condition S + F + A exceeds the corresponding mean under condition S. Use $\alpha = .05$.

9.42 Tetrachlorodibenzo-p-dioxin (TCDD) is a highly toxic substance found in industrial wastes. A study was conducted to determine the amount of TCDD present in the tissues of bullfrogs inhabiting the Rocky Branch Creek in central Arkansas, an area known to be contaminated by TCDD (*Chemosphere*, Feb. 1986). The level of TCDD (in parts per trillion) was measured in several specific tissues of four female bull frogs; the ratio of TCDD in the tissue to TCDD in the leg muscle of the frog was recorded for each. The relative ratios of contaminant for two tissues, the liver and the ovaries, are given for each of the four frogs in the accompanying table. According to the researchers, "the data set suggests that the [mean] relative level of TCDD in the ovaries of female frogs is higher than the [mean] level in the liver of the frogs." Test this claim using $\alpha = .05$.

Frog	A	B	C	D
Liver	11.0	14.6	14.3	12.2
Ovaries	34.2	41.2	32.5	26.2

Source: Korfmacher, W. A., Hansen, E. B., and Rowland, K. L. "Tissue distribution of 2,3,7,8-TCDD in bullfrogs obtained from a 2,3,7,8-TCDD-contaminated area." *Chemosphere*, Vol. 15, No. 2, Feb. 1986, p. 125. Reprinted with permission. Copyright 1986, Pergamon Press, Ltd.

9.43 Merck Research Labs conducted an experiment to evaluate the effect of a new drug using the Single-T Swim maze. Nineteen impregnated dam rats were captured and allocated a dosage of 12.5 milligrams of the drug. One male and one female pup were randomly selected from each resulting litter to perform in the swim maze. Each rat pup is placed in water at one end of the maze and allowed to swim until it successfully escapes at the opposite end. If the rat pup fails to escape after a certain period of time, it is placed at the beginning end of the maze and given another attempt to escape. The experiment is repeated until three successful escapes are accomplished by each rat pup. The number of swims required by each pup to perform three successful escapes is reported in the table on page 464. Is there sufficient evidence of a difference between the mean number of swims required by male and female rat pups? Use the MINITAB printout on page 464 to conduct the test (at $\alpha = .10$).

Litter	*Male*	*Female*	*Litter*	*Male*	*Female*
1	8	5	11	6	5
2	8	4	12	6	3
3	6	7	13	12	5
4	6	3	14	3	8
5	6	5	15	3	4
6	6	3	16	8	12
7	3	8	17	3	6
8	5	10	18	6	4
9	4	4	19	9	5
10	4	4			

Source: Thomas E. Bradstreet, Merck Research Labs, BL 3-2, West Point, Penn. 19486.

```
TEST OF MU = 0.000 VS MU N.E. 0.000

              N      MEAN     STDEV   SE MEAN        T   P VALUE
SwimDiff     19     0.368     3.515     0.806     0.46      0.65
```

9.44 Refer to the *Journal of Environmental Engineering* (Feb. 1986) study of winter heat loss in wastewater treatment clarifiers, Exercise 8.40. The data, reproduced in the table, were used to compare the mean day-long clear-sky solar radiation levels (in BTU/sq. ft.) at two midwest sites. A SAS printout for a test to compare the means follows. Interpret the results of the test.

Date	*St. Joseph, Mo.*	*Iowa Great Lakes*
December 21	782	593
January 6	965	672
January 21	948	750
February 6	1,181	988
February 21	1,414	1,226
March 7	1,633	1,462
March 21	1,852	1,698

Source: Wall, D. J., and Peterson, G. "Model for winter heat loss in uncovered clarifiers." *Journal of Environmental Engineering*, Vol. 112, No. 1, Feb. 1986, p. 128.

```
Analysis Variable : RADDIFF

N Obs              T  Prob>|T|
------------------------------
    7    11.7649303    0.0001
------------------------------
```

9.10 Testing a Population Proportion

In Section 9.3, we gave several examples of a statistical test of hypothesis for a population proportion p. When the sample size is large, the sample proportion of successes $\hat{p}$ is approximately normal and the general formulas for conducting a large-sample z test (given in Section 9.3) can be applied.

The procedure for testing a hypothesis about a population proportion p based on a large sample from the target population is described in the box. (Recall that p represents the probability of success in a binomial experiment.) For the procedure to be valid, the sample size must be sufficiently large to guarantee approximate normality of the sampling distribution of the sample proportion, $\hat{p}$. As with confidence intervals, a general rule of thumb for determining whether n is "sufficiently large" is that both $n\hat{p}$ and $n\hat{q}$ are greater than or equal to 4.

Large-Sample Test of Hypothesis About a Population Proportion

One-Tailed Test	*Two-Tailed Test*
H_0: $p = p_0$	H_0: $p = p_0$
H_a: $p > p_0$ [or H_a: $p < p_0$]	H_a: $p \neq p_0$

Test statistic: $$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

where $q_0 = 1 - p_0$

Rejection region: $z > z_\alpha$ [or $z < -z_\alpha$]	*Rejection region:* $\lvert z\rvert > z_{\alpha/2}$

Assumption: The sample size n is sufficiently large so that the approximation is valid. As a rule of thumb, the condition of "sufficiently large" will be satisfied when $n\hat{p} \geq 4$ and $n\hat{q} \geq 4$.

EXAMPLE 9.16

Controversy surrounds the use of weathering steel in the construction of highway bridges. Critics have recently cited serious corrosive problems with weathering steel and are currently urging states to prohibit its use in bridge construction. On the other hand, the steel corporations claim that these charges are exaggerated and report that 95% of all weathering steel bridges in operation show "good" performance, with no major corrosive damage. To test this claim, a team of engineers and steel industry experts evaluated 60 randomly selected weathering steel bridges and found 54 of them showing "good" performance. Is there evidence, at $\alpha = .05$, that the true proportion of weathering steel highway bridges that show "good" performance is less than .95, the figure quoted by the steel corporations?

Solution

The parameter of interest is a population proportion, p. We want to test

$$H_0:\ p = .95$$
$$H_a:\ p < .95$$

where p is the true proportion of all weathering steel highway bridges that show "good" performance.

At significance level $\alpha = .05$, the null hypothesis will be rejected if

$$z < -z_{.05}$$

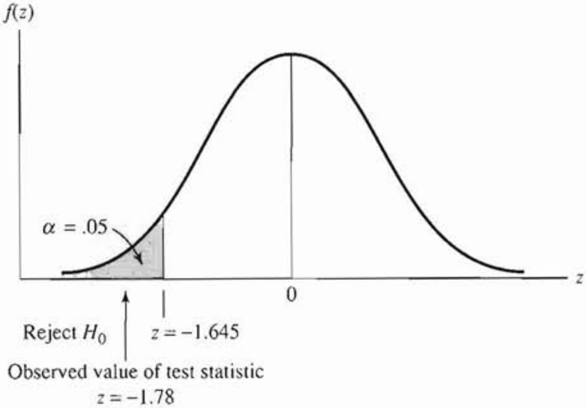
that is, H_0 will be rejected if

$$z < -1.645 \quad \text{(see Figure 9.18)}$$

The sample proportion of bridges that show "good" performance is

$$\hat{p} = \frac{54}{60} = .90$$

FIGURE 9.18 ▶
Rejection region for Example 9.16



Thus, the test statistic has the value

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}} = \frac{.90 - .95}{\sqrt{(.95)(.05)/60}} = -1.78$$

The null hypothesis can be rejected (at $\alpha = .05$), since the computed value of z falls within the rejection region. There is sufficient evidence to support the hypothesis that the proportion of weathering steel highway bridges that show "good" performance is less than .95. [Note that both $n\hat{p} = 60(.90) = 54$ and $n\hat{q} = 60(.10) = 6$ exceed 4. Thus, the sample size is clearly large enough to guarantee the validity of the hypothesis test.]

Although small-sample procedures are available for testing hypotheses about a population proportion, the details are omitted from our discussion. It is our experience that they are of limited utility, since most surveys of binomial populations (for example, opinion polls) performed in the real world use samples that are large enough to employ the techniques of this section.

EXERCISES

9.45 Researchers at the University of Rochester studied the friction that occurs in the paper-feeding process of a photocopier (*Journal of Engineering for Industry*, May 1993). The experiment involved monitoring the displacement of individual sheets of paper in a stack fed through the copier. If no sheet except the top one moved more than 25% of the total stroke distance, the feed was considered successful. In a stack of 100 sheets of paper, the feeding process was successful 94 times. The success rate of the feeder is designed to be .90. Test to determine whether the true success rate of the feeder exceeds .90. Use $\alpha = .10$.

9.46 Staying too long in a spa pool can result in overheating, which in the case of a pregnant woman, may cause fetal malformation. But how long is too long? Based on their work in this area, several researchers hypothesize that 75% of women, immersed in a spa with water temperature of 40°C, will become uncomfortably hot when their ear canal (core) temperature reaches 40°C. As a result, subjective discomfort is suggested as a possible safeguard against overheating. This finding was apparently contradicted by an Australian study of 24 healthy, nonpregnant women (*New England Journal of Medicine*, Sept. 20, 1990). Only 11 of the 24 women (46%) were uncomfortably hot when their core temperature reached 40°C. Test the hypothesis that the true percentage of healthy, nonpregnant women who become uncomfortably hot when their core temperature reaches 40°C is less than 75%. Use $\alpha = .10$.

9.47 Distortions that occur on a computer graphics terminal screen are often due to data being lost in the communications linkage process between the terminal and the computer. A manufacturer of a new data-communications error controller claims that the chance of losing data with the controller in operation is only .01. To test this claim, the communications link between a graphics terminal and computer is monitored with the error controller in operation. Of a random sample of 200 on-screen graphic items, six were distorted because of data errors in the communications link. Does the sample evidence refute the manufacturer's claim? Use $\alpha = .05$.

9.48 The National Science Foundation, in a survey of 2,237 engineering graduate students who earned their Ph.D.degrees, found that 607 were U.S. citizens; the majority (1,630) of the Ph.D. degrees were awarded to foreign nationals (*Science*, Sept. 24, 1993). Conduct a test to determine whether the true proportion of engineering Ph.D. degrees awarded to foreign nationals exceeds .5. Use $\alpha = .01$.

9.49 Concerned about airport and airline security, the Federal Aviation Administration (FAA) has begun imposing sanctions against airlines that fail security tests. One series of tests conducted at Los Angeles International Airport (LAX) showed that security guards detected only 72 of the 100 mock weapons carried on by FAA inspectors or included in their carry-on luggage (*Gainesville Sun*, Dec. 11, 1987). According to the FAA, this "detection rate was well below the national rate of .80." Is there sufficient evidence to conclude that the mock weapon detection rate at LAX is less than the national rate of .80? Test using $\alpha = .10$.

9.50 As part of the evaluation for an environmental impact statement of proposed hydroelectric design on the Stikine River in British Columbia, researchers conducted preliminary investigations of the effects of human-induced disturbances on the behavior of the resident mountain goat population (*Environmental Management*, Mar. 1983). Goat responses to exploration activities, including close-flying helicopters, fixed-wing aircraft, human bipedal movement, and loud blasts from geological drilling activities, were recorded for $n = 804$ goats. The researchers observed that 265 goats displayed a severe flight response to local rock or plant cover. Test the hypothesis that over 30% of the resident mountain goats will show a severe response to human-induced disturbances. Use $\alpha = .05$.

9.51 Architects and engineers, faced with public-sector (i.e., government) cuts, are turning to private-sector clients to fill an increasing share of their workloads. According to some researchers, the decrease in popularity of public-sector work among small, medium, and large architecture–engineering (A–E) firms has been dramatic. Two years ago, one-third of all A–E firms reported they relied on public sector projects for most (if not all) of their work. In a recent survey of 60 A–E firms, 10 indicated that they depended so heavily on government contracts. Do the sample data provide sufficient evidence to conclude that the percentage of A–E firms that rely heavily on public-sector clients has declined during the past 2 years? Use $\alpha = .05$.

9.11 Testing the Difference Between Two Population Proportions

The method for performing a large-sample test of hypothesis about $(p_1 - p_2)$, the difference between two binomial proportions, is outlined in the accompanying box.

When testing the null hypothesis that $(p_1 - p_2)$ equals some specified difference—say, D_0—we make a distinction between the case $D_0 = 0$ and the case $D_0 \neq 0$. For the special case $D_0 = 0$, i.e., when we are testing H_0: $(p_1 - p_2) = 0$ or, equivalently, H_0: $p_1 = p_2$, the best estimate of $p_1 = p_2 = p$ is found by dividing the total number of successes in the combined samples by the total number of observations in the two samples. That is, if y_1 is the number of successes in sample 1 and y_2 is the number of successes in sample 2, then

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2}$$

In this case, the best estimate of the standard deviation of the sampling distribution of $(\hat{p}_1 - \hat{p}_2)$ is found by substituting $\hat{p}$ for both p_1 and p_2:

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \approx \sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

For all cases in which $D_0 \neq 0$ [for example, when testing H_0: $(p_1 - p_2) = .2$], we use $\hat{p}_1$ and $\hat{p}_2$ in the formula for $\sigma_{(\hat{p}_1 - \hat{p}_2)}$. However, in most practical situations, we will want to test for a difference between proportions—that is, we will want to test H_0: $(p_1 - p_2) = 0$.

Large-Sample Test of Hypothesis About $(p_1 - p_2)$: Independent Samples

One-Tailed Test

H_0: $(p_1 - p_2) = D_0$

H_a: $(p_1 - p_2) > D_0$
[or H_a: $(p_1 - p_2) < D_0$]

Two-Tailed Test

H_0: $(p_1 - p_2) = D_0$

H_a: $(p_1 - p_2) \neq D_0$

$$\textit{Test statistic:} \quad z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}$$

Rejection region: $z > z_\alpha$
[or $z < -z_\alpha$]

Rejection region: $|z| > z_{\alpha/2}$

When $D_0 \neq 0$,

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} \approx \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

where $\hat{q}_1 = 1 - \hat{p}_1$ and $\hat{q}_2 = 1 - \hat{p}_2$.

When $D_0 = 0$,

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} \approx \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where the total number of successes in the combined sample is $(y_1 + y_2)$ and

$$\hat{p}_1 = \hat{p}_2 = \hat{p} = \frac{y_1 + y_2}{n_1 + n_2}$$

Assumption: The sample sizes, n_1 and n_2, are sufficiently large. This will be satisfied if $n_1\hat{p}_1 \geq 4$, $n_1\hat{q}_1 \geq 4$, and $n_2\hat{p}_2 \geq 4$, $n_2\hat{q}_2 \geq 4$.

The sample sizes n_1 and n_2 must be sufficiently large to ensure that the sampling distributions of $\hat{p}_1$ and $\hat{p}_2$, and hence of the difference $(\hat{p}_1 - \hat{p}_2)$, are approximately normal. The rule of thumb used to determine if the sample sizes are "sufficiently large" is the same as that given in Section 8.9, namely, that the quantities $n_1\hat{p}_1$, $n_2\hat{p}_2$, $n_1\hat{q}_1$, and $n_2\hat{q}_2$ are all greater than or equal to 4. [*Note:* If the sample sizes are not sufficiently large, p_1 and p_2 can be compared using a technique to be discussed in Chapter 10.]

EXAMPLE 9.17

Recently there have been intensive campaigns encouraging people to save energy by carpooling to work. Some cities have created an incentive for carpooling by designating certain highway traffic lanes as "car-pool only" (i.e., only cars with two or more passengers can use these lanes). To evaluate the effectiveness of this plan, toll booth personnel in one city monitored 2,000 randomly selected cars prior to establishing car-pool-only lanes, and 1,500 cars after the car-pool-only lanes were established. The results of the study are shown in Table 9.5, where y_1 and y_2 represent the numbers of cars with two or more passengers (i.e., car-pool riders) in the "before" and "after" samples, respectively. Do the data indicate that the fraction of cars with car-pool riders has increased over this period? Use $\alpha = .05$.

TABLE 9.5 Results of Carpooling Study, Example 9.16

	Before Car-Pool Lanes Established	After Car-Pool Lanes Established
Sample Size	$n_1 = 2{,}000$	$n_2 = 1{,}500$
Car-Pool Riders	$y_1 = 655$	$y_2 = 576$

Solution

If we define p_1 and p_2 as the true proportions of cars with car-pool riders before and after establishing car-pool lanes, respectively, the elements of our test are:

$$H_0\text{:}\quad (p_1 - p_2) = 0$$
$$H_a\text{:}\quad (p_1 - p_2) < 0$$

(The test is one-tailed since we are interested only in determining whether the proportion of cars with car-pool riders has increased, i.e., whether $p_2 > p_1$.)

Test statistic: $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - 0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}}$

Rejection region: $\alpha = .05$

$z < -z_\alpha = -z_{.05} = -1.645$ (see Figure 9.19)

We now calculate the sample proportions of cars with car-pool riders:

$$\hat{p}_1 = \frac{652}{2{,}000} = .326 \qquad \hat{p}_2 = \frac{576}{1{,}500} = .384$$

The test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sigma_{(\hat{p}_1 - \hat{p}_2)}} \approx \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

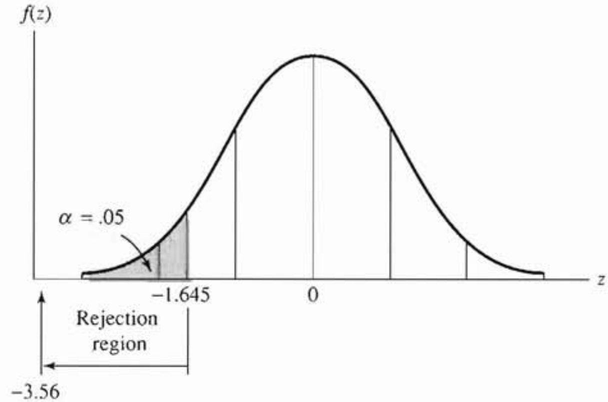


FIGURE 9.19 ▶
Rejection region for Example 9.17

where

$$\hat{p} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{652 + 576}{2{,}000 + 1{,}500} = .351$$

Thus,

$$z = \frac{.326 - .384}{\sqrt{(.351)(.649)\left(\frac{1}{2{,}000} + \frac{1}{1{,}500}\right)}} = \frac{-.058}{.0163} = -3.56$$

Since $z = -3.56$ falls in the rejection region, there is sufficient evidence at $\alpha = .05$ to conclude that the proportion of all cars with car-pool riders has increased after establishing car-pool lanes. We could place a confidence interval on $(p_1 - p_2)$ if we were interested in estimating the extent of the increase.

EXERCISES

9.52 Scientists have linked a catastrophic decline in the number of frogs inhabiting the world to ultraviolet radiation from the earth's tattered ozone layer (*Tampa Tribune*, Mar. 1, 1994). The Pacific tree frog, however, is not believed to be in decline because it produces an enzyme that appears to protect its eggs from ultraviolet radiation. Researchers at Oregon State University compared the hatching rates of two groups of Pacific tree frog eggs. One group of eggs was shielded with ultraviolet-blocking sun shades, whereas the second group was not. The number of eggs successfully hatched in each group is provided in the table. Compare the hatching rates of the two groups of Pacific tree frog eggs with a test of hypothesis. Use $\alpha = .01$.

	Sun-Shaded Eggs	*Unshaded Eggs*
Total Number	70	80
Number Hatched	34	31

9.53 Calcium blockers are among several classes of medicines commonly prescribed to relieve high blood pressure. A study in Denmark has found that calcium blockers may also be effective in reducing the risk of heart attacks (*Tampa Tribune*, Mar. 23, 1990). A total of 897 Danish patients, each recovering from a heart attack, were given a daily dose of the drug Verapamil, a calcium blocker. After 18 months of follow-up, 146 of these patients had recurring heart attacks. In a control group of 878 people—each of whom took placebos—180 had a heart attack. Do the data provide sufficient evidence to infer that calcium blockers are effective in reducing the risk of heart attacks? Test using $\alpha = .01$.

9.54 Every 10 years the Mechanics Division of ASEE conducts a nationwide survey on undergraduate mechanics education at colleges and universities. In 1985, 66 of the 100 colleges surveyed covered fluid statics in their undergraduate engineering program, compared to 43% in the 1975 survey (*Engineering Education*, Apr. 1986). Assuming that 100 colleges were also surveyed in 1975, conduct a test to determine whether the percentage of colleges covering fluid statics increased from 1975 to 1985. Use $\alpha = .01$.

9.55 A study was conducted to determine the impact of a multifunction workstation (MFWS) on the way managers work (*Datamation*, Feb. 15, 1986). Two groups of managers at a St. Louis-based defense agency took part in the survey: a test group consisting of 12 managers who currently use MFWS software and a control group of 25 non-MFWS users. One question on the survey concerned the information sources of the managers. In the test group (MFWS users), 4 of the 12 managers reported that their major source of information is the computer, whereas 2 of the 25 in the control group (non-MFWS users) rely on the computer as their major source of information.

a. Is there evidence of a difference between the proportions of MFWS users and non-MFWS users who rely on the computer as their major information source? Test using $\alpha = .10$.

b. Are the sample sizes large enough for the approximation procedure, part **a**, to be valid?

9.56 Home solar heating systems can be categorized into two groups, *passive* solar heating systems and *active* solar heating systems. In a passive solar heating system, the house itself is a solar energy collector, whereas in an active solar heating system, elaborate mechanical equipment is used to convert the sun's rays into heat. Consider the difference between the proportions of passive solar and active solar heating systems that require less than 200 gallons of oil per year in fuel consumption. Independent random samples of 50 passive and 50 active solar-heated homes are selected and the numbers that required less than 200 gallons of oil last year are noted, with the results given in the table. Is there evidence of a difference between the proportions of passive and active solar-heated homes that required less than 200 gallons of oil in fuel consumption last year? Test at a level of significance of $\alpha = .02$.

	Passive Solar	*Active Solar*
Number of homes	50	50
Number that required less than 200 gallons of oil last year	37	46

9.57 In 1982, 371 manufacturing and retailing companies were surveyed to determine the extent to which logistics information systems were implemented. A follow-up survey of 459 firms was conducted in 1987 to measure the 5-year trend in computerization of logistics information (*Industrial Engineering*, July 1990). One of the survey items focused on the percentage of firms that had computerized external market data. From 1982 to 1987, this percentage increased from 25% to 33%. Use this information to test for a significant increase in the percentage of firms with computerized external market data over the 5-year period. Test using $\alpha = .05$.

9.12 Testing a Population Variance

Recall from Section 8.10 that the pivotal statistic for estimating a population variance σ^2 does not possess a normal (z) distribution. Therefore, we cannot apply the procedure outlined in Section 9.4 when testing hypotheses about σ^2.

When the sample is selected from a normal population, however, the pivotal statistic possesses a chi-square (χ^2) distribution and the test can be conducted as outlined in the box. Note that the assumption of normality is required regardless of whether the sample size n is large or small.

Test of Hypothesis About a Population Variance σ^2

One-Tailed Test

H_0: $\sigma^2 = \sigma_0^2$

H_a: $\sigma^2 > \sigma_0^2$

[or H_a: $\sigma^2 < \sigma_0^2$]

Two-Tailed Test

H_0: $\sigma^2 = \sigma_0^2$

H_a: $\sigma^2 \neq \sigma_0^2$

Test statistic: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

Rejection region (one-tailed): $\chi^2 > \chi^2_{\alpha}$ (or $\chi^2 < \chi^2_{1-\alpha}$)

Rejection region (two-tailed): $\chi^2 < \chi^2_{1-\alpha/2}$ or $\chi^2 > \chi^2_{\alpha/2}$

where χ^2_{α} and $\chi^2_{1-\alpha}$ are values of χ^2 that locate an area of α to the right and α to the left, respectively, of a chi-square distribution based on $(n - 1)$ degrees of freedom.

[*Note:* σ_0^2 is our symbol for the particular numerical value specified for σ^2 in the null hypothesis.]

Assumption: The population from which the random sample is selected has an approximate normal distribution.

EXAMPLE 9.18

Refer to Example 8.15 concerning the variability of the amount of fill at a cannery. Suppose regulatory agencies specify that the standard deviation of the amount of fill should be less than .1 ounce. The quality control supervisor sampled $n = 10$ cans and mesured the amount of fill in each. The data are reproduced here. Does this information provide sufficient evidence to indicate that the standard deviation σ of the fill measurements is less than .1 ounce?

7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02

Solution

Since the null and alternative hypotheses must be stated in terms of σ^2 (rather than σ), we will want to test the null hypothesis that $\sigma^2 = .01$ against the alternative that $\sigma^2 < .01$. Therefore, the elements of the test are

H_0: $\sigma^2 = .01$ (i.e., $\sigma = .1$)

H_a: $\sigma^2 < .01$ (i.e., $\sigma < .1$)

Assumption: The populaton of fill amounts is approximately normal.

Test statistic: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

Rejection region: The smaller the value of s^2 we observe, the stronger the evidence in favor of H_a. Thus, we reject H_0 for "small values" of the test statistic. With $\alpha = .05$ and 9 df, the χ^2 value for rejection is found in Table 8 of Appendix II and pictured in Figure 9.20. We will reject H_0 if $\chi^2 < 3.32511$. (Remember that the area given in Table 8 of Appendix II is the area to the *right* of the numerical value in the table. Thus, to determine the lower-tail value that has $\alpha = .05$ to its *left*, we use the $\chi^2_{.95}$ column in Table 8.)

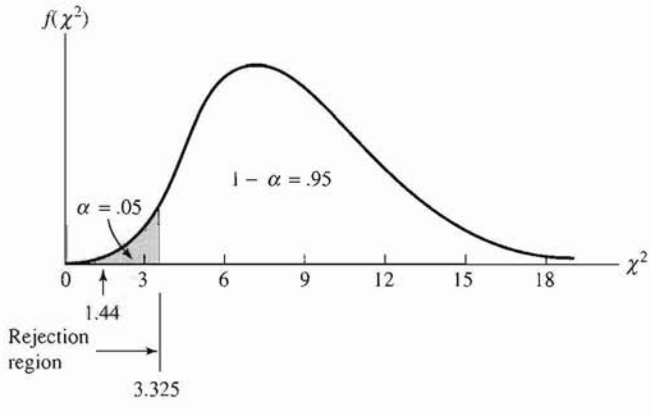


FIGURE 9.20 ► **Rejection region for Example 9.18**

To compute the test statistic, we need to find the sample standard deviation, s. Numerical descriptive statistics for the sample data are provided in the SAS printout shown in Figure 9.21. The value of s, shaded in Figure 9.21, is $s = .043$. Substituting $s = .043$, $n = 10$, and $\sigma_0^2 = .01$ into the formula for the test statistic, we obtain

$$\chi^2 = \frac{(10-1)(.043)^2}{.01} = 1.66$$

Conclusion: Since the test statistic, $\chi^2 = 1.66$, is less than 3.32511, the supervisor can conclude (at $\alpha = .05$) that the variance of the population of all amounts of fill is less than .01 ($\sigma < .1$). If this procedure is repeatedly used, it will incorrectly reject H_0 only 5% of the time. Thus, the quality control supervisor is confident in the decision that the cannery is operating within the desired limits of variability.

FIGURE 9.21 ▶
SAS printout: Descriptive statistics for Example 9.18

```
Variable=FILL
                              Moments

           N              10    Sum Wgts          10
           Mean        7.989    Sum            79.89
           Std Dev  0.043063    Variance    0.001854
           Skewness  -0.8538    Kurtosis    0.479371
           USS      638.2579    CSS          0.01669
           CV       0.539032    Std Mean    0.013618
           T:Mean=0 586.6587    Prob>|T|      0.0001
           Sgn Rank     27.5    Prob>|S|      0.0020
           Num ^= 0       10

                          Quantiles(Def=5)

            100% Max    8.04         99%        8.04
             75% Q3     8.02         95%        8.04
             50% Med   7.995         90%       8.035
             25% Q1     7.96         10%        7.93
              0% Min     7.9          5%         7.9
                                      1%         7.9

            Range       0.14
            Q3-Q1       0.06
            Mode        7.96
```

EXERCISES

9.58 Refer to the *Journal for Engineering for Industry* (May 1993) study of deep hole drilling under drill chip congestion, Exercise 9.21. Test to determine whether the true standard deviation of drill chip lengths differs from 75 mm. Recall that for $n = 50$ drill chips, $s = 50.2$.

9.59 Recording electrical activity of the brain is important in clinical problems as well as in neurophysiological research. To improve the signal-to-noise ratio (SNR) in the electrical activity, it is necessary to repeatedly stimulate subjects and average the responses—a procedure that assumes that single responses are homogeneous. A study was conducted to test the homogeneous signal theory (*IEEE Engineering in Medicine and Biology Magazine*, Mar. 1990). The null hypothesis is that the variance of the SNR readings of subjects equals the "expected" level under the homogeneous signal theory. For this study, the "expected" level was assumed to be .54. If the SNR variance exceeds this level, the researchers will conclude that the signals are nonhomogeneous.

a. Set up the null and alternative hypotheses for the researchers.
b. SNRs recorded for a sample of 41 normal children ranged from .03 to 3.0. Use this information to obtain an estimate of the sample standard deviation. [*Hint:* Assume that the distribution of SNRs is normal, and that most of the SNRs in the population will fall within $\mu \pm 2\sigma$, i.e., from $\mu - 2\sigma$ to $\mu + 2\sigma$. Note that the range of the interval equals 4σ.]
c. Use the estimate of s in part **b** to conduct the test of part **a**. Test using $\alpha = .10$.

9.60 The most common method of disinfecting water for potable use is free residual chlorination. Recently, preammoniation (i.e., the addition of ammonia to the water prior to applying free chlorine) has received considerable attention as an alternative treatment. In one study, 44 water specimens treated with preammoniation were found to have a mean effluent turbidity of 1.8 and a standard deviation of .16 (*American Water Works Journal*, Jan. 1986). Is there sufficient evidence to indicate that the variance of the effluent turbidity in water specimens disinfected by the preammoniation method exceeds .0016? (The value .0016 represents the known effluent turbidity variance of water specimens treated with free chlorine.) Test using $\alpha = .01$.

9.61 In any canning process, a manufacturer will lose money if the cans contain either significantly more or significantly less than is claimed on the label. Accordingly, canners pay close attention to the amount of their product being dispensed by the can-filling machines. Consider a company that produces a fast-drying rubber cement in 32-ounce aluminum cans. A quality control inspector is interested in testing whether the variance of the amount of rubber cement dispensed into the cans is more than .3. If so, the dispensing machine is in need of adjustment. Since inspection of the canning process requires that the dispensing machines be shut down, and shutdowns for any lengthy period of time cost the company thousands of dollars in lost revenue, the inspector is able to obtain a random sample of only 10 cans for testing. After measuring the weights of their contents, the inspector computes the following summary statistics:

$\bar{x} = 31.55$ ounces $\quad s = .48$ ounce

a. Does the sample evidence indicate that the dispensing machines are in need of adjustment? Test at significance level $\alpha = .05$.

b. What assumption is necessary for the hypothesis test of part **a** to be valid?

9.62 Polychlorinated biphenyls (PCBs), used in the manufacture of large electrical transformers and capacitors, are extremely hazardous contaminants when released into the environment. The Environmental Protection Agency (EPA) is experimenting with a new device for measuring PCB concentration in fish. To check the precision of the new instrument, seven PCB readings were taken on the same fish sample. The data are recorded here (in parts per million):

6.2 5.8 5.7 6.3 5.9 5.8 6.0

Suppose the EPA requires an instrument that yields PCB readings with a variance of less than .1. Does the new instrument meet the EPA's specifications? Test at $\alpha = .05$.

9.13 Testing the Ratio of Two Population Variances

As in the one-sample case, the pivotal statistic for comparing two population variances, σ_1^2 and σ_2^2, has a nonnormal sampling distribution. Recall from Section 8.11 that the ratio of the sample variances s_1^2/s_2^2 possesses, under certain conditions, an F distribution.

The elements of the hypothesis test for the ratio of two population variances, σ_1^2/σ_2^2, are given in the box.

Test of Hypothesis for the Ratio of Two Population Variances σ_1^2/σ_2^2: Independent Samples

One-Tailed Test	*Two-Tailed Test*
H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$	H_0: $\dfrac{\sigma_1^2}{\sigma_2^2} = 1$
H_a: $\dfrac{\sigma_1^2}{\sigma_2^2} > 1$	H_a: $\dfrac{\sigma_1^2}{\sigma_2^2} \neq 1$
$\left[\text{or, } H_a\text{: } \dfrac{\sigma_1^2}{\sigma_2^2} < 1\right]$	
Test statistic: $F = \dfrac{s_1^2}{s_2^2}$ $\left[\text{or, } F = \dfrac{s_2^2}{s_1^2}\right]$	*Test statistic:* $F = \dfrac{\text{Larger sample variance}}{\text{Smaller sample variance}} = \begin{cases} \dfrac{s_1^2}{s_2^2} & \text{when } s_1^2 > s_2^2 \\ \dfrac{s_2^2}{s_1^2} & \text{when } s_2^2 > s_1^2 \end{cases}$
Rejection region: $F > F_\alpha$	*Rejection region:* $F > F_{\alpha/2}$

where F_α and $F_{\alpha/2}$ are values that locate area α and $\alpha/2$, respectively, in the upper tail of the F distribution with ν_1 = numerator degrees of freedom (i.e., the df for the sample variance in the numerator) and ν_2 = denominator degrees of freedom (i.e., the df for the sample variance in the denominator).

Assumptions:
1. Both of the populations from which the samples are selected have relative frequency distributions that are approximately normal.
2. The random samples are selected in an independent manner from the two populations.

EXAMPLE 9.19

Heavy doses of ethylene oxide (ETO) in rabbits have been shown to alter significantly the DNA structure of cells. Although it is a known mutagen and suspected carcinogen, ETO is used quite frequently in sterilizing hospital supplies. A study was conducted to investigate the effect of ETO on hospital personnel involved with the sterilization process. Thirty-one subjects were randomly selected and assigned to one of two tasks. Eighteen subjects were assigned the task of opening the sterilization package that contains ETO (task 1). The remaining 13 subjects were assigned the task of opening and unloading the sterilizer gun filled with ETO (task 2). After the tasks were performed, researchers measured the amount of ETO (in milligrams) present in the bloodstream of each subject. A summary of the results appears in Table 9.6. Do the data provide sufficient evidence to indicate a difference in the variability of the ETO levels in subjects assigned to the two tasks? Test using $\alpha = .10$.

TABLE 9.6 Summary Data for Example 9.19

	Task 1	Task 2
Sample Size	18	13
Mean	5.90	5.60
Standard Deviation	1.93	3.10

Solution

Let

σ_1^2 = Population variance of ETO levels in subjects assigned task 1

σ_2^2 = Population variance of ETO levels in subjects assigned task 2

For this test to yield valid results, we must assume that both samples of ETO levels come from normal populations and that the samples are independent.

The hypotheses of interest are then

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \quad (\sigma_1^2 = \sigma_2^2)$$

$$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \quad (\sigma_1^2 \neq \sigma_2^2)$$

The nature of the F tables given in Appendix II affects the form of the test statistic. To form the rejection region for a two-tailed F test we want to make certain that the upper tail is used, because only the upper-tail values of F are shown in Tables 9–12 of Appendix II. To accomplish this, **we will always place the larger sample variance in the numerator of the F test statistic**. This has the effect of doubling the tabulated value for α, since we double the probability that the F ratio will fall in the upper tail by always placing the larger sample variance in the numerator. That is, we make the test two-tailed by putting the larger variance in the numerator rather than establishing rejection regions in both tails.

Thus, for our example, we have a denominator s_1^2 with df $= n_1 - 1 = 17$ and a numerator s_2^2 with df $= n_2 - 1 = 12$. Therefore, the test statistic will be

$$F = \frac{\text{Larger sample variance}}{\text{Smaller sample variance}} = \frac{s_2^2}{s_1^2}$$

and we will reject H_0: $\sigma_1^2 = \sigma_2^2$ for $\alpha = .10$ when the calculated value of F exceeds the tabulated value:

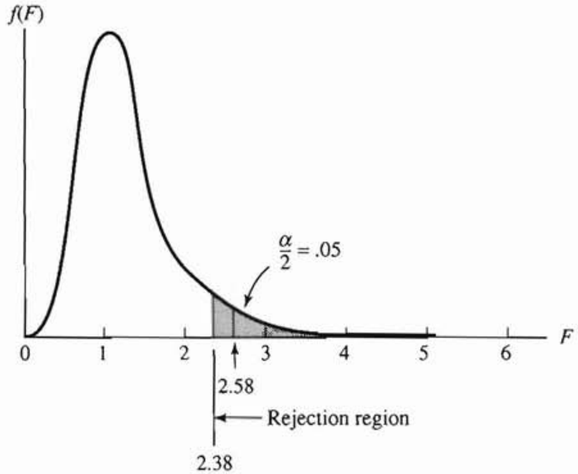
$$F_{\alpha/2} = F_{.05} = 2.38$$

We can now calculate the value of the test statistic and complete the analysis:

$$F = \frac{s_2^2}{s_1^2} = \frac{(3.10)^2}{(1.93)^2} = \frac{9.61}{3.72} = 2.58$$

When we compare this to the rejection region shown in Figure 9.22, we see that $F = 2.58$ falls in the rejection region. Therefore, the data provide sufficient evidence to indicate that the population variances differ. It appears that hospital personnel involved with opening the sterilization package (task 1) have less variable ETO levels than those involved with opening and unloading the sterilizer gun (task 2).

FIGURE 9.22 ▶
Rejection region for Example 9.19



What would you have concluded in Example 9.19 if the value of F calculated from the samples had not fallen in the rejection region? Would you conclude that the null hypothesis of equal variances is true? No, because then you risk the possibility of a Type II error (failing to reject H_0 if H_a is true) without knowing the value of β, the probability of failing to reject H_0: $\sigma_1^2 = \sigma_2^2$ if in fact it is false. Since we will not consider the calculation of β for specific alternatives, when the F statistic does not fall in the rejection region, we simply conclude that insufficient sample evidence exists to refute the null hypothesis that $\sigma_1^2 = \sigma_2^2$.

Example 9.19 illustrates the technique for calculating the test statistic and rejection region for a two-tailed test to avoid the problem of locating an F value in the lower tail of the F distribution. In a one-tailed test this is much easier to accomplish since we can control how we specify the ratio of the population variances in H_0 and H_a. That is, we can always make a one-tailed test an *upper-tailed* test. For example, if we want to test whether σ_1^2 is greater than σ_2^2, then we write the alternative hypothesis as

$$H_a\text{:}\ \frac{\sigma_1^2}{\sigma_2^2} > 1 \quad (\text{i.e., } \sigma_1^2 > \sigma_2^2)$$

and the appropriate test statistic is $F = s_1^2/s_2^2$. Conversely, if we want to test whether σ_1^2 is less than σ_2^2 (i.e, whether σ_2^2 is greater than σ_1^2), we write

$$H_a\colon \ \frac{\sigma_2^2}{\sigma_1^2} > 1 \quad (\text{i.e., } \sigma_2^2 > \sigma_1^2)$$

and the corresponding test statistic is $F = s_2^2/s_1^2$.

EXERCISES

9.63 Refer to Exercise 9.35. Recall that an *Environmental Science & Technology* study was conducted to compare the mean oxon/thion ratios at a California orchard under two weather conditions—foggy and clear/cloudy. Test the assumption of equal variances required for the comparison of means to be valid. Use $\alpha = .05$.

Date	*Condition*	*Thion*	*Oxon*	*Oxon/Thion Ratio*
Jan. 15	Fog	38.2	10.3	.270
17	Fog	28.6	6.9	.241
18	Fog	30.2	6.2	.205
19	Fog	23.7	12.4	.523
20	Fog	62.3	(Air sample lost)	—
20	Clear	74.1	45.8	.618
21	Fog	88.2	9.9	.112
21	Clear	46.4	27.4	.591
22	Fog	135.9	44.8	.330
23	Fog	102.9	27.8	.270
23	Cloudy	28.9	6.5	.225
25	Fog	46.9	11.2	.239
25	Clear	44.3	16.6	.375

Source: Selber, J. N., et al. "Air and fog deposition residues of four organophosphate insecticides used on dormant orchards in the San Joaquin Valley, California." *Environmental Science & Technology*, Vol. 27, No. 10, Oct. 1993, p. 2240 (Table V).

9.64 Wet samplers are standard devices used to measure the chemical composition of precipitation. The accuracy of the wet deposition readings, however, may depend on the number of samplers stationed in the field. Experimenters in The Netherlands collected wet deposition measurements using anywhere from one to eight identical wet samplers (*Atmospheric Environment*, Vol. 24A, 1990). For each sampler (or sampler combination) data was collected every 24 hours for an entire year; thus, 365 readings were collected per sampler (or sampler combination). When one wet sampler was used, the standard deviation of the hydrogen readings (measured as percentage relative to the average reading from all eight samplers) was 6.3%. When three wet samplers were used, the standard deviation of the hydrogen readings (measured as percentage relative to the average reading from all eight samplers) was 2.6%. Conduct a test to compare the variation in hydrogen readings for the two sampling schemes (i.e., one wet sampler versus three wet samplers). Test using $\alpha = .05$.

9.65 An experiment was conducted to study the effect of reinforced flanges on the torsional capacity of reinforced concrete T-beams (*Journal of the American Concrete Institute*, Jan.–Feb. 1986). Several different types of

T-beams were used in the experiment, each type having a different flange width. The beams were tested under combined torsion and bending until failure (cracking). One variable of interest is the cracking torsion moment at the top of the flange of the T-beam. Cracking torsion moments for eight beams with 70-cm slab widths and eight beams with 100-cm slab widths follow:

70-cm slab width: 6.00, 7.20, 10.20, 13.20, 11.40, 13.60, 9.20, 11.20
100-cm slab width: 6.80, 9.20, 8.80, 13.20, 11.20, 14.90, 10.20, 11.80

a. Is there evidence of a difference in the variation in the cracking torsion moments of the two types of T-beams? Use $\alpha = .10$.
b. What assumptions are required for the test to be valid?

9.66 Refer to the general trace organic monitoring study discussed in Exercise 7.21. The total organic carbon (TOC) level was measured in water samples collected at two sewage treatment sites in England. The accompanying table gives the summary information on the TOC levels (measured in mg/l) found in the rivers adjacent to the two sewage facilities. Since the river at the Foxcote sewage treatment works was subject to periodic spillovers, not far upstream of the plant's intake, it is believed that the TOC levels found at Foxcote will have greater variation than the levels at Bedford. Does the sample information support this hypothesis? Test at $\alpha = .05$.

Bedford	*Foxcote*
$n_1 = 61$	$n_2 = 52$
$\bar{y}_1 = 5.35$	$\bar{y}_2 = 4.27$
$s_1 = .96$	$s_2 = 1.27$

Source: Pinchin, M. J. "A study of the trace organics profiles of raw and potable water systems." *Journal of the Institute of Water Engineers & Scientists*, Vol. 40, No. 1, Feb. 1986, p. 87.

9.67 Refer to the speechreading study introduced in Exercise 9.41. A second experiment was conducted to compare the variability in the sentence perception of normal-hearing individuals with no prior experience in speechreading to those with experience in speechreading. The sample consisted of 24 inexperienced and 12 experienced subjects. All subjects were asked to verbally reproduce sentences under several conditions, one of which was speechreading supplemented with sound-pressure information. A summary of the results (percentage of correct syllables) for the two groups is given in the table. Conduct a test to determine whether the variance in the percentage of correctly reproduced syllables differs between the two groups of speechreaders. Test using $\alpha = .10$.

Inexperienced Speechreaders	*Experienced Speechreaders*
$n_1 = 24$	$n_2 = 12$
$\bar{y}_1 = 87.1$	$\bar{y}_2 = 86.1$
$s_1 = 8.7$	$s_2 = 12.4$

Source: Breeuwer, M., and Plomp, R. "Speechreading supplemented with auditorily presented speech parameters." *Journal of the Acoustical Society of America*, Vol. 79, No. 2, Feb. 1986, p. 487.

OPTIONAL EXERCISES

9.68 Suppose we want to test H_0: $\sigma_1^2 = \sigma_2^2$ versus H_a: $\sigma_1^2 \neq \sigma_2^2$. Show that the rejection region given by

$$\frac{s_1^2}{s_2^2} > F_{\alpha/2} \quad \text{or} \quad \frac{s_1^2}{s_2^2} < F_{(1-\alpha/2)}$$

where F depends on $\nu_1 = (n_1 - 1)$ df and $\nu_2 = (n_2 - 1)$ df, is equivalent to the rejection region given by

$$\frac{s_1^2}{s_2^2} > F_{\alpha/2} \quad \text{where } F \text{ depends on } \nu_1 \text{ numerator df and } \nu_2 \text{ denominator df}$$

or

$$\frac{s_2^2}{s_1^2} > F^*_{\alpha/2} \quad \text{where } F^* \text{ depends on } \nu_2 \text{ numerator df and } \nu_1 \text{ denominator df}$$

[*Hint:* Use the fact (proof omitted) that

$$F_{(1-\alpha/2)} = \frac{1}{F^*_{\alpha/2}}$$

where F depends on ν_1 numerator df and ν_2 denominator df and F^* depends on ν_2 numerator df and ν_1 denominator df.]

9.69 Use the results of Optional Exercise 9.68 to show that

$$P\left(\frac{\text{Larger sample variance}}{\text{Smaller sample variance}} > F_{\alpha/2}\right) = \alpha$$

where F depends on numerator df = [(Sample size for numerator sample variance) − 1] and denominator df = [(Sample size for denominator sample variance) − 1]. [*Hint:* First write

$$P\left(\frac{\text{Larger sample variance}}{\text{Smaller sample variance}} > F_{\alpha/2}\right) = P\left(\frac{s_1^2}{s_2^2} > F_{\alpha/2} \quad \text{or} \quad \frac{s_2^2}{s_1^2} > F_{\alpha/2}\right)$$

Then use the fact that $P(F > F_{\alpha/2}) = \alpha/2$.]

9.14 Summary

This chapter presents the basic concepts of a statistical **test of a hypothesis** about one or more population parameters. Tests of hypotheses are used when the ultimate practical objective of an inference is to reach a decision about the value(s) of the parameter(s). We can evaluate the goodness of the inference in terms of α and β, the probabilities of making incorrect decisions.

The close relationship between estimation and hypothesis testing is apparent when we compare the statistics employed for these two purposes. The statistics used to construct confidence intervals for parameters in Chapter 8 were then used to test hypotheses about the same parameters in Chapter 9. These tests are summarized in Tables 9.7a and 9.7b.

In the following chapters, we will present some very useful methodology for analyzing multivariable experiments. As you will subsequently learn, the confidence intervals and tests that we will employ are based on an assumption of normality. Thus, the statistics that we will use to construct confidence intervals and test hypotheses possess sampling distributions that are the familiar t, χ^2, and F distributions of Chapters 7, 8, and 9.

TABLE 9.7a Summary of Hypothesis Tests: One-Sample Case

Parameter (θ)	Null Hypothesis (H_0)	Point Estimator ($\hat{\theta}$)	Test Statistic	Sample Size	Additional Assumptions
μ	$\mu = \mu_0$	$\bar{y}$	$z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}} \approx \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$	$n \geq 30$	None
			$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$ where t is based on $\nu = (n - 1)$ degrees of freedom	$n < 30$	Normal population
p	$p = p_0$	$\hat{p} = \frac{y}{n}$	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	n large enough so that $n\hat{p} \geq 4$ and $n\hat{q} \geq 4$	None
σ^2	$\sigma^2 = \sigma_0^2$	s^2	$\chi^2 = \frac{(n - 1)s^2}{\sigma_0^2}$ where χ^2 has a chi-square distribution with $\nu = (n - 1)$ degrees of freedom	All n	Normal population

TABLE 9.7b Summary of Hypothesis Tests: Two-Sample Case

Parameter (θ)	Null Hypothesis (H_0)	Point Estimator ($\hat{\theta}$)	Test Statistic	Sample Size	Additional Assumptions
$(\mu_1 - \mu_2)$ Independent samples	$(\mu_1 - \mu_2) = D_0$ (If we want to detect a difference between μ_1 and μ_2, then $D_0 = 0$.)	$(\bar{y} - \bar{y}_2)$	$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \approx \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$n_1 \geq 30$, $n_2 \geq 30$	None
			$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where t is based on $\nu = n_1 + n_2 - 2$ degrees of freedom and $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	Either $n_1 < 30$ or $n_2 < 30$ or both	Both populations normal with equal variances ($\sigma_1^2 = \sigma_2^2$) (For situations in which $\sigma_1^2 \neq \sigma_2^2$, see the modifications listed in the box on page 455.)

(*continued*)

TABLE 9.7b Summary of Hypothesis Tests: Two-Sample Case, continued

Parameter (θ)	Null Hypothesis (H_0)	Point Estimator ($\hat{\theta}$)	Test Statistic	Sample Size	Additional Assumptions
$\mu_d = (\mu_1 - \mu_2)$ Matched pairs	$\mu_d = D_0$ (If we want to detect a difference between μ_1 and μ_2, then $D_0 = 0$.)	$\bar{d} = \Sigma_{i=1}^{n} d_i/n$ Mean of sample differences	$t = \dfrac{\bar{d} - D_0}{s_d/\sqrt{n_d}}$ where t is based on $\nu = (n_d - 1)$ degrees of freedom	All n_d (If $n_d \geq 30$, then the standard normal (z) test may be used.)	Population of differences d_i is normal
$(p_1 - p_2)$	$(p_1 - p_2) = D_0$ (If we want to detect a difference between p_1 and p_2, then $D_0 = 0$.)	$(\hat{p}_1 - \hat{p}_2)$	For $D_0 = 0$: $z = \dfrac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ where $\hat{p} = \dfrac{y_1 + y_2}{n_1 + n_2}$ For $D_0 \neq 0$: $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\dfrac{\hat{p}_1\hat{q}_1}{n_1} + \dfrac{\hat{p}_2\hat{q}_2}{n_2}}}$	n_1 and n_2 large enough so that $n_1\hat{p}_1 \geq 4$, $n_1\hat{q}_1 \geq 4$ and $n_2\hat{p}_2 \geq 4$, $n_2\hat{q}_2 \geq 4$	Independent samples
$\dfrac{\sigma_1^2}{\sigma_2^2}$	$\dfrac{\sigma_1^2}{\sigma_2^2} = 1$ (i.e., $\sigma_1^2 = \sigma_2^2$)	$\dfrac{s_1^2}{s_2^2}$	For H_a: $\sigma_1^2 > \sigma_2^2$: $F = \dfrac{s_1^2}{s_2^2}$ For H_a: $\sigma_2^2 > \sigma_1^2$: $F = \dfrac{s_2^2}{s_1^2}$ For H_a: $\sigma_1^2 \neq \sigma_2^2$: $F = \dfrac{\text{Larger } s^2}{\text{Smaller } s^2}$ where the distribution of F is based on ν_1 = numerator degrees of freedom and ν_2 = denominator degrees of freedom	All n_1 and n_2	Independent random samples from normal populations

SUPPLEMENTARY EXERCISES

9.70 One of the keys to occupational therapy is patient motivation. A study was conducted to determine whether *purposeful activity* (defined as tasks that are goal-directed) provides intrinsic motivation to exercise performance (*Journal of Occupational Therapy*, Mar. 1984). Twenty-six females were recruited to take part in the study. Each female subject was instructed to perform two similar exercises, jumping rope (the purposeful activity) and jumping without a rope (the nonpurposeful activity), until their perceived exertion level reached

17 on the RPE scale (i.e., until they had worked their bodies "very hard"). The length of time (in minutes) that each subject jumped was then recorded for each of the two exercises and the difference d_i (computed by subtracting the length of jumping time without rope from the length of jumping time with rope) was calculated. A summary of the 26 differences is provided here:

$$\bar{d} = 41.84 \text{ seconds}$$
$$s_d = 110.28 \text{ seconds}$$

One theory held by occupational therapists is that those performing a purposeful activity are more motivated, and hence, tend to fatigue less easily. Test the hypothesis that the mean exercise time for the purposeful activity (jumping with a rope) exceeds the mean exercise time for the nonpurposeful activity (jumping without a rope). Use $\alpha = .05$.

9.71 Suppose you want to determine whether users of data processors have a preference between word processors A and B. If users have no preference for either of the two word processors (i.e., if the two systems are identical), then the probability p that a user prefers system A is $p = .5$. Let y be the number of users in a sample of 10 who prefer system A, and suppose you want to test H_0: $p = .5$ against H_a: $p \neq .5$. One possible test procedure is to reject H_0 if $y \leq 1$ or $y \geq 8$.

a. Find α for this test.
b. Find β if $p = .4$. What is the power of the test?
c. Find β if $p = .8$. What is the power of the test?

9.72 The quality control department of a paper company measures the brightness (a measure of reflectance) of finished paper on a periodic basis throughout the day. Two instruments that are available to measure the paper specimens are subject to error, but they can be adjusted so that the mean readings for a control paper specimen are the same for both instruments. Suppose you are concerned about the precision of the two instruments—namely, that instrument 2 is less precise than instrument 1. To check this theory, five measurements of a single paper sample are made on both instruments. The data are shown in the table. Do the data provide sufficient evidence to indicate that instrument 2 is less precise than instrument 1? Test using $\alpha = .05$.

Instrument 1	*Instrument 2*
29	26
28	34
30	30
28	32
30	28

9.73 The testing department of a tire and rubber company schedules truck and passenger tires for durability tests. Currently, tires are scheduled twice weekly on flexible processors (machines that can handle either truck or passenger tires) using the shortest processing time (SPT) approach. Under SPT, the tire with the shortest processing time is scheduled first. Company researchers have developed a new scheduling rule which they believe will reduce the average flow time (i.e., the average completion time of a test) and lead to a reduction in the average tardiness of a scheduled test. To compare the two scheduling rules, 64 tires were randomly selected and divided into two groups of equal size. One set of tires was scheduled using SPT, the other using the proposed rule. A summary of the flow times and tardiness (in hours) of the tire tests is provided in the table on page 486.

	Flow Time		Tardiness	
	Mean	*Variance*	*Mean*	*Variance*
SPT	158.28	8,532.80	5.26	452.09
Proposed Rule	117.07	5,208.53	4.52	319.41

a. Is there sufficient evidence at $\alpha = .05$ to conclude that the average flow time is less under the proposed scheduling rule than under the SPT approach?
b. Is there sufficient evidence at $\alpha = .05$ to conclude that the proposed scheduling rule will lead to a reduction in the average tardiness of tire tests?

9.74 Refer to the reinforced concrete T-beam cracking experiment described in Exercise 9.65. The experimental results were compared to the theoretical results obtained using the failure surface method of predicting ultimate load capacity. The actual and theoretical ultimate torsion moments for six T-beams with 40-cm slab widths are given in the table. Conduct a test to determine whether the experimental mean ultimate torsion moment differs from the theoretical mean ultimate torsion moment. Use $\alpha = .05$.

T-Beam	1	2	3	4	5	6
Experimental result	4.70	5.20	5.40	5.40	4.30	4.80
Theoretical result	4.63	4.65	5.60	5.60	3.62	3.62

Source: Zararis, P. D., and Penelis, C. Jr. "Reinforced concrete T-beams in torsion and bending." *Journal of the American Concrete Institute*, Vol. 83, No. 1, Jan.–Feb. 1986, p. 153.

9.75 A problem that occurs with certain types of mining is that some byproducts tend to be mildly radioactive and these products sometimes get into our freshwater supply. The EPA has issued regulations concerning a limit on the amount of radioactivity in supplies of drinking water. Particularly, the maximum level for naturally occurring radiation is 5 picocuries per liter of water. A random sample of 24 water specimens from a city's water supply produced the sample statistics $\bar{y} = 4.61$ picocuries per liter and $s = .87$ picocurie per liter.
a. Do these data provide sufficient evidence to indicate that the mean level of radiation is safe (below the maximum level set by the EPA)? Test using $\alpha = .01$.
b. Why should you want to use a small value of α for the test in part **a**?
c. Calculate the value of β for the test if $\mu_a = 4.5$ picocuries per liter of water.
d. Calculate and interpret the p-value for the test.

9.76 Usually, when trees grown in greenhouses are replanted in their natural habitat, there is only a 50% survival rate. However, a recent General Telephone and Electronics (GTE) advertisement claimed that trees grown in a particular environment ideal for plant growth have a 95% survival rate when replanted. These trees are grown inside a mountain in Idaho where the air temperature, carbon dioxide content, and humidity are all constant, and there are no major disease or insect problems. A key growth ingredient—light—is supplied by specially made GTE Sylvania Super-Metalarc lamps. These lights help the young trees develop a more fibrous root system that aids in the transplantation. Suppose that we want to challenge GTE's claim, i.e., we want to test whether the true proportion of all trees grown inside the Idaho mountain that survive when replanted in their natural habitat is less than .95. We randomly sample 50 of the trees grown in the controlled environment, replant the trees in their natural habitat, and observe that 46 of the trees survive. Perform the test at a level of significance of $\alpha = .01$.

9.77 A *parallel processor*, or *paracomputer*, consists of autonomous processing elements (PEs) sharing a central memory. Researchers at New York University have recently designed such a paracomputer, called the NYU

Ultracomputer. To assess the impact of network delay on overall ultracomputer performance, the researchers simulated central memory access time for sample instructions from a parallel version of a NASA weather program. Two sets of access times were simulated—one set processed with 16 processing elements, the other set with 48 processing elements. With 16 PEs, the average central memory access time was 8.94 seconds, whereas with 48 PEs the average central memory access time was 8.83 seconds. Assume that $n = 1{,}000$ instructions were simulated for each of the two programs, with standard deviations equal to 3.10 and 3.50, respectively. This information was not provided in the researchers' report. Is there sufficient evidence to indicate a difference between the average central memory access times of instructions processed with 16 and 48 PEs? Test using $\alpha = .05$.

9.78 In the manufacture of machinery, it is essential to utilize parts that conform to specifications. In the past, diameters of the ball bearings produced by a certain manufacturer had a variance of .00156. To cut costs, the manufacturer instituted a less expensive production method. The variance of the diameters of 100 randomly sampled bearings produced by the new process was .00211. Do the data provide sufficient evidence to indicate that diameters of ball bearings produced by the new process are more variable than those produced by the old process? Test at $\alpha = .05$.

9.79 The ion balance of our atmosphere has a significant effect on human health. A high concentration of positive ions in a room can induce fatigue, stress, and respiratory problems in the room's occupants. However, research has shown that introduction of additional negative ions into the room's atmosphere (through a negative ion generator), in combination with constant ventilation, restores the natural balance of ions that is conducive to human health. One experiment was conducted as follows. One hundred employees of a large factory were randomly selected and divided into two groups of 50 each. Both groups were told that they would be working in an atmosphere with an ion balance controlled through negative ion generators. However, unknown to the employees, the generators were switched on only in the experimental group's work area. At the end of the day, the number of employees reporting migraine, nausea, fatigue, faintness, or some other physical discomfort was recorded for each group. The results are summarized in the table.

	Experimental Group (Ion generators on)	*Control Group (Ion generators off)*
Number in Sample	$n_1 = 50$	$n_2 = 50$
Number in Sample Who Experience Some Type of Physical Discomfort	3	12

a. Perform a test of hypothesis to determine whether the proportion of employees in the experimental group who experience some type of physical discomfort at the end of the day is significantly less than the corresponding proportion for the control group. Use a significance level of $\alpha = .03$.

b. Compute the p-value for this test.

9.80 The use of computer equipment in business is growing at a phenomenal rate. A recent study revealed that 184 of 616 working adults now regularly use a personal computer, microcomputer, computer terminal, or word processor on the job (*Journal of Advertising Research*, Apr./May 1984). Is this sufficient evidence to indicate that the proportion of all working adults who regularly use computer equipment on the job exceeds 25%? Test using $\alpha = .05$.

9.81 The means and standard deviations shown in the table summarize information on the strengths (modules of rupture at ground line, in pounds per square inch) for two types of wooden poles used by the utility industry. Do the data provide sufficient evidence to indicate a difference in the variance of the strengths of wooden poles made from coastal Douglas fir and southern pine? Test using $\alpha = .02$.

Species	*Sample Size*	*Sample Mean*	*Sample Standard Deviation*
Coastal Douglas fir	118	8,380	644.62
Southern pine	147	8,870	611.72

Source: Goodman, J. R., Vanderbilt, M. D., and Criswell, M. E. "Reliability-based design of wood transmission line structures." *Journal of Structural Engineering*, Vol. 109, No. 3, 1983, pp. 690–704.

9.82 The accompanying table provides data on the theoretical (calculated) and experimental values of the vapor pressures for dibenzothiophene, a heterocycloaromatic compound similar to those found in coal tar. If the theoretical model for vapor pressure is a good model of reality, the true mean difference between the experimental and calculated values of vapor pressure for a given temperature will equal 0.

Temperature	*Vapor Pressure*		*Temperature*	*Vapor Pressure*	
(°C)	*Experimental*	*Calculated*	*(°C)*	*Experimental*	*Calculated*
100.60	.282	.276	116.69	.669	.695
101.36	.314	.307	119.38	.834	.805
104.60	.335	.350	121.08	.890	.882
106.44	.404	.390	123.61	1.01	1.01
108.70	.422	.444	124.90	1.07	1.08
110.96	.513	.505	127.74	1.26	1.25
112.62	.554	.554	130.24	1.42	1.43
115.21	.642	.640	131.75	1.55	1.54

Source: Edwards, D. R., and Prausnitz, J. M. "Vapor pressures of some sulphur-containing, coal-related compounds." *Journal of Chemical and Engineering Data*, Vol. 26, 1981, pp. 121–124. Copyright 1981 American Chemical Society. Reprinted with permission.

a. Do the data provide sufficient evidence to indicate that the mean difference differs from 0? Test using $\alpha = .05$.

b. Calculate and interpret the p-value for the test.

9.83 A machine is set to produce bolts with a mean length of 1 inch. Bolts that are too long or too short do not meet the customer's specifications and must be rejected. To avoid producing too many rejects, the bolts produced by the machine are sampled from time to time and tested as a check to determine whether the machine is still operating properly, i.e., producing bolts with a mean length of 1 inch. Suppose 50 bolts have been sampled, and $\bar{y} = 1.02$ inches and $s = .04$ inch. Does the sample evidence indicate that the machine is producing bolts with a mean length not equal to 1 inch; i.e., is the production process out of control? Test using $\alpha = .01$.

9.84 Heat stress in dairy cows can have a dramatic negative effect on milk production. High temperatures tend to reduce a cow's food intake, which in turn reduces milk yield. Researchers in the IFAS Dairy Research Unit and the Department of Agricultural Engineering at the University of Florida have developed design criteria for the construction of shade structures that they believe will help alleviate heat stress for dairy cows. In one experiment, 31 Holstein cows in the last trimester of pregnancy were divided into two groups. Sixteen cows were given access to a shade structure and the remaining 15 cows were denied shade. Researchers recorded the 100-day milk yield (in pounds) of each cow after calving. The mean milk yields of the two groups are shown in the accompanying table. Is there sufficient evidence to indicate a difference between the mean milk yields of cows given access to shade and cows denied shade? Use $\alpha = .10$. (Assume the standard deviations of milk yields are equal to 40 pounds for both groups.)

	Shade	*No Shade*
Sample Size	16	15
Mean	367.4	330.8

Source: "Minimizing heat stress for dairy cows." *Florida Agricultural Research* 83, Vol. 2, No. 1, Winter 1983, pp. 10–13.

COMPUTER LAB: Testing Means

In this section, we present the computer commands for conducting tests of hypotheses concerning population means. Both packages, SAS and MINITAB, can perform t tests about μ, $(\mu_1 - \mu_2)$ for independent samples, and $(\mu_1 - \mu_2)$ for paired samples. (Remember, for large samples, the t and z statistics are nearly equivalent.) Tests about variances and proportions are not available in SAS or MINITAB.

SAS

a. One-Sample Test—Test H_0: $\mu = 8.5$ in Example 9.8

Command line

```
1   DATA BONES;              }
2   INPUT RATIO @@;          } Data entry instructions
3   TESTRAT=RATIO-8.5;       }
4   CARDS;
    10.73 8.48  8.52         }
      :     :     :          } Input data values
      :     :     :          } (3 observations per line)
     9.93 8.17 12.00         }
5   PROC MEANS T PRT;        }
6   VAR TESTRAT;             } Student's t test
```

COMMAND 3 The transformed variable TESTRAT is computed by subtracting the hypothesized mean ($\mu = 8.5$) from each value of RATIO.

COMMANDS 5–6 The PROC MEANS statement commands SAS to conduct a t test on the values of the variable TESTRAT (specified in line 6). SAS will test the null hypothesis H_0: $\mu_{\text{TESTRAT}} = 0$, which is equivalent to testing H_0: $\mu_{\text{RATIO}} = 8.5$.

OUTPUT The p-value reported in SAS is a *two-tailed* observed significance level. Divide this reported value in half to obtain the p-value for a one-tailed test. [*Note:* The SAS output for this program is displayed in Figure 9.23a.]

b. Two-Sample Test, Independent Samples —Test H_0: $\mu_1 - \mu_2 = 0$ in Example 9.14

Command line

```
1   DATA DISKS;                     } Data entry instructions
2   INPUT DRIVE TIME @@;            }
3   CARDS;                          }
    1 59 1 73 1 74 1 61             ] Input data values
    : : : : : : : :                 ] (4 observations per line)
    : : : : : : : :                 ]
    2 86 2 53 2 68 2 39             ]
4   PROC TTEST;                     } Student's t test
5   CLASS DRIVE; VAR TIME;          }
```

COMMAND 2 TIME is the variable of interest. DRIVE is a grouping variable that takes on two values (e.g., 1 and 2).

COMMANDS 4–5 The TTEST procedure conducts a *t* test on the difference in means of the variable TIME for the two groups identified by DRIVE.

OUTPUT SAS calculates the *t* value for both the equal population variances case and the unequal variances case. [*Note:* The SAS output for this program is displayed in Figure 9.23b.]

c. Two-Sample Test, Paired Samples—Test H_0: $\mu_d = 0$ in Example 9.15

Command line

```
1   DATA CLOUD;                     } Data entry instructions
2   INPUT SEED UNSEED;              }
3   DIFF=SEED-UNSEED                }
4   CARDS;
    1.75 1.62                       ] Input data values
      :    :                        ] (1 observation per line)
      :    :                        ]
    2.42 2.33                       ]
5   PROC MEANS T PRT;               } Student's t test
6   VAR DIFF;                       }
```

COMMANDS 2–3 The variables SEED and UNSEED contain the measurements for each member of the matched pair. The difference, DIFF, is computed in line 3.

OUTPUT [*Note:* The SAS output for this program is displayed in Figure 9.23c.]

FIGURE 9.23 ▶
SAS output for computer Lab

a.

```
Analysis Variable : TESTRAT

N Obs           T    Prob>|T|
-----------------------------
   41   4.0303238      0.0002
-----------------------------
```

b.

TTEST PROCEDURE

Variable: TIME

DRIVE	N	Mean	Std Dev	Std Error	Minimum	Maximum
1	13	68.23076923	18.65991178	5.17532836	33.00000000	102.0000000
2	15	53.80000000	15.80777386	4.08154966	34.00000000	86.0000000

Variances	T	DF	Prob>|T|
Unequal	2.1894	23.7	0.0387
Equal	2.2163	26.0	0.0356

For H0: Variances are equal, F' = 1.39 DF = (12,14) Prob>F' = 0.5482

c.

Analysis Variable : DIFF

N Obs	T	Prob>|T|
6	3.0066442	0.0299

MINITAB

a. One-Sample t Test—Test H_0: $\mu = 8.5$ in Example 9.8

Command line		
1	SET RATIOS IN C1	Data entry instruction
2	NAME C1='RATIO'	
	10.73 8.48 8.52 ⋮ ⋮ ⋮ 9.93 8.17 12.00	Input data values (3 observations per line)
3	TTEST OF MU=50 ON C1;	Student's t test
4	ALTERNATIVE=+1.	

COMMANDS 3–4 The TTEST procedure performs a t test on the difference between the mean of the variable read in C1 and the hypothesized value specified in the MU= subcommand (line 3). The subcommand ALTERNATIVE=+1 (line 4) requests that a one-tailed upper-tailed test be performed. Use ALTERNATIVE=−1 for a lower-tailed test. If the subcommand is not used, a two-tailed test is performed.

OUTPUT [*Note:* The MINITAB output for this program is displayed in Figure 9.24a on page 493.]

b. Two-Sample Test, Independent Samples—Test H_0: $\mu_1 - \mu_2 = 0$ in Example 9.14

Command line		
1	`SET DISK1 DATA IN C1`	Data entry instruction

	`59 73 74 61`
	`.  .  .  .`
	`.  .  .  .`
	`.  .  .  .`
		Input data values (4 observations per line)

2	`SET DISK2 DATA IN C2`	

	`71 63 40 34`
	`.  .  .  .`
	`.  .  .  .`
	`.  .  .  .`
		Input data values (4 observations per line)

3	`TWOSAMPLE T C1 C2;`	Student's t test
4	`  POOLED.`	

COMMANDS 3–4 TWOSAMPLE performs a t test on the difference between the means of the data in C1 and C2. The subcommand POOLED (line 4) requests that a pooled sample variance be used. (This is appropriate when the population variances are equal.) If you want MINITAB to adjust the t statistic and degrees of freedom for the unequal variances case, omit the POOLED subcommand.

GENERAL Use the ALTERNATIVE subcommand to obtain a one-tailed test.

OUTPUT [*Note:* The MINITAB output for this program is displayed in Figure 9.24b.]

c. Two-Sample Test, Paired Samples—Test H_0: $\mu_d = 0$ in Example 9.15

Command line		
1	`READ DATA IN C1 C2`	Data entry instruction

	`1.75 1.62`
	`  .    .`
	`  .    .`
	`2.42 2.33`
		Input data values (1 observation per line)

2	`SUBTRACT C2 FROM C1, PUT IN C3`	
3	`NAME C3='DIFF'`	
4	`TTEST OF MU=0 ON DATA IN C3`	Student's t test

COMMANDS 1–2 The data in columns C1 and C2 are the measurements for each member of the matched pair. C3 contains the difference between the measurements.

COMMAND 4 TTEST performs a t test on the mean of the differences in C3.

GENERAL Use the ALTERNATIVE subcommand to obtain a one-tailed test.

OUTPUT [*Note:* The MINITAB output for this program is displayed in Figure 9.24c.]

FIGURE 9.24 ▶ MINITAB output for Computer Lab

a.

```
TEST OF MU = 8.500 VS MU G.T. 8.500

                N      MEAN     STDEV   SE MEAN        T     P VALUE
RATIO          41     9.258     1.204     0.188     4.03      0.0001
```

b.

```
TWOSAMPLE T FOR disk1 VS disk2
        N       MEAN       STDEV    SE MEAN
disk1  13       68.2        18.7        5.2
disk2  15       53.8        15.8        4.1

95 PCT CI FOR MU disk1 - MU disk2: (1.0, 27.8)

TTEST MU disk1 = MU disk2 (VS NE): T= 2.22  P=0.036  DF=  26

POOLED STDEV =        17.2
```

c.

```
TEST OF MU = 0.0000 VS MU N.E. 0.0000

                N      MEAN     STDEV   SE MEAN        T     P VALUE
DIFF            6    0.1633    0.1331    0.0543     3.01       0.030
```

References

Freedman, D., Pisani, R., and Purves, R., *Statistics*. New York: W. W. Norton and Co., 1978.

Hoel, P. G. *Introduction to Mathematical Statistics*, 6th ed. New York: Wiley, 1987.

Hogg, R. V., and Craig, A. T. *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan, 1978.

McClave, J. T., and Dietrich, F. H. II. *Statistics*, 6th ed. San Francisco: Dellen, 1994.

Mendenhall, W. *Introduction to Probability and Statistics*, 8th ed. Boston: Duxbury, 1990.

Mendenhall, W., Wackerly, D. D., and Scheaffer, R. L. *Mathematical Statistics with Applications*, 3rd ed. Boston: Duxbury, 1989.

Mood, A. M., Graybill, F. A., and Boes, D. *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill, 1974.

Snedecor, G. W., and Cochran, W. G. *Statistical Methods*, 7th ed. Ames, Iowa: Iowa State University Press, 1980.