A Review of Basic Concepts (Optional)

Chapter 1

Contents

Objectives

1. Review some basic concepts of sampling.
2. Review methods for describing both qualitative and quantitative data.
3. Review inferential statistical methods: confidence intervals and hypothesis tests.

Although we assume students have had a prerequisite introductory course in statistics, courses vary somewhat in content and in the manner in which they present statistical concepts. To be certain that we are starting with a common background, we use this chapter to review some basic definitions and concepts. Coverage is optional.

1.1 Statistics and Data

According to *The Random House College Dictionary* (2001 ed.), 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**—a science that will enable you to be proficient data producers and efficient data users.

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

Data are obtained by measuring some characteristic or property of the objects (usually people or things) of interest to us. These objects upon which the measurements (or observations) are made are called **experimental units**, and the properties being measured are called **variables** (since, in virtually all studies of interest, the property varies from one observation to another).

Definition 1.2 An **experimental unit** is an object (person or thing) upon which we collect data.

Definition 1.3 A **variable** is a characteristic (property) of the experimental unit with outcomes (data) that vary from one observation to the next.

All data (and consequently, the variables we measure) are either **quantitative** or **qualitative** in nature. Quantitative data are data that can be measured on a naturally occurring numerical scale. In general, qualitative data take values that are nonnumerical; they can only be classified into categories. The statistical tools that we use to analyze data depend on whether the data are quantitative or qualitative. Thus, it is important to be able to distinguish between the two types of data.

Definition 1.4 **Quantitative data** are observations measured on a naturally occurring numerical scale.

Definition 1.5 Nonnumerical data that can only be classified into one of a group of categories are said to be **qualitative data**.

Example 1.1

Chemical and manufacturing plants often discharge toxic waste materials such as DDT into nearby rivers and streams. These toxins can adversely affect the plants and animals inhabiting the river and the riverbank. The U.S. Army Corps of Engineers conducted a study of fish in the Tennessee River (in Alabama) and its three tributary creeks: Flint Creek, Limestone Creek, and Spring Creek. A total of 144 fish were captured, and the following variables were measured for each:

1. River/creek where each fish was captured
2. Number of miles upstream where the fish was captured
3. Species (channel catfish, largemouth bass, or smallmouth buffalofish)
4. Length (centimeters)
5. Weight (grams)
6. DDT concentration (parts per million)

The data are saved in the FISHDDT file. Data for 10 of the 144 captured fish are shown in Table 1.1.

(a) Identify the experimental units.
(b) Classify each of the five variables measured as quantitative or qualitative.

Solution

(a) Because the measurements are made for each fish captured in the Tennessee River and its tributaries, the experimental units are the 144 captured fish.
(b) The variables upstream that capture location, length, weight, and DDT concentration are quantitative because each is measured on a natural numerical scale: upstream in miles from the mouth of the river, length in centimeters, weight in grams, and DDT in parts per million. In contrast, river/creek and species cannot be measured quantitatively; they can only be classified into categories (e.g., channel catfish, largemouth bass, and smallmouth buffalofish for species). Consequently, data on river/creek and species are qualitative. ■

FISHDDT

Table 1.1 Data collected by U.S. Army Corps of Engineers (selected observations)

River/Creek	Upstream	Species	Length	Weight	DDT
FLINT	5	CHANNELCATFISH	42.5	732	10.00
FLINT	5	CHANNELCATFISH	44.0	795	16.00
SPRING	1	CHANNELCATFISH	44.5	1133	2.60
TENNESSEE	275	CHANNELCATFISH	48.0	986	8.40
TENNESSEE	275	CHANNELCATFISH	45.0	1023	15.00
TENNESSEE	280	SMALLMOUTHBUFF	49.0	1763	4.50
TENNESSEE	280	SMALLMOUTHBUFF	46.0	1459	4.20
TENNESSEE	285	LARGEMOUTHBASS	25.0	544	0.11
TENNESSEE	285	LARGEMOUTHBASS	23.0	393	0.22
TENNESSEE	285	LARGEMOUTHBASS	28.0	733	0.80

1.1 Exercises

1.1 College application data. Colleges and universities are requiring an increasing amount of information about applicants before making acceptance and financial aid decisions. Classify each of the following types of data required on a college application as quantitative or qualitative.

(a) High school GPA
(b) Country of citizenship
(c) Applicant's score on the SAT or ACT
(d) Gender of applicant
(e) Parents' income
(f) Age of applicant

1.2 Fuel Economy Guide. The data in the accompanying table were obtained from the *Model Year 2009 Fuel Economy Guide* for new automobiles.

(a) Identify the experimental units.
(b) State whether each of the variables measured is quantitative or qualitative.

1.3 Ground motion of earthquakes. In the *Journal of Earthquake Engineering* (November 2004), a team of civil and environmental engineers studied the ground motion characteristics of 15 earthquakes that occurred around the world between 1940 and 1995. Three (of many) variables measured on each earthquake were the type of ground motion (short, long, or forward directive), earthquake magnitude (Richter scale), and peak ground acceleration (feet per second). One of the goals of the study was to estimate the inelastic spectra of any ground motion cycle.

(a) Identify the experimental units for this study.
(b) Identify the variables measured as quantitative or qualitative.

1.4 Use of herbal medicines. *The American Association of Nurse Anesthetists Journal* (February 2000) published the results of a study on the use of herbal medicines before surgery. Each of 500

MODEL NAME	MFG	TRANSMISSION TYPE	ENGINE SIZE (LITERS)	NUMBER OF CYLINDERS	EST. CITY MILEAGE (MPG)	EST. HIGHWAY MILEAGE (MPG)
TSX	Acura	Automatic	2.4	4	21	30
Jetta	VW	Automatic	2.0	4	29	40
528i	BMW	Manual	3.0	6	18	28
Fusion	Ford	Automatic	3.0	6	17	25
Camry	Toyota	Manual	2.4	4	21	31
Escalade	Cadillac	Automatic	6.2	8	12	19

Source: Model Year 2009 Fuel Economy Guide, U.S. Dept. of Energy, U.S. Environmental Protection Agency (www.fueleconomy.gov).

surgical patients was asked whether they used herbal or alternative medicines (e.g., garlic, ginkgo, kava, fish oil) against their doctor's advice before surgery. Surprisingly, 51% answered "yes."

(a) Identify the experimental unit for the study.
(b) Identify the variable measured for each experimental unit.
(c) Is the data collected quantitative or qualitative?

1.5 **Drinking-water quality study.** *Disasters* (Vol. 28, 2004) published a study of the effects of a tropical cyclone on the quality of drinking water on a remote Pacific island. Water samples (size 500 milliliters) were collected approximately 4 weeks after Cyclone Ami hit the island. The following variables were recorded for each water sample. Identify each variable as quantitative or qualitative.

(a) Town where sample was collected
(b) Type of water supply (river intake, stream, or borehole)
(c) Acidic level (pH scale, 1–14)
(d) Turbidity level (nephalometric turbidity units [NTUs])
(e) Temperature (degrees Centigrade)
(f) Number of fecal coliforms per 100 milliliters
(g) Free chlorine-residual (milligrams per liter)
(h) Presence of hydrogen sulphide (yes or no)

1.6 **Accounting and Machiavellianism.** *Behavioral Research in Accounting* (January 2008) published a study of Machiavellian traits in accountants. *Machiavellian* describes negative character traits that include manipulation, cunning, duplicity, deception, and bad faith. A questionnaire was administered to a random sample of 700 accounting alumni of a large southwestern university. Several variables were measured, including age, gender, level of education, income, job satisfaction score, and Machiavellian ("Mach") rating score. What type of data (quantitative or qualitative) is produced by each of the variables measured?

1.2 Populations, Samples, and Random Sampling

When you examine a data set in the course of your study, you will be doing so because the data characterize a group of experimental units of interest to you. In statistics, the data set that is collected for all experimental units of interest is called a **population**. This data set, which is typically large, either exists in fact or is part of an ongoing operation and hence is conceptual. Some examples of statistical populations are given in Table 1.2.

Definition 1.6 A **population data set** is a collection (or set) of data measured on all experimental units of interest to you.

Many populations are too large to measure (because of time and cost); others cannot be measured because they are partly conceptual, such as the set of quality

Table 1.2 Some typical populations

Variable	Experimental Units	Population Data Set	Type
a. Starting salary of a graduating Ph.D. biologist	All Ph.D. biologists graduating this year	Set of starting salaries of all Ph.D. biologists who graduated this year	Existing
b. Breaking strength of water pipe in Philadelphia	All water pipe sections in Philadelphia	Set of breakage rates for all water pipe sections in Philadelphia	Existing
c. Quality of an item produced on an assembly line	All manufactured items	Set of quality measurements for all items manufactured over the recent past and in the future	Part existing, part conceptual
d. Sanitation inspection level of a cruise ship	All cruise ships	Set of sanitation inspection levels for all cruise ships	Existing

measurements (population c in Table 1.2). Thus, we are often required to select a subset of values from a population and to make **inferences** about the population based on information contained in a **sample**. This is one of the major objectives of modern statistics.

> **Definition 1.7** A **sample** is a subset of data selected from a population.

> **Definition 1.8** A **statistical inference** is an estimate, prediction, or some other generalization about a population based on information contained in a sample.

Example 1.2

According to the research firm Magnum Global (2008), the average age of viewers of the major networks' television news programming is 50 years. Suppose a cable network executive hypothesizes that the average age of cable TV news viewers is less than 50. To test her hypothesis, she samples 500 cable TV news viewers and determines the age of each.

(a) Describe the population.
(b) Describe the variable of interest.
(c) Describe the sample.
(d) Describe the inference.

Solution

(a) The population is the set of units of interest to the cable executive, which is the set of all cable TV news viewers.
(b) The age (in years) of each viewer is the variable of interest.
(c) The sample must be a subset of the population. In this case, it is the 500 cable TV viewers selected by the executive.
(d) The inference of interest involves the *generalization* of the information contained in the sample of 500 viewers to the population of all cable news viewers. In particular, the executive wants to estimate the average age of the viewers in order to determine whether it is less than 50 years. She might accomplish this by calculating the average age in the sample and using the sample average to estimate the population average. ■

Whenever we make an inference about a population using sample information, we introduce an element of uncertainty into our inference. Consequently, it is important to report the **reliability** of each inference we make. Typically, this is accomplished by using a probability statement that gives us a high level of confidence that the inference is true. In Example 1.2, we could support the inference about the average age of all cable TV news viewers by stating that the population average falls within 2 years of the calculated sample average with "95% confidence." (Throughout the text, we demonstrate how to obtain this measure of reliability—and its meaning—for each inference we make.)

> **Definition 1.9** A **measure of reliability** is a statement (usually quantified with a probability value) about the degree of uncertainty associated with a statistical inference.

The level of confidence we have in our inference, however, will depend on how **representative** our sample is of the population. Consequently, the sampling procedure plays an important role in statistical inference.

> **Definition 1.10** A **representative sample** exhibits characteristics typical of those possessed by the population.

The most common type of sampling procedure is one that gives every different sample of fixed size in the population an equal probability (chance) of selection. Such a sample—called a **random sample**—is likely to be representative of the population.

> **Definition 1.11** A **random sample** of n experimental units is one selected from the population in such a way that every different sample of size n has an equal probability (chance) of selection.

How can a random sample be generated? If the population is not too large, each observation may be recorded on a piece of paper and placed in a suitable container. After the collection of papers is thoroughly mixed, the researcher can remove n pieces of paper from the container; the elements named on these n pieces of paper are the ones to be included in the sample. Lottery officials utilize such a technique in generating the winning numbers for Florida's weekly 6/52 Lotto game. Fifty-two white ping-pong balls (the population), each identified from 1 to 52 in black numerals, are placed into a clear plastic drum and mixed by blowing air into the container. The ping-pong balls bounce at random until a total of six balls "pop" into a tube attached to the drum. The numbers on the six balls (the random sample) are the winning Lotto numbers.

This method of random sampling is fairly easy to implement if the population is relatively small. It is not feasible, however, when the population consists of a large number of observations. Since it is also very difficult to achieve a thorough mixing, the procedure only approximates random sampling. Most scientific studies, however, rely on computer software (with built-in random-number generators) to automatically generate the random sample. Almost all of the popular statistical software packages available (e.g., SAS, SPSS, MINITAB) have procedures for generating random samples.

1.2 Exercises

1.7 Guilt in decision making. The effect of guilt emotion on how a decision-maker focuses on the problem was investigated in the *Journal of Behavioral Decision Making* (January 2007). A total of 155 volunteer students participated in the experiment, where each was randomly assigned to one of three emotional states (guilt, anger, or neutral) through a reading/writing task. Immediately after the task, the students were presented with a decision problem (e.g., whether or not to spend money on repairing a very old car). The researchers found that a higher proportion of students in the guilty-state group chose not to repair the car than those in the neutral-state and anger-state groups.

(a) Identify the population, sample, and variables measured for this study.

(b) What inference was made by the researcher?

1.8 Use of herbal medicines. Refer to the *American Association of Nurse Anesthetists Journal* (February 2000) study on the use of herbal medicines before surgery, Exercise 1.4 (p. 3). The 500 surgical patients that participated in the study were randomly selected from surgical patients at several metropolitan hospitals across the country.

(a) Do the 500 surgical patients represent a population or a sample? Explain.
(b) If your answer was sample in part a, is the sample likely to be representative of the population? If you answered population in part a, explain how to obtain a representative sample from the population.

1.9 Massage therapy for athletes. Does a massage enable the muscles of tired athletes to recover from exertion faster than usual? To answer this question, researchers recruited eight amateur boxers to participate in an experiment (*British Journal of Sports Medicine*, April 2000). After a 10-minute workout in which each boxer threw 400 punches, half the boxers were given a 20-minute massage and half just rested for 20 minutes. Before returning to the ring for a second workout, the heart rate (beats per minute) and blood lactate level (micromoles) were recorded for each boxer. The researchers found no difference in the means of the two groups of boxers for either variable.

(a) Identify the experimental units of the study.
(b) Identify the variables measured and their type (quantitative or qualitative).
(c) What is the inference drawn from the analysis?
(d) Comment on whether this inference can be made about all athletes.

1.10 Gallup Youth Poll. A Gallup Youth Poll was conducted to determine the topics that teenagers most want to discuss with their parents. The findings show that 46% would like more discussion about the family's financial situation, 37% would like to talk about school, and 30% would like to talk about religion. The survey was based on a national sampling of 505 teenagers, selected at random from all U.S. teenagers.

(a) Describe the sample.
(b) Describe the population from which the sample was selected.
(c) Is the sample representative of the population?
(d) What is the variable of interest?
(e) How is the inference expressed?
(f) Newspaper accounts of most polls usually give a *margin of error* (e.g., plus or minus 3%) for the survey result. What is the purpose of the margin of error and what is its interpretation?

1.11 Insomnia and education. Is insomnia related to education status? Researchers at the Universities of Memphis, Alabama at Birmingham, and Tennessee investigated this question in the *Journal of Abnormal Psychology* (February 2005). Adults living in Tennessee were selected to participate in the study using a random-digit telephone dialing procedure. Two of the many variables measured for each of the 575 study participants were number of years of education and insomnia status (normal sleeper or chronic insomnia). The researchers discovered that the fewer the years of education, the more likely the person was to have chronic insomnia.

(a) Identify the population and sample of interest to the researchers.
(b) Describe the variables measured in the study as quantitative or qualitative.
(c) What inference did the researchers make?

1.12 Accounting and Machiavellianism. Refer to the *Behavioral Research in Accounting* (January 2008) study of Machiavellian traits in accountants, Exercise 1.6 (p. 6). Recall that a questionnaire was administered to a random sample of 700 accounting alumni of a large southwestern university; however, due to nonresponse and incomplete answers, only 198 questionnaires could be analyzed. Based on this information, the researchers concluded that Machiavellian behavior is not required to achieve success in the accounting profession.

(a) What is the population of interest to the researcher?
(b) Identify the sample.
(c) What inference was made by the researcher?
(d) How might the nonresponses impact the inference?

1.3 Describing Qualitative Data

Consider a study of aphasia published in the *Journal of Communication Disorders* (March 1995). Aphasia is the "impairment or loss of the faculty of using or understanding spoken or written language." Three types of aphasia have been identified by researchers: Broca's, conduction, and anomic. They wanted to determine whether one type of aphasia occurs more often than any other, and, if so, how often. Consequently, they measured aphasia type for a sample of 22 adult aphasiacs. Table 1.3 gives the type of aphasia diagnosed for each aphasiac in the sample.

APHASIA

Table 1.3 Data on 22 adult aphasiacs

Subject	Type of Aphasia
1	Broca's
2	Anomic
3	Anomic
4	Conduction
5	Broca's
6	Conduction
7	Conduction
8	Anomic
9	Conduction
10	Anomic
11	Conduction
12	Broca's
13	Anomic
14	Broca's
15	Anomic
16	Anomic
17	Anomic
18	Conduction
19	Broca's
20	Anomic
21	Conduction
22	Anomic

Source: Reprinted from *Journal of Communication Disorders,* Mar. 1995, Vol. 28, No. 1, E. C. Li, S. E. Williams, and R. D. Volpe, "The effects of topic and listener familiarity of discourse variables in procedural and narrative discourse tasks," p. 44 (Table 1) Copyright © 1995, with permission from Elsevier.

For this study, the variable of interest, aphasia type, is qualitative in nature. Qualitative data are nonnumerical in nature; thus, the value of a qualitative variable can only be classified into categories called *classes*. The possible aphasia types—Broca's, conduction, and anomic—represent the classes for this qualitative variable. We can summarize such data numerically in two ways: (1) by computing the *class frequency*—the number of observations in the data set that fall into each class; or (2) by computing the *class relative frequency*—the proportion of the total number of observations falling into each class.

Definition 1.12 A **class** is one of the categories into which qualitative data can be classified.

Definition 1.13 The **class frequency** is the number of observations in the data set falling in a particular class.

Definition 1.14 The **class relative frequency** is the class frequency divided by the total number of observations in the data set, i.e.,

$$\text{class relative frequency} = \frac{\text{class frequency}}{n}$$

Examining Table 1.3, we observe that 5 aphasiacs in the study were diagnosed as suffering from Broca's aphasia, 7 from conduction aphasia, and 10 from anomic aphasia. These numbers—5, 7, and 10—represent the class frequencies for the three classes and are shown in the summary table, Table 1.4.

Table 1.4 also gives the relative frequency of each of the three aphasia classes. From Definition 1.14, we know that we calculate the relative frequency by dividing the class frequency by the total number of observations in the data set. Thus, the relative frequencies for the three types of aphasia are

$$\text{Broca's: } \frac{5}{22} = .227$$

$$\text{Conduction: } \frac{7}{22} = .318$$

$$\text{Anomic: } \frac{10}{22} = .455$$

From these relative frequencies we observe that nearly half (45.5%) of the 22 subjects in the study are suffering from anomic aphasia.

Although the summary table in Table 1.4 adequately describes the data in Table 1.3, we often want a graphical presentation as well. Figures 1.1 and 1.2 show two of the most widely used graphical methods for describing qualitative data—bar graphs and pie charts. Figure 1.1 shows the frequencies of aphasia types in a **bar graph** produced with SAS. Note that the height of the rectangle, or "bar," over each class is equal to the class frequency. (Optionally, the bar heights can be proportional to class relative frequencies.)

Table 1.4 Summary table for data on 22 adult aphasiacs

Class (Type of Aphasia)	Frequency (Number of Subjects)	Relative Frequency (Proportion)
Broca's	5	.227
Conduction	7	.318
Anomic	10	.455
Totals	22	1.000

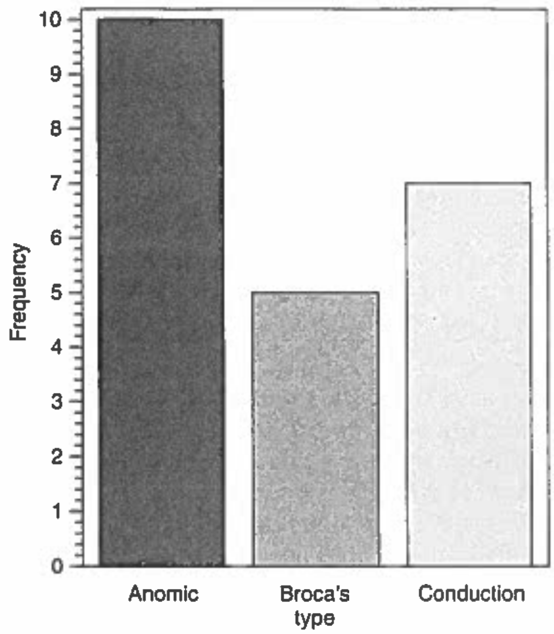


Figure 1.1 SAS bar graph for data on 22 aphasiacs

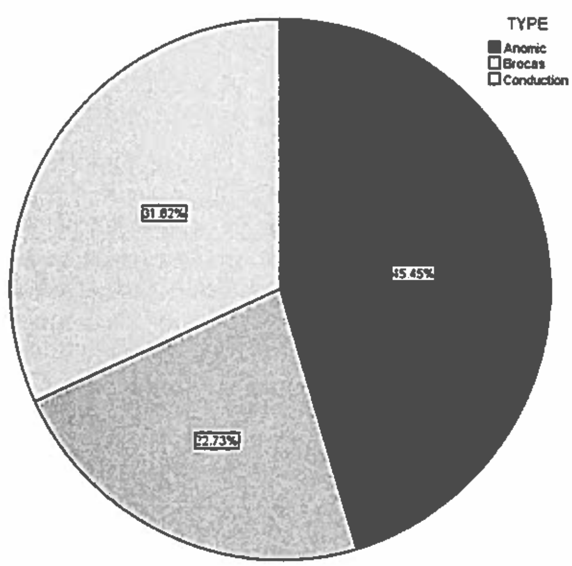


Figure 1.2 SPSS pie chart for data on 22 aphasiacs

In contrast, Figure 1.2 shows the relative frequencies of the three types of aphasia in a **pie chart** generated with SPSS. Note that the pie is a circle (spanning 360°) and the size (angle) of the "pie slice" assigned to each class is proportional to the class relative frequency. For example, the slice assigned to anomic aphasia is 45.5% of 360°, or (.455)(360°) = 163.8°.

1.3 Exercises

1.13 Estimating the rhino population. The International Rhino Federation estimates that there are 17,800 rhinoceroses living in the wild in Africa and Asia. A breakdown of the number of rhinos of each species is reported in the accompanying table.

RHINO SPECIES	POPULATION ESTIMATE
African Black	3,610
African White	11,330
(Asian) Sumatran	300
(Asian) Javan	60
(Asian) Indian	2,500
Total	17,800

Source: International Rhino Federation, March 2007.

(a) Construct a relative frequency table for the data.
(b) Display the relative frequencies in a bar graph.
(c) What proportion of the 17,800 rhinos are African rhinos? Asian?

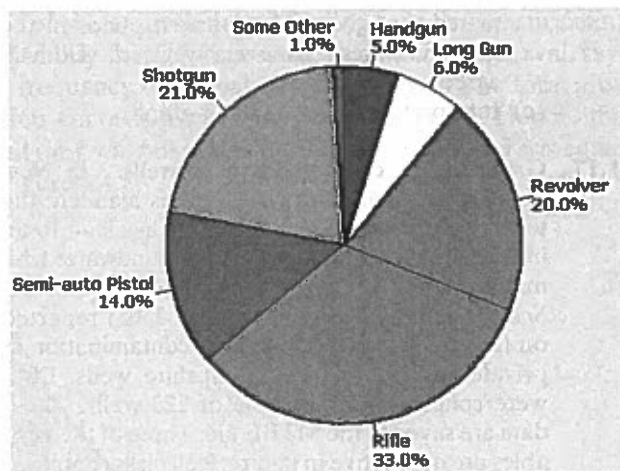
1.14 Blogs for Fortune 500 firms. Website communication through blogs and forums is becoming a key marketing tool for companies. The *Journal of Relationship Marketing* (Vol. 7, 2008) investigated the prevalence of blogs and forums at Fortune 500 firms with both English and Chinese websites. Of the firms that provided blogs/forums as a marketing tool, the accompanying table gives a breakdown on the entity responsible for creating the blogs/forums. Use a graphical method to describe the data summarized in the table. Interpret the graph.

BLOG/FORUM	PERCENTAGE OF FIRMS
Created by company	38.5
Created by employees	34.6
Created by third party	11.5
Creator not identified	15.4

Source: "Relationship Marketing in Fortune 500 U.S. and Chinese Web Sites," Karen E. Mishra and Li Cong, *Journal of Relationship Marketing,* Vol. 7, No. 1, 2008, reprinted by permission of the publisher (Taylor and Francis, Inc.)

1.15 National Firearms Survey. In the journal *Injury Prevention* (January 2007), researchers from the Harvard School of Public Health reported on the size and composition of privately held firearm stock in the United States. In a representative household telephone survey of 2,770 adults, 26% reported that they own at least one gun. The accompanying graphic summarizes the types of firearms owned.

(a) What type of graph is shown?
(b) Identify the qualitative variable described in the graph.
(c) From the graph, identify the most common type of firearms.



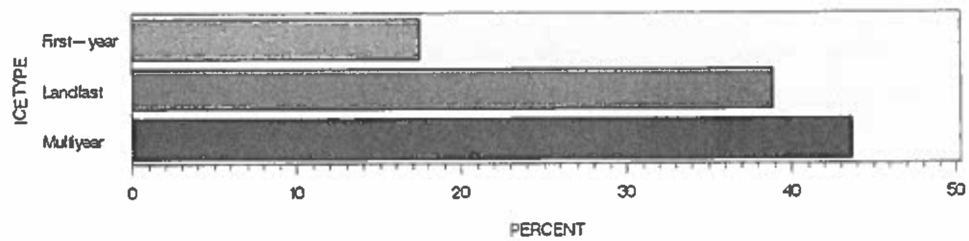
PONDICE

1.16 Characteristics of ice melt ponds. The National Snow and Ice Data Center (NSIDC) collects data on the albedo, depth, and physical characteristics of ice melt ponds in the Canadian arctic. Environmental engineers at the University of Colorado are using these data to study how climate impacts the sea ice. Data for 504 ice melt ponds located in the Barrow Strait in the Canadian arctic are saved in the PONDICE file. One variable of interest is the type of ice observed for each pond. Ice type is classified as first-year ice, multiyear ice, or landfast ice. A SAS summary table and horizontal bar graph that describe the ice types of the 504 melt ponds are shown at the top of the next page.

(a) Of the 504 melt ponds, what proportion had landfast ice?

The FREQ Procedure

ICETYPE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
First-year	88	17.46	88	17.46
Landfast	196	38.89	284	56.35
Multiyear	220	43.65	504	100.00



(b) The University of Colorado researchers estimated that about 17% of melt ponds in the Canadian arctic have first-year ice. Do you agree?

(c) Interpret the horizontal bar graph.

1.17 Groundwater contamination in wells. In New Hampshire, about half the counties mandate the use of reformulated gasoline. This has lead to an increase in the contamination of groundwater with methyl *tert*-butyl ether (MTBE). *Environmental Science and Technology* (January 2005) reported on the factors related to MTBE contamination in private and public New Hampshire wells. Data were collected for a sample of 223 wells. These data are saved in the MTBE file. Three of the variables are qualitative in nature: well class (public or private), aquifer (bedrock or unconsolidated), and detectible level of MTBE (below limit or detect). [Note: A detectible level of MTBE occurs if the MTBE value exceeds .2 micrograms per liter.] The data for 10 selected wells are shown in the accompanying table.

(a) Apply a graphical method to all 223 wells to describe the well class distribution.

(b) Apply a graphical method to all 223 wells to describe the aquifer distribution.

(c) Apply a graphical method to all 223 wells to describe the detectible level of MTBE distribution.

(d) Use two bar charts, placed side by side, to compare the proportions of contaminated wells for private and public well classes. What do you infer?

MTBE (selected observations)

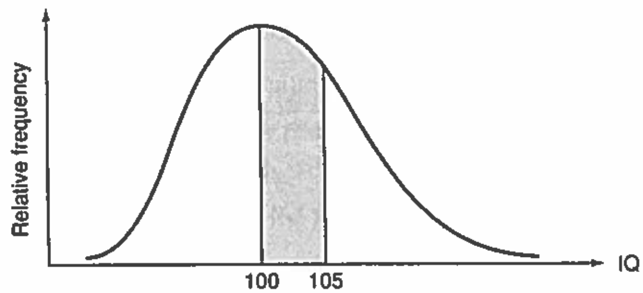
WELL CLASS	AQUIFER	DETECT MTBE
Private	Bedrock	Below Limit
Private	Bedrock	Below Limit
Public	Unconsolidated	Detect
Public	Unconsolidated	Below Limit
Public	Unconsolidated	Below Limit
Public	Unconsolidated	Below Limit
Public	Unconsolidated	Detect
Public	Unconsolidated	Below Limit
Public	Unconsolidated	Below Limit
Public	Bedrock	Detect
Public	Bedrock	Detect

Source: Ayotte, J. D., Argue, D. M., and McGarry, F. J. "Methyl *tert*-butyl ether occurrence and related factors in public and private wells in southeast New Hampshire," *Environmental Science and Technology*, Vol. 39, No. 1, Jan. 2005. Reprinted with permission.

1.4 Describing Quantitative Data Graphically

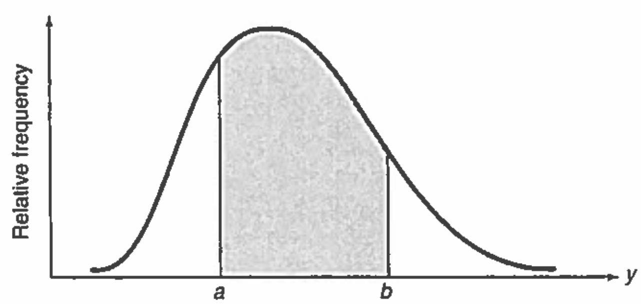
A useful graphical method for describing quantitative data is provided by a relative frequency distribution. Like a bar graph for qualitative data, this type of graph shows the proportions of the total set of measurements that fall in various intervals on the scale of measurement. For example, Figure 1.3 shows the intelligence quotients (IQs) of identical twins. The area over a particular interval under a relative frequency distribution curve is proportional to the fraction of the total number of measurements that fall in that interval. In Figure 1.3, the fraction of the total number of identical twins with IQs that fall between 100 and 105 is proportional to the shaded area. **If we take the total area under the distribution curve as equal to 1, then the shaded area is equal to the fraction of IQs that fall between 100 and 105.**

Figure 1.3 Relative frequency distribution: IQs of identical twins



Throughout this text we denote the quantitative variable measured by the symbol y. Observing a single value of y is equivalent to selecting a single measurement from the population. The probability that it will assume a value in an interval, say, a to b, is given by its relative frequency or **probability distribution**. The total area under a probability distribution curve is always assumed to equal 1. Hence, the probability that a measurement on y will fall in the interval between a and b is equal to the shaded area shown in Figure 1.4.

Figure 1.4 Probability distribution for a quantitative variable



Since the theoretical probability distribution for a quantitative variable is usually unknown, we resort to obtaining a sample from the population: Our objective is to describe the sample and use this information to make inferences about the probability distribution of the population. **Stem-and-leaf plots** and **histograms** are two of the most popular graphical methods for describing quantitative data. Both display the frequency (or relative frequency) of observations that fall into specified intervals (or classes) of the variable's values.

For small data sets (say, 30 or fewer observations) with measurements with only a few digits, stem-and-leaf plots can be constructed easily by hand. Histograms, on the other hand, are better suited to the description of larger data sets, and they permit greater flexibility in the choice of classes. Both, however, can be generated using the computer, as illustrated in the following examples.

Example 1.3

The Environmental Protection Agency (EPA) performs extensive tests on all new car models to determine their highway mileage ratings. The 100 measurements in Table 1.5 represent the results of such tests on a certain new car model.

A visual inspection of the data indicates some obvious facts. For example, most of the mileages are in the 30s, with a smaller fraction in the 40s. But it is difficult to provide much additional information without resorting to a graphical method of summarizing the data. A stem-and-leaf plot for the 100 mileage ratings, produced using MINITAB, is shown in Figure 1.5. Interpret the figure.

EPAGAS

Table 1.5 EPA mileage ratings on 100 cars

36.3	41.0	36.9	37.1	44.9	36.8	30.0	37.2	42.1	36.7
32.7	37.3	41.2	36.6	32.9	36.5	33.2	37.4	37.5	33.6
40.5	36.5	37.6	33.9	40.2	36.4	37.7	37.7	40.0	34.2
36.2	37.9	36.0	37.9	35.9	38.2	38.3	35.7	35.6	35.1
38.5	39.0	35.5	34.8	38.6	39.4	35.3	34.4	38.8	39.7
36.3	36.8	32.5	36.4	40.5	36.6	36.1	38.2	38.4	39.3
41.0	31.8	37.3	33.1	37.0	37.6	37.0	38.7	39.0	35.8
37.0	37.2	40.7	37.4	37.1	37.8	35.9	35.6	36.7	34.5
37.1	40.3	36.7	37.0	33.9	40.1	38.0	35.2	34.8	39.5
39.9	36.9	32.9	33.8	39.8	34.0	36.8	35.0	38.1	36.9

Figure 1.5 MINITAB stem-and-leaf plot for EPA gas mileages

```
Stem-and-leaf of MPG  N  = 100
Leaf Unit = 0.10

  1    30  0
  2    31  8
  6    32  5799
 12    33  126899
 18    34  024588
 29    35  01235667899
 49    36  01233445566777888999
(21)   37  000011122334456677899
 30    38  0122345678
 20    39  00345789
 12    40  0123557
  5    41  002
  2    42  1
  1    43
  1    44  9
```

Solution

In a stem-and-leaf plot, each measurement (mpg) is partitioned into two portions, a *stem* and a *leaf*. MINITAB has selected the digit to the right of the decimal point to represent the leaf and the digits to the left of the decimal point to represent the stem. For example, the value 36.3 mpg is partitioned into a stem of 36 and a leaf of 3, as illustrated below:

Stem	Leaf
36	3

The stems are listed in order in the second column of the MINITAB plot, Figure 1.5, starting with the smallest stem of 30 and ending with the largest stem of 44.

The respective leaves are then placed to the right of the appropriate stem row in increasing order.* For example, the stem row of 32 in Figure 1.5 has four leaves—5, 7, 9, and 9—representing the mileage ratings of 32.5, 32.7, 32.9, and 32.9, respectively. Notice that the stem row of 37 (representing MPGs in the 37's) has the most leaves (21). Thus, 21 of the 100 mileage ratings (or 21%) have values in the 37's. If you examine stem rows 35, 36, 37, 38, and 39 in Figure 1.5 carefully, you will also find that 70 of the 100 mileage ratings (70%) fall between 35.0 and 39.9 mpg. ■

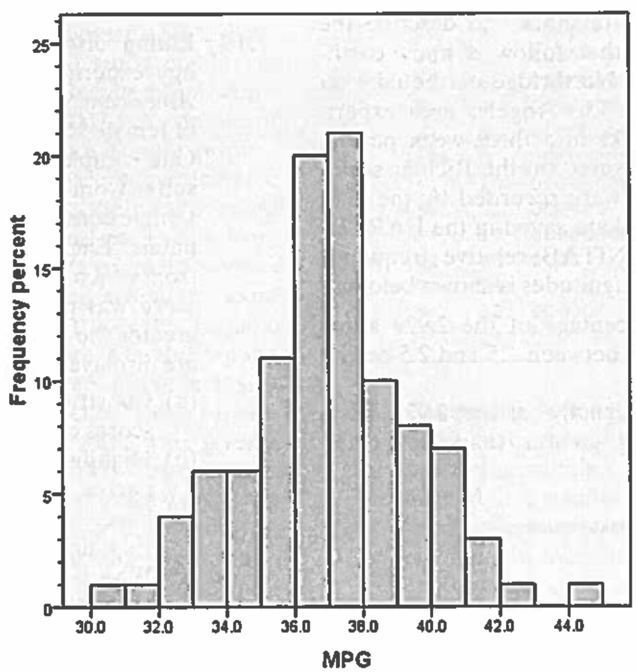
Example 1.4

Refer to Example 1.3. Figure 1.6 is a relative frequency histogram for the 100 EPA gas mileages (Table 1.5) produced using SPSS.

(a) Interpret the graph.

(b) Visually estimate the proportion of mileage ratings in the data set between 36 and 38 MPG.

Figure 1.6 SPSS histogram for 100 EPA gas mileages



Solution

(a) In constructing a histogram, the values of the mileages are divided into the intervals of equal length (1 MPG), called **classes**. The endpoints of these classes are shown on the horizontal axis of Figure 1.6. The relative frequency (or percentage) of gas mileages falling in each class interval is represented by the vertical bars over the class. You can see from Figure 1.6 that the mileages tend to pile up near 37 MPG; in fact, the class interval from 37 to 38 MPG has the greatest relative frequency (represented by the highest bar).

Figure 1.6 also exhibits **symmetry** around the center of the data—that is, a tendency for a class interval to the right of center to have about the same relative frequency as the corresponding class interval to the left of center. This

* The first column in the MINITAB stem-and-leaf plot gives the cumulative number of measurements in the nearest "tail" of the distribution beginning with the stem row.

is in contrast to **positively skewed** distributions (which show a tendency for the data to tail out to the right due to a few extremely large measurements) or to **negatively skewed** distributions (which show a tendency for the data to tail out to the left due to a few extremely small measurements).

(b) The interval 36–38 MPG spans two mileage classes: 36–37 and 37–38. The proportion of mileages between 36 and 38 MPG is equal to the sum of the relative frequencies associated with these two classes. From Figure 1.6 you can see that these two class relative frequencies are .20 and .21, respectively. Consequently, the proportion of gas mileage ratings between 36 and 38 MPG is (.20 + .21) = .41, or 41%.

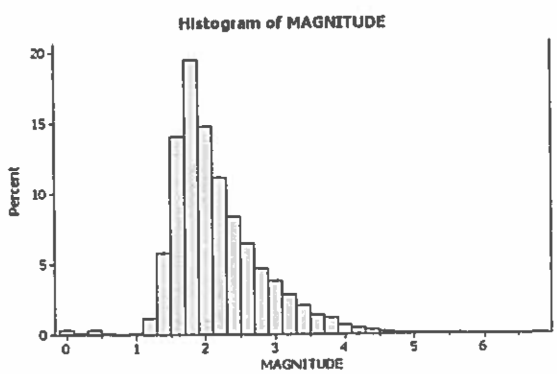
1.4 Exercises

EARTHQUAKE

1.18 Earthquake aftershock magnitudes. Seismologists use the term "aftershock" to describe the smaller earthquakes that follow a main earthquake. Following the Northridge earthquake on January 17, 1994, the Los Angeles area experienced 2,929 aftershocks in a three-week period. The magnitudes (measured on the Richter scale) for these aftershocks were recorded by the U.S. Geological Survey and are saved in the EARTHQUAKE file. A MINITAB relative frequency histogram for these magnitudes is shown below.

(a) Estimate the percentage of the 2,929 aftershocks measuring between 1.5 and 2.5 on the Richter scale.

(b) Estimate the percentage of the 2,929 aftershocks measuring greater than 3.0 on the Richter scale.



(c) Is the aftershock data distribution skewed right, skewed left, or symmetric?

1.19 Eating disorder study. Data from a psychology experiment were reported and analyzed in *American Statistician* (May 2001). Two samples of female students participated in the experiment. One sample consisted of 11 students known to suffer from the eating disorder bulimia; the other sample consisted of 14 students with normal eating habits. Each student completed a questionnaire from which a "fear of negative evaluation" (FNE) score was produced. (The higher the score, the greater the fear of negative evaluation.) The data are displayed in the table at the bottom of the page.

(a) Construct a stem-and-leaf display for the FNE scores of all 25 female students.

(b) Highlight the bulimic students on the graph, part a. Does it appear that bulimics tend to have a greater fear of negative evaluation? Explain.

(c) Why is it important to attach a measure of reliability to the inference made in part b?

1.20 Data on postmortem intervals. *Postmortem interval* (PMI) is defined as the elapsed time between death and an autopsy. Knowledge of PMI is considered essential when conducting medical research on human cadavers. The data in the table (p. 17) are the PMIs of 22 human brain specimens obtained at autopsy in a recent study (*Brain and Language*, June 1995). Graphically describe the PMI data with a stem-and-leaf plot. Based on the plot, make a summary statement about the PMI of the 22 human brain specimens.

BULIMIA

Bulimic students:	21	13	10	20	25	19	16	21	24	13	14			
Normal students:	13	6	16	13	8	19	23	18	11	19	7	10	15	20

Source: Randles, R. H. "On neutral responses (zeros) in the sign test and ties in the Wilcoxon-Mann-Whitney test," *American Statistician*, Vol. 55, No. 2, May 2001 (Figure 3).

BRAINPMI

Postmortem intervals for 22 human brain specimens

5.5	14.5	6.0	5.5	5.3	5.8	11.0	6.1
7.0	14.5	10.4	4.6	4.3	7.2	10.5	6.5
3.3	7.0	4.1	6.2	10.4	4.9		

Source: Reprinted from *Brain and Language,* Vol. 49, Issue 3, T. L. Hayes and D. A. Lewis, "Anatomical Specialization of the Anterior Motor Speech Area: Hemispheric Differences in Magnopyramidal Neurons," p. 292 (Table 1), Copyright © 1995, with permission of Elsevier.

1.21 Is honey a cough remedy? Coughing at night is a common symptom of an upper respiratory tract infection, yet there is no accepted therapeutic cure. Does a teaspoon of honey before bed really calm a child's cough? To test the folk remedy, pediatric researchers at Pennsylvania State University carried out a designed study conducted over two nights (*Archives of Pediatrics and Adolescent Medicine*, December 2007.) A sample of 105 children who were ill with an upper respiratory tract infection and their parents participated in the study. On the first night, the parents rated their children's cough symptoms on a scale from 0 (no problems at all) to 6 (extremely severe) in five different areas. The total symptoms score (ranging from 0 to 30 points) was the variable of interest for the 105 patients. On the second night, the parents were instructed to give their sick child a dosage of liquid "medicine" prior to bedtime. Unknown to the parents, some were given a dosage of dextromethorphan (DM)—an over-the-counter cough medicine—while others were given a similar dose of honey. Also, a third group of parents (the control group) gave their sick children no dosage at all. Again, the parents rated their children's cough symptoms, and the improvement in total cough symptoms score was determined for each child. The data (improvement scores) for the study are shown in the accompanying table, followed by a MINITAB stem-and-leaf plot of the data. Shade the leaves for the honey dosage group on the stem-and-leaf plot. What conclusions can pediatric researchers draw from the graph? Do you agree with the statement (extracted from the article), "honey may be a preferable treatment for the cough and sleep difficulty associated with childhood upper respiratory tract infection"?

```
Stem-and-leaf of TotalScore  N  = 105
Leaf Unit = 0.10

  1    0  0
  4    1  000
  4    2
  7    3  000
 16    4  000000000
 20    5  0000
 28    6  00000000
 41    7  0000000000000
 52    8  00000000000
(13)   9  0000000000000
 40   10  0000000000
 30   11  000000
 24   12  0000000000000
 11   13  0000
  7   14  0
  6   15  00000
  1   16  0
```

1.22 Comparing voltage readings. A Harris Corporation/University of Florida study was undertaken to determine whether a manufacturing process performed at a remote location could be established locally. Test devices (pilots) were setup at both the old and new locations, and voltage readings on the process were obtained. A "good" process was considered to be one with voltage readings of at least 9.2 volts (with larger readings better than smaller readings). The first table on p. 18 contains voltage readings for 30 production runs at each location.

HONEYCOUGH

Honey Dosage:	12 11 15 11 10 13 10 4 15 16 9 14 10 6 10 8 11 12 12 8
	12 9 11 15 10 15 9 13 8 12 10 8 9 5 12
DM Dosage:	4 6 9 4 7 7 7 9 12 10 11 6 3 4 9 12 7 6 8 12 12 4 12
	13 7 10 13 9 4 4 10 15 9
No Dosage (Control):	5 8 6 1 0 8 12 8 7 7 1 6 7 7 12 7 9 7 9 5 11 9 5
	6 8 8 6 7 10 9 4 8 7 3 1 4 3

Source: Paul, I. M., et al. "Effect of honey, dextromethorphan, and no treatment on nocturnal cough and sleep quality for coughing children and their parents," *Archives of Pediatrics and Adolescent Medicine*, Vol. 161, No. 12, Dec. 2007 (data simulated).

VOLTAGE

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.

(a) Construct a relative frequency histogram for the voltage readings of the old process.

(b) Construct a stem-and-leaf display for the voltage readings of the old process. Which of the two graphs in parts a and b is more informative?

(c) Construct a frequency histogram for the voltage readings of the new process.

(d) Compare the two graphs in parts a and c. (You may want to draw the two histograms on the same graph.) Does it appear that the manufacturing process can be established locally (i.e., is the new process as good as or better than the old)?

1.23 Sanitation inspection of cruise ships. To minimize the potential for gastrointestinal disease outbreaks, all passenger cruise ships arriving at U.S. ports are subject to unannounced sanitation inspections. Ships are rated on a 100-point scale by the Centers for Disease Control and Prevention. A score of 86 or higher indicates that the ship is providing an accepted standard of sanitation. The May 2006 sanitation scores for 169 cruise ships are saved in the SHIPSANIT file. The first five and last five observations in the data set are listed in the accompanying table.

(a) Generate a stem-and-leaf display of the data. Identify the stems and leaves of the graph.

(b) Use the stem-and-leaf display to estimate the proportion of ships that have an accepted sanitation standard.

(c) Locate the inspection score of 84 (*Sea Bird*) on the stem-and-leaf display.

(d) Generate a histogram for the data.

(e) Use the histogram to estimate the proportion of ships that have an accepted sanitation standard.

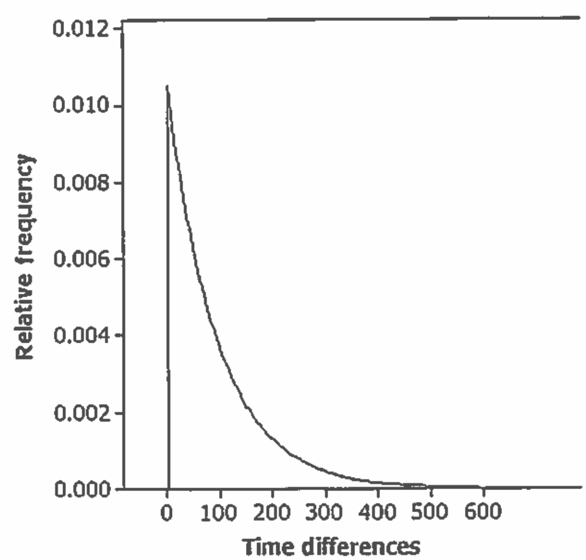
SHIPSANIT (selected observations)

SHIP NAME	SANITATION SCORE
Adventure of the Seas	95
Albatross	96
Amsterdam	98
Arabella	94
Arcadia	98
.	.
.	.
Wind Surf	95
Yorktown Clipper	91
Zaandam	98
Zenith	94
Zuiderdam	94

Source: National Center for Environmental Health, Centers for Disease Control and Prevention, May 24, 2006.

PHISHING

1.24 Phishing attacks to email accounts. *Phishing* is the term used to describe an attempt to extract personal/financial information (e.g., PIN numbers, credit card information, bank account numbers) from unsuspecting people through fraudulent email. An article in *Chance* (Summer 2007) demonstrates how statistics can help identify phishing attempts and make e-commerce safer. Data from an actual phishing attack against an organization were used to determine whether the attack may have been an "inside job" that originated within the company. The company setup a publicized email account—called a "fraud box"—that enabled employees to notify them if they suspected an email phishing attack.



The interarrival times, that is, the time differences (in seconds), for 267 fraud box email notifications were recorded. *Chance* showed that if there is minimal or no collaboration or collusion from within the company, the interarrival times would have a frequency distribution similar to the one shown in the accompanying figure (p. 18). The 267 interarrival times are saved in the PHISHING file. Construct a frequency histogram for the interarrival times. Is the data skewed to the right? Give your opinion on whether the phishing attack against the organization was an "inside job."

1.5 Describing Quantitative Data Numerically

Numerical descriptive measures provide a second (and often more powerful) method for describing a set of quantitative data. These measures, which locate the center of the data set and its spread, actually enable you to construct an approximate mental image of the distribution of the data set.

Note: Most of the formulas used to compute numerical descriptive measures require the summation of numbers. For instance, we may want to sum the observations in a data set, or we may want to square each observation and then sum the squared values. The symbol Σ (sigma) is used to denote a summation operation.

For example, suppose we denote the n sample measurements on a random variable y by the symbols $y_1, y_2, y_3, \ldots, y_n$. Then the sum of all n measurements in the sample is represented by the symbol

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

This is read "summation y, y_1 to y_n" and is equal to the value

$$y_1 + y_2 + y_3 + \cdots + y_n$$

One of the most common measures of central tendency is the **mean**, or arithmetic average, of a data set. Thus, if we denote the sample measurements by the symbols $y_1, y_2, y_3, \ldots,$ the sample mean is defined as follows:

Definition 1.15 The **mean** of a sample of n measurements $y_1, y_2, \ldots, y_n$ is

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

The mean of a population, or equivalently, the expected value of y, $E(y)$, is usually unknown in a practical situation (we will want to infer its value based on the sample data). Most texts use the symbol μ to denote the mean of a population. Thus, we use the following notation:

Notation

Sample mean: $\bar{y}$
Population mean: $E(y) = \mu$

The spread or variation of a data set is measured by its **range**, its **variance**, or its **standard deviation**.

Definition 1.16 The **range** of a sample of n measurements $y_1, y_2, \ldots, y_n$ is the difference between the largest and smallest measurements in the sample.

Example 1.5

If a sample consists of measurements 3, 1, 0, 4, 7, find the sample mean and the sample range.

Solution

The sample mean and range are

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{15}{5} = 3$$

$$\text{Range} = 7 - 0 = 7$$

The variance of a set of measurements is defined to be the average of the *squares of the deviations* of the measurements about their mean. Thus, the population variance, which is usually unknown in a practical situation, would be the mean or expected value of $(y - \mu)^2$, or $E[(y - \mu)^2]$. We use the symbol σ^2 to represent the variance of a population:

$$E[(y - \mu)^2] = \sigma^2$$

The quantity usually termed the **sample variance** is defined in the box.

Definition 1.17 The **variance** of a sample of n measurements $y_1, y_2, \ldots, y_n$ is defined to be

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

Note that the sum of squares of deviations in the sample variance is divided by $(n - 1)$, rather than n. Division by n produces estimates that tend to underestimate σ^2. Division by $(n - 1)$ corrects this problem.

Example 1.6

Refer to Example 1.5. Calculate the sample variance for the sample 3, 1, 0, 4, 7.

Solution

We first calculate

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 75 - 5(3)^2 = 30$$

where $\bar{y} = 3$ from Example 1.4. Then

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1} = \frac{30}{4} = 7.5$$

The concept of a variance is important in theoretical statistics, but its square root, called a **standard deviation**, is the quantity most often used to describe data variation.

> **Definition 1.18** The **standard deviation** of a set of measurements is equal to the square root of their variance. Thus, the standard deviations of a sample and a population are
>
> **Sample standard deviation:** s
> **Population standard deviation:** σ

The standard deviation of a set of data takes on meaning in light of a theorem (Tchebysheff's theorem) and a rule of thumb.* Basically, they give us the following guidelines:

> **Guidelines for Interpreting a Standard Deviation**
>
> 1. For *any* data set (population or sample), at least three-fourths of the measurements will lie within 2 standard deviations of their mean.
> 2. For *most* data sets of moderate size (say, 25 or more measurements) with a mound-shaped distribution, approximately 95% of the measurements will lie within 2 standard deviations of their mean.

Example 1.7

Often, travelers who have no intention of showing up fail to cancel their hotel reservations in a timely manner. These travelers are known, in the parlance of the hospitality trade, as "no-shows." To protect against no-shows and late cancellations, hotels invariably overbook rooms. A study reported in the *Journal of Travel Research* examined the problems of overbooking rooms in the hotel industry. The data in Table 1.6, extracted from the study, represent daily numbers of late cancellations and no-shows for a random sample of 30 days at a large (500-room) hotel. Based on this sample, how many rooms, at minimum, should the hotel overbook each day?

NOSHOWS

Table 1.6 Hotel no-shows for a sample of 30 days

18	16	16	16	14	18	16	18	14	19
15	19	9	20	10	10	12	14	18	12
14	14	17	12	18	13	15	13	15	19

Source: Toh, R. S. "An inventory depletion overbooking model for the hotel industry," *Journal of Travel Research*, Vol. 23, No. 4, Spring 1985, p. 27. The *Journal of Travel Research* is published by the Travel and Tourism Research Association (TTRA) and the Business Research Division, University of Colorado at Boulder.

Solution

To answer this question, we need to know the range of values where most of the daily numbers of no-shows fall. We must compute $\bar{y}$ and s, and examine the shape of the relative frequency distribution for the data.

* For a more complete discussion and a statement of Tchebysheff's theorem, see the references listed at the end of this chapter.

Figure 1.7 is a MINITAB printout that shows a stem-and-leaf display and descriptive statistics of the sample data. Notice from the stem-and-leaf display that the distribution of daily no-shows is mound-shaped, and only slightly skewed on the low (top) side of Figure 1.7. Thus, guideline 2 in the previous box should give a good estimate of the percentage of days that fall within 2 standard deviations of the mean.

Figure 1.7 MINITAB printout: Describing the no-show data, Example 1.6

Stem-and-Leaf Display: Noshows

```
Stem-and-leaf of Noshows   N  = 30
Leaf Unit = 0.10

  1    9 0
  3   10 00
  3   11
  6   12 000
  8   13 00
 13   14 00000
(3)   15 000
 14   16 0000
 10   17 0
  9   18 00000
  4   19 000
  1   20 0
```

Descriptive Statistics: Noshows

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Noshows	30	15.133	15.000	15.231	2.945	0.538

Variable	Minimum	Maximum	Q1	Q3
Noshows	9.000	20.000	13.000	18.000

The mean and standard deviation of the sample data, shaded on the MINITAB printout, are $\bar{y} = 15.133$ and $s = 2.945$. From guideline 2 in the box, we know that about 95% of the daily number of no-shows fall within 2 standard deviations of the mean, that is, within the interval

$$\bar{y} \pm 2s = 15.133 \pm 2(2.945)$$
$$= 15.133 \pm 5.890$$

or between 9.243 no-shows and 21.023 no-shows. (If we count the number of measurements in this data set, we find that actually 29 out of 30, or 96.7%, fall in this interval.)

From this result, the large hotel can infer that there will be at least 9.243 (or, rounding up, 10) no-shows per day. Consequently, the hotel can overbook at least 10 rooms per day and still be highly confident that all reservations can be honored. ■

Numerical descriptive measures calculated from sample data are called **statistics.** Numerical descriptive measures of the population are called **parameters.** In a practical situation, we will not know the population relative frequency distribution (or equivalently, the population distribution for y). We will usually assume that it has unknown numerical descriptive measures, such as its mean μ and standard deviation σ, and by inferring (using **sample statistics**) the values of these parameters, we infer the nature of the population relative frequency distribution. Sometimes we will assume that we know the shape of the population relative frequency distribution and use this information to help us make our inferences. When we do this, we are postulating a model for the population relative frequency distribution, and we must keep in mind that the validity of the inference may depend on how well our model fits reality.

> **Definition 1.19** Numerical descriptive measures of a population are called **parameters**.

> **Definition 1.20** A **sample statistic** is a quantity calculated from the observations in a sample.

1.5 Exercises

EARTHQUAKE

1.25 Earthquake aftershock magnitudes. Refer to Exercise 1.18 (p. 16) and U.S. Geological Survey data on aftershocks from a major California earthquake. The EARTHQUAKE file contains the magnitudes (measured on the Richter scale) for 2,929 aftershocks. A MINITAB printout with descriptive statistics of magnitude is shown at the bottom of the page.

(a) Locate and interpret the mean of the magnitudes for the 2,929 aftershocks.

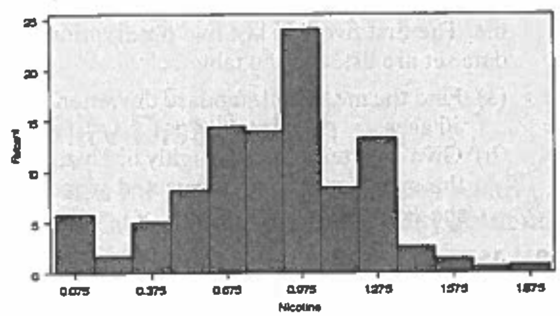
(b) Locate and interpret the range of the magnitudes for the 2,929 aftershocks.

(c) Locate and interpret the standard deviation of the magnitudes for the 2,929 aftershocks.

(d) If the target of your interest is these specific 2,929 aftershocks, what symbols should you use to describe the mean and standard deviation?

FTC

1.26 FTC cigarette rankings. Periodically, the Federal Trade Commission (FTC) ranks domestic cigarette brands according to tar, nicotine, and carbon monoxide content. The test results are obtained by using a sequential smoking machine to "smoke" cigarettes to a 23-millimeter butt length. The tar, nicotine, and carbon monoxide concentrations (rounded to the nearest milligram) in the residual "dry" particulate matter of the smoke are then measured. The accompanying SAS printouts describe the nicotine contents of 500 cigarette brands. (The data are saved in the FTC file.)



The MEANS Procedure

Analysis Variable : NICOTINE

N	Mean	Std Dev	Minimum	Maximum
500	0.8425000	0.3455250	0.0500000	1.9000000

(a) Examine the relative frequency histogram for nicotine content. Use the rule of thumb to describe the data set.

(b) Locate $\bar{y}$ and s on the printout, then compute the interval $\bar{y} \pm 2s$.

(c) Based on your answer to part a, estimate the percentage of cigarettes with nicotine contents in the interval formed in part b.

(d) Use the information on the SAS histogram to determine the actual percentage of nicotine contents that fall within the interval formed

Descriptive Statistics: MAGNITUDE

Variable	N	Mean	StDev	Variance	Minimum	Median	Maximum	Range
MAGNITUDE	2929	2.1197	0.6636	0.4403	0.0000	2.0000	6.7000	6.7000

Variable	Mode	N for Mode
MAGNITUDE	1.8	298

in part b. Does your answer agree with your estimate of part c?

SHIPSANIT

1.27 Sanitation inspection of cruise ships. Refer to the Centers for Disease Control and Prevention study of sanitation levels for 169 international cruise ships, Exercise 1.23 (p. 18). (Recall that sanitation scores range from 0 to 100.)

(a) Find $\bar{y}$ and s for the 169 sanitation scores.
(b) Calculate the interval $\bar{y} \pm 2s$.
(c) Find the percentage of scores in the data set that fall within the interval, part b. Does the result agree with the rule of thumb given in this section?

1.28 Most powerful business women in America. *Fortune* (October 16, 2008) published a list of the 50 most powerful women in business in the United States. The data on age (in years) and title of each of these 50 women are stored in the WPOWER50 file. The first five and last five observations of the data set are listed in the table below.

(a) Find the mean and standard deviation of these 50 ages.
(b) Give an interval that is highly likely to contain the age of a randomly selected woman from the *Fortune* list.

1.29 Ammonia from vehicle exhaust. Three-way catalytic converters have been installed in new vehicles in order to reduce pollutants from motor vehicle exhaust emissions. However, these converters unintentionally increase the level of ammonia in the air. *Environmental Science and Technology* (September 1, 2000) published a study on the ammonia levels near the exit ramp of a San Francisco highway tunnel. The data in the next table represent daily ammonia concentrations (parts per million) on eight randomly selected days during afternoon drive-time in the summer of 1999.

AMMONIA

1.53	1.50	1.37	1.51	1.55	1.42	1.41	1.48

(a) Find and interpret the mean daily ammonia level in air in the tunnel.
(b) Find the standard deviation of the daily ammonia levels. Interpret the result.
(c) Suppose the standard deviation of the ammonia levels during morning drive-time at the exit ramp is 1.45 ppm. Which time, morning or afternoon drive-time, has more variable ammonia levels?

1.30 Animal-assisted therapy for heart patients. Medical researchers at an *American Heart Association Conference* (November 2005) presented a study to gauge whether animal-assisted therapy can improve the physiological responses of heart patients. A team of nurses from the UCLA Medical Center randomly divided 76 heart patients into three groups. Each patient in group T was visited by a human volunteer accompanied by a trained dog; each patient in group V was visited by a volunteer only; and the patients in group C were not visited at all. The anxiety level of each patient was measured (in points) both before and after the visits. The next table (p. 25) gives summary statistics for the drop in anxiety level for patients in the three groups. Suppose the anxiety level of a patient selected from the study had a drop of 22.5 points. Which group is the patient more likely to have come from? Explain.

WPOWER50 (selected observations)

RANK	NAME	AGE	COMPANY	TITLE
1	Indra Nooyi	52	PepsiCo	CEO/Chairman
2	Irene Rosenfeld	55	Kraft Foods	CEO/Chairman
3	Pat Woertz	55	Archer Daniels Midland	CEO/Chairman
4	Anne Mulcahy	55	Xerox	CEO/Chairman
5	Angela Braley	47	Wellpoint	CEO/President
.	.	.		
.	.	.		
46	Lorrie Norrington	48	eBay	CEO
47	Terri Dial	58	Citigroup	CEO
48	Lynn Elsenhans	52	Sunoco	CEO/President
49	Cathie Black	64	Hearst Magazines	President
50	Marissa Mayer	33	Google	VP

Source: Fortune, Oct. 16, 2008.

Summary table for Exercise 1.30

	SAMPLE SIZE	MEAN DROP	STD. DEV.
Group T: Volunteer + Trained Dog	26	10.5	7.6
Group V: Volunteer only	25	3.9	7.5
Group C: Control group (no visit)	25	1.4	7.5

Source: Cole, K., et al. "Animal assisted therapy decreases hemodynamics, plasma epinephrine and state anxiety in hospitalized heart failure patients," *American Heart Association Conference,* Dallas, Texas, Nov. 2005.

1.31 Improving SAT scores. The National Education Longitudinal Survey (NELS) tracks a nationally representative sample of U.S. students from eighth grade through high school and college. Research published in *Chance* (Winter 2001) examined the Standardized Admission Test (SAT) scores of 265 NELS students who paid a private tutor to help them improve their scores. The table below summarizes the changes in both the SAT-Mathematics and SAT-Verbal scores for these students.

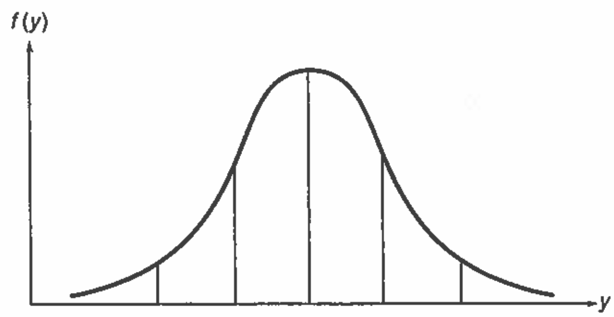
	SAT-MATH	SAT-VERBAL
Mean change in score	19	7
Standard deviation of score changes	65	49

(a) Suppose one of the 265 students who paid a private tutor is selected at random. Give an interval that is likely to contain this student's change in the SAT-Math score.
(b) Repeat part a for the SAT-Verbal score.
(c) Suppose the selected student's score increased on one of the SAT tests by 140 points. Which test, the SAT-Math or SAT-Verbal, is the one most likely to have the 140-point increase? Explain.

1.6 The Normal Probability Distribution

One of the most commonly used models for a theoretical population relative frequency distribution for a quantitative variable is the **normal probability distribution,** as shown in Figure 1.8. The normal distribution is symmetric about its mean μ, and its spread is determined by the value of its standard deviation σ. Three normal curves with different means and standard deviations are shown in Figure 1.9.

Figure 1.8 A normal probability distribution



Computing the area over an interval under the normal probability distribution can be a difficult task.* Consequently, we will use the computed areas listed in Table 1 of Appendix D. A partial reproduction of this table is shown in Table 1.7. As you can see from the normal curve above the table, the entries give areas under the normal curve between the mean of the distribution and a standardized distance

$$z = \frac{y - \mu}{\sigma}$$

* Students with knowledge of calculus should note that the probability that y assumes a value in the interval $a < y < b$ is $P(a < y < b) = \int_a^b f(y)dy$, assuming the integral exists. The value of this definite integral can be obtained to any desired degree of accuracy by approximation procedures. For this reason, it is tabulated for the user.

to the right of the mean. Note that z is the number of standard deviations σ between μ and y. The distribution of z, which has mean $\mu = 0$ and standard deviation $\sigma = 1$, is called a **standard normal distribution.**

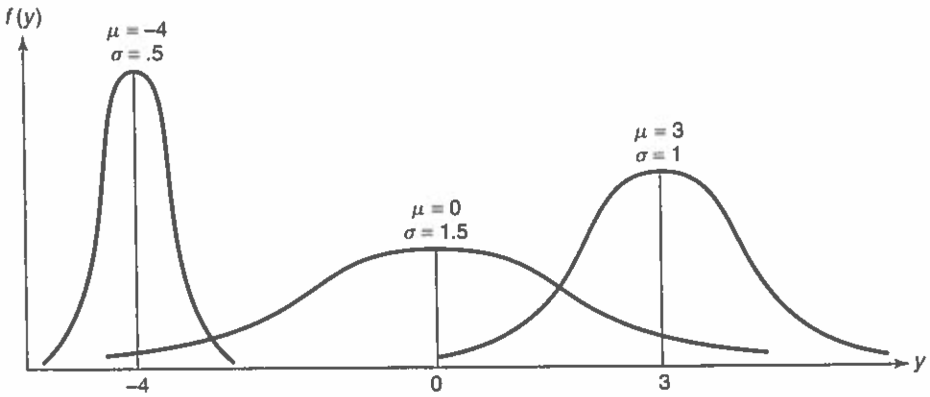
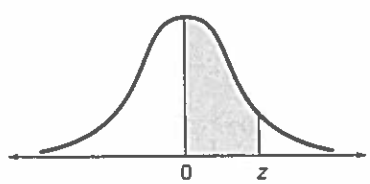


Figure 1.9 Several normal distributions with different means and standard deviations

Table 1.7 Reproduction of part of Table 1 of Appendix D



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441

Example 1.8

Suppose y is a normal random variable with $\mu = 50$ and $\sigma = 15$. Find $P(30 < y < 70)$, the probability that y will fall within the interval $30 < y < 70$.

Solution

Refer to Figure 1.10. Note that $y = 30$ and $y = 70$ lie the same distance from the mean $\mu = 50$, with $y = 30$ below the mean and $y = 70$ above it. Then, because the normal curve is symmetric about the mean, the probability A_1 that y falls between $y = 30$ and $\mu = 50$ is equal to the probability A_2 that y falls between $\mu = 50$ and $y = 70$. The z score corresponding to $y = 70$ is

$$z = \frac{y-\mu}{\sigma} = \frac{70-50}{15} = 1.33$$

Therefore, the area between the mean $\mu = 50$ and the point $y = 70$ is given in Table 1 of Appendix D (and Table 1.7) at the intersection of the row corresponding to $z = 1.3$ and the column corresponding to .03. This area (probability) is $A_2 = .4082$. Since $A_1 = A_2$, A_1 also equals .4082; it follows that the probability that y falls in the interval $30 < y < 70$ is $P(30 < y < 70) = 2(.4082) = .8164$. The z scores corresponding to $y = 30$ ($z = -1.33$) and $y = 70$ ($z = 1.33$) are shown in Figure 1.11. ■

Figure 1.10 Normal probability distribution: $\mu = 50, \sigma = 15$

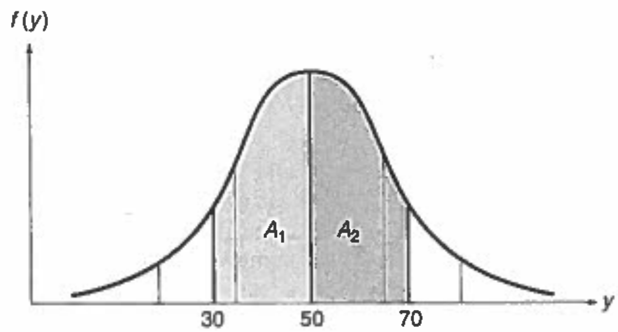
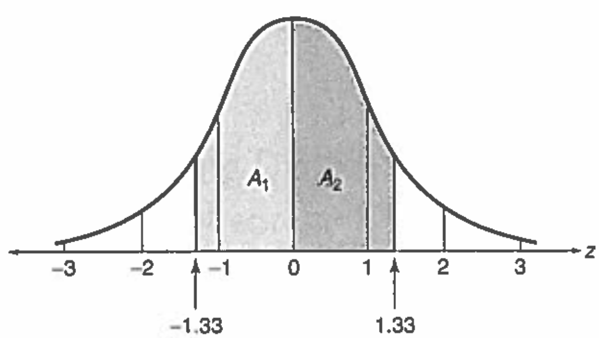


Figure 1.11 A distribution of z scores (a standard normal distribution)



Example 1.9

Use Table 1 of Appendix D to determine the area to the right of the z score 1.64 for the standard normal distribution. That is, find $P(z \geq 1.64)$.

Solution

The probability that a normal random variable will fall more than 1.64 standard deviations to the right of its mean is indicated in Figure 1.12. Because the normal distribution is symmetric, half of the total probability (.5) lies to the right of the mean and half to the left. Therefore, the desired probability is

$$P(z \geq 1.64) = .5 - A$$

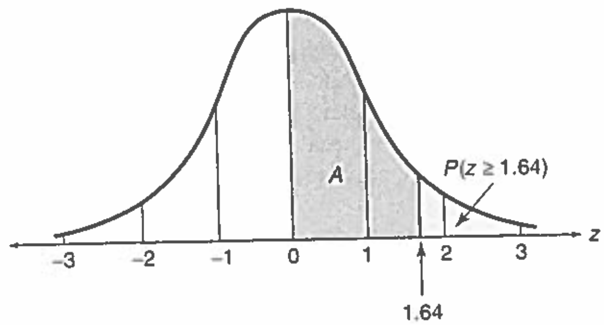


Figure 1.12 Standard normal distribution: $\mu = 0$, $\sigma = 1$

where A is the area between $\mu = 0$ and $z = 1.64$, as shown in the figure. Referring to Table 1, we find that the area A corresponding to $z = 1.64$ is .4495. So

$$P(z \geq 1.64) = .5 - A = .5 - .4495 = .0505$$

We will not be making extensive use of the table of areas under the normal curve, but you should know some of the common tabulated areas. In particular, you should note that the area between $z = -2.0$ and $z = 2.0$, which gives the probability that y falls in the interval $\mu - 2\sigma < y < \mu + 2\sigma$, is .9544 and agrees with guideline 2 of Section 1.4.

1.6 Exercises

1.32 Normal Probabilities. Use Table 1 of Appendix D to find each of the following:

(a) $P(-1 \leq z \leq 1)$ (b) $P(-1.96 \leq z \leq 1.96)$
(c) $P(-1.645 \leq z \leq 1.645)$ (d) $P(-3 \leq z \leq 3)$

1.33 Normal probabilities. Given that the random variable y has a normal probability distribution with mean 100 and variance 64, draw a sketch (i.e., graph) of the frequency function of y. Locate μ and the interval $\mu \pm 2\sigma$ on the graph. Find the following probabilities:

(a) $P(\mu - 2\sigma \leq y \leq \mu + 2\sigma)$ (b) $P(y \geq 108)$
(c) $P(y \leq 92)$ (d) $P(92 \leq y \leq 116)$
(e) $P(92 \leq y \leq 96)$ (f) $P(76 \leq y \leq 124)$

1.34 Transmission delays in wireless technology. Resource reservation protocol (RSVP) was originally designed to establish signaling links for stationary networks. In *Mobile Networks and Applications* (December 2003), RSVP was applied to mobile wireless technology (e.g., a PC notebook with wireless LAN card for Internet access). A simulation study revealed that the transmission delay (measured in milliseconds) of an RSVP linked wireless device has an approximate normal distribution with mean $\mu = 48.5$ milliseconds and $\sigma = 8.5$ milliseconds.

(a) What is the probability that the transmission delay is less than 57 milliseconds?
(b) What is the probability that the transmission delay is between 40 and 60 milliseconds?

1.35 Alkalinity of water. The alkalinity level of water specimens collected from the Han River in Seoul, Korea, has a mean of 50 milligrams per liter and a standard deviation of 3.2 milligrams per liter (*Environmental Science and Engineering*, September 1, 2000). Assume the distribution of alkalinity levels is approximately normal and find the probability that a water specimen collected from the river has an alkalinity level

(a) exceeding 45 milligrams per liter.
(b) below 55 milligrams per liter.
(c) between 51 and 52 milligrams per liter.

1.36 Range of women's heights. In *Chance* (Winter 2007), Yale Law School professor Ian Ayres published the results of a study he conducted with his son and daughter on whether college students could estimate a range for women's heights. The students were shown a graph of a normal distribution of heights and asked: "The average height of women over 20 years old in the United States is 64 inches. Using your intuition, please give your best estimate of the range of heights that would include 90% of women over 20 years old. Please make sure that the center of the range is the average height of 64 inches." The standard deviation of heights for women over 20 years old is known to be 2.6 inches. Find the range of interest.

1.37 **Psychological experiment on alcohol and threats.** A group of Florida State University psychologists examined the effects of alcohol on the reactions of people to a threat (*Journal of Abnormal Psychology*, Vol. 107, 1998). After obtaining a specified blood alcohol level, experimental subjects were placed in a room and threatened with electric shocks. Using sophisticated equipment to monitor the subjects' eye movements, the startle response (measured in milliseconds) was recorded for each subject. The mean and standard deviation of the startle responses were 37.9 and 12.4, respectively. Assume that the startle response y for a person with the specified blood alcohol level is approximately normally distributed.

(a) Find the probability that y is between 40 and 50 milliseconds.
(b) Find the probability that y is less than 30 milliseconds.
(c) Give an interval for y, centered around 37.9 milliseconds, so that the probability that y falls in the interval is .95.

1.38 **Modeling length of gestation.** Based on data from the National Center for Health Statistics, N. Wetzel used the normal distribution to model the length of gestation for pregnant U.S. women (*Chance*, Spring 2001). Gestation length has a mean of 280 days with a standard deviation of 20 days.

(a) Find the probability that gestation length is between 275.5 and 276.5 days. (This estimates the probability that a woman has her baby 4 days earlier than the "average" due date.)
(b) Find the probability that gestation length is between 258.5 and 259.5 days. (This estimates the probability that a woman has her baby 21 days earlier than the "average" due date.)
(c) Find the probability that gestation length is between 254.5 and 255.5 days. (This estimates the probability that a woman has her baby 25 days earlier than the "average" due date.)
(d) The *Chance* article referenced a newspaper story about three sisters who all gave birth on the same day (March 11, 1998). Karralee had her baby 4 days early; Marrianne had her baby 21 days early; and Jennifer had her baby 25 days early. Use the results, parts a–c, to estimate the probability that three women have their babies 4, 21, and 25 days early, respectively. Assume the births are independent events. [Hint: If events A, B, and C are independent, then P(A and B and C) = P(A) × P(B) × P(C).]

1.39 **Mean shifts on a production line.** *Six Sigma* is a comprehensive approach to quality goal setting that involves statistics. An article in *Aircraft Engineering and Aerospace Technology* (Vol. 76, No. 6, 2004) demonstrated the use of the normal distribution in Six Sigma goal setting at Motorola Corporation. Motorola discovered that the average defect rate for parts produced on an assembly line varies from run to run, and is approximately normally distributed with a mean equal to 3 defects per million. Assume that the goal at Motorola is for the average defect rate to vary no more than 1.5 standard deviations above or below the mean of 3. How likely is it that the goal will be met?

1.7 Sampling Distributions and the Central Limit Theorem

Since we use sample statistics to make inferences about population parameters, it is natural that we would want to know something about the reliability of the resulting inferences. For example, if we use a statistic to estimate the value of a population mean μ, we will want to know how close to μ our estimate is likely to fall. To answer this question, we need to know the probability distribution of the statistic.

The probability distribution for a statistic based on a random sample of n measurements could be generated in the following way. For purposes of illustration, we suppose we are sampling from a population with $\mu = 10$ and $\sigma = 5$, the sample statistic is $\bar{y}$, and the sample size is $n = 25$. Draw a single random sample of 25 measurements from the population and suppose that $\bar{y} = 9.8$. Return the measurements to the population and try again. That is, draw another random sample of $n = 25$ measurements and see what you obtain for an outcome. Now, perhaps, $\bar{y} = 11.4$. Replace these measurements, draw another sample of $n = 25$ measurements, calculate $\bar{y}$, and so on. If this sampling process were repeated over and over again an infinitely large number of times, you would generate an infinitely large

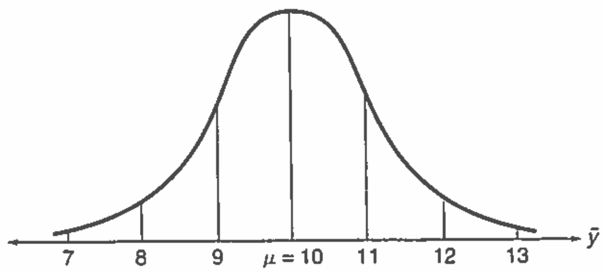


Figure 1.13 Sampling distribution for $\bar{y}$ based on a sample of $n = 25$ measurements

number of values of $\bar{y}$ that could be arranged in a relative frequency distribution. This distribution, which would appear as shown in Figure 1.13, is the probability distribution (or **sampling distribution**, as it is commonly called) of the statistic $\bar{y}$.

> **Definition 1.21** The **sampling distribution** of a sample statistic calculated from a sample of n measurements is the probability distribution of the statistic.

In actual practice, the sampling distribution of a statistic is obtained mathematically or by simulating the sampling on a computer using the procedure described previously.

If $\bar{y}$ has been calculated from a sample of $n = 25$ measurements selected from a population with mean $\mu = 10$ and standard deviation $\sigma = 5$, the sampling distribution shown in Figure 1.13 provides all the information you may wish to know about its behavior. For example, the probability that you will draw a sample of 25 measurements and obtain a value of $\bar{y}$ in the interval $9 \leq \bar{y} \leq 10$ will be the area under the sampling distribution over that interval.

Generally speaking, if we use a statistic to make an inference about a population parameter, we want its sampling distribution to center about the parameter (as is the case in Figure 1.13) and the standard deviation of the sampling distribution, called the **standard error of estimate**, to be as small as possible.

Two theorems provide information on the sampling distribution of a sample mean.

Theorem 1.1 If $y_1, y_2, \ldots, y_n$ represent a random sample of n measurements from a large (or infinite) population with mean μ and standard deviation σ, then, regardless of the form of the population relative frequency distribution, the mean and standard error of estimate of the sampling distribution of $\bar{y}$ will be

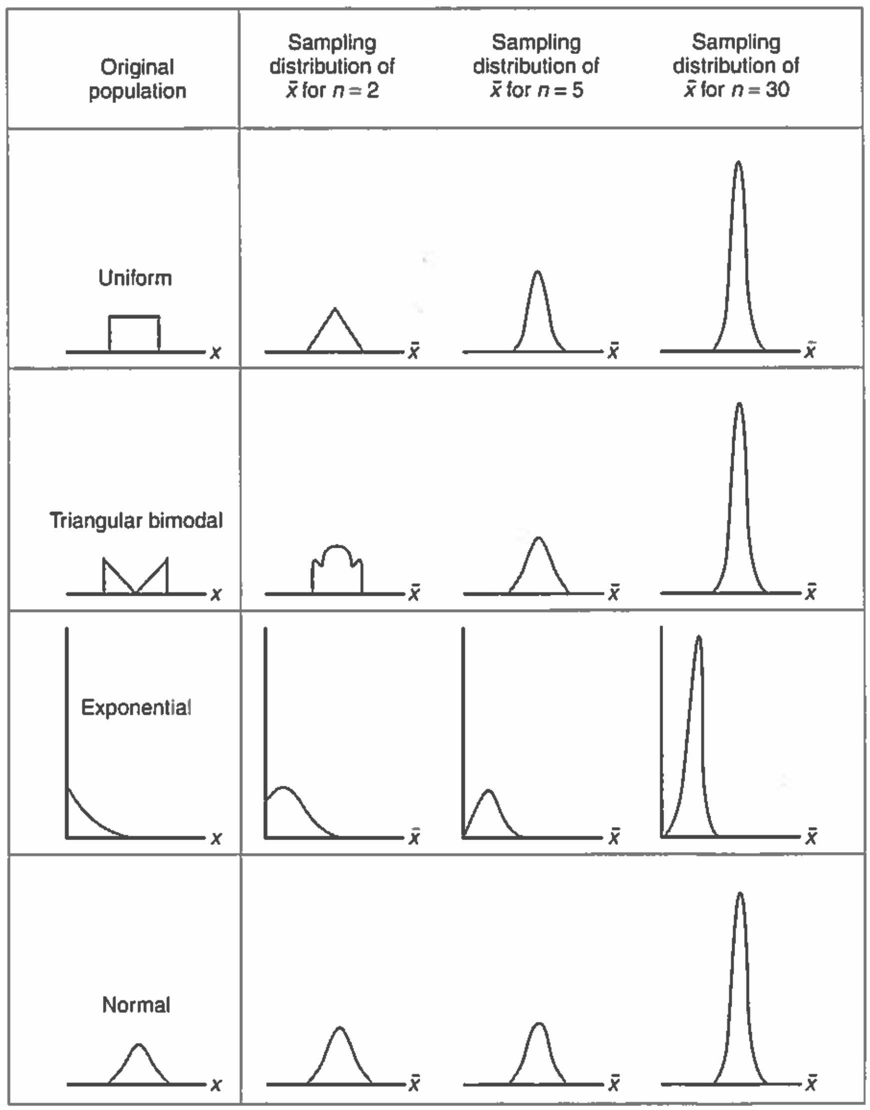
Mean: $E(\bar{y}) = \mu_{\bar{y}} = \mu$
Standard error of estimate: $\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$

Theorem 1.2 **The Central Limit Theorem** For large sample sizes, the mean $\bar{y}$ of a sample from a population with mean μ and standard deviation σ has a sampling distribution that is approximately normal, **regardless of the probability distribution of the sampled population.** The larger the sample size, the better will be the normal approximation to the sampling distribution of $\bar{y}$.

Theorems 1.1 and 1.2 together imply that for sufficiently large samples, the sampling distribution for the sample mean $\bar{y}$ will be approximately normal with mean μ and standard error $\sigma_{\bar{y}} = \sigma/\sqrt{n}$. The parameters μ and σ are the mean and standard deviation of the sampled population.

How large must the sample size n be so that the normal distribution provides a good approximation for the sampling distribution of $\bar{y}$? The answer depends on the shape of the distribution of the sampled population, as shown by Figure 1.14. Generally speaking, the greater the skewness of the sampled population distribution, the larger the sample size must be before the normal distribution is an adequate approximation for the sampling distribution of $\bar{y}$. For most sampled populations, sample sizes of $n \geq 30$ will suffice for the normal approximation to be reasonable. We will use the normal approximation for the sampling distribution of $\bar{y}$ when the sample size is at least 30.

Figure 1.14 Sampling distributions of $\bar{x}$ for different populations and different sample sizes



Example 1.10

Suppose we have selected a random sample of $n = 25$ observations from a population with mean equal to 80 and standard deviation equal to 5. It is known that the population is not extremely skewed.

(a) Sketch the relative frequency distributions for the population and for the sampling distribution of the sample mean, $\bar{y}$.

(b) Find the probability that $\bar{y}$ will be larger than 82.

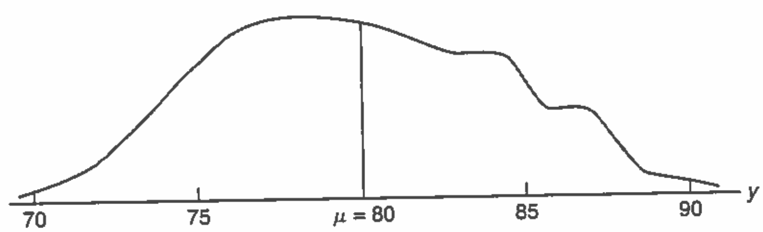
Solution

(a) We do not know the exact shape of the population relative frequency distribution, but we do know that it should be centered about $\mu = 80$, its spread should be measured by $\sigma = 5$, and it is not highly skewed. One possibility is shown in Figure 1.15(a). From the central limit theorem, we know that the sampling distribution of $\bar{y}$ will be approximately normal since the sampled population distribution is not extremely skewed. We also know that the sampling distribution will have mean and standard deviation

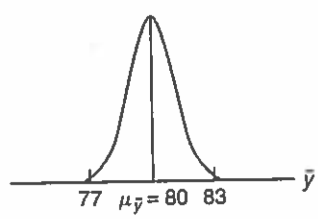
$$\mu_{\bar{y}} = \mu = 80 \quad \text{and} \quad \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$$

The sampling distribution of $\bar{y}$ is shown in Figure 1.15(b).

Figure 1.15 A population relative frequency distribution and the sampling distribution for $\bar{y}$



(a) Population relative frequency distribution



(b) Sampling distribution of $\bar{y}$

(b) The probability that $\bar{y}$ will exceed 82 is equal to the highlighted area in Figure 1.15. To find this area, we need to find the z-value corresponding to $\bar{y} = 82$. Recall that the standard normal random variable z is the difference between any normally distributed random variable and its mean, expressed in units of its standard deviation. Since $\bar{y}$ is a normally distributed random variable with mean $\mu_{\bar{y}} = \mu$ and standard deviation $\sigma_{\bar{y}} = \sigma/\sqrt{n}$, it follows that the standard normal z-value corresponding to the sample mean, $\bar{y}$, is

$$z = \frac{\text{(Normal random variable)} - \text{(Mean)}}{\text{Standard Deviation}} = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}}$$

Therefore, for $\bar{y} = 82$, we have

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}} = \frac{82 - 80}{1} = 2$$

The area A in Figure 1.16 corresponding to $z = 2$ is given in the table of areas under the normal curve (see Table 1 of Appendix C) as .4772. Therefore, the tail area corresponding to the probability that $\bar{y}$ exceeds 82 is

$$P(\bar{y} > 82) = P(z > 2) = .5 - .4772 = .0228$$

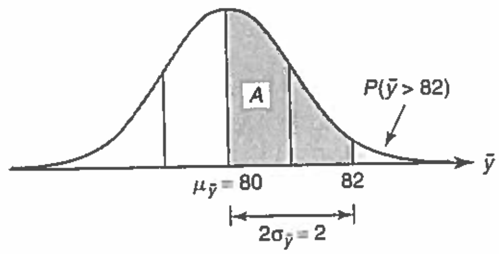


Figure 1.16 The sampling distribution of $\bar{y}$

The central limit theorem can also be used to justify the fact that the *sum* of the sample measurements possesses a sampling distribution that is approximately normal for large sample sizes. In fact, since many statistics are obtained by summing or averaging random quantities, the central limit theorem helps to explain why many statistics have mound-shaped (or approximately normal) sampling distributions.

As we proceed, we encounter many different sample statistics, and we need to know their sampling distributions to evaluate the reliability of each one for making inferences. These sampling distributions are described as the need arises.

1.8 Estimating a Population Mean

We can make an inference about a population parameter in two ways:

1. Estimate its value.
2. Make a decision about its value (i.e., test a hypothesis about its value).

In this section, we illustrate the concepts involved in estimation, using the estimation of a population mean as an example. Tests of hypotheses will be discussed in Section 1.9.

To estimate a population parameter, we choose a sample statistic that has two desirable properties: (1) a sampling distribution that centers about the parameter and (2) a small standard error. If the mean of the sampling distribution of a statistic equals the parameter we are estimating, we say that the statistic is an **unbiased estimator** of the parameter. If not, we say that it is **biased**.

In Section 1.7, we noted that the sampling distribution of the sample mean is approximately normally distributed for moderate to large sample sizes and that it possesses a mean μ and standard error $\sigma/\sqrt{n}$. Therefore, as shown in Figure 1.17,

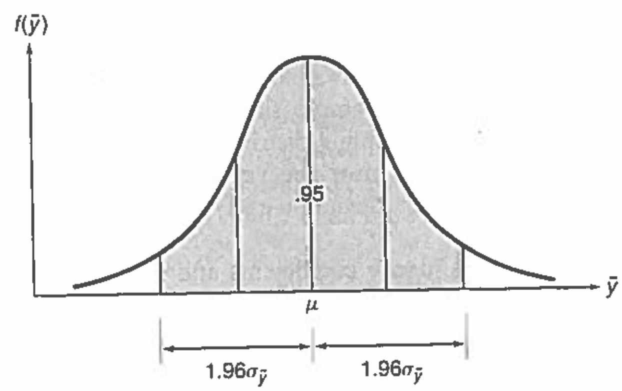


Figure 1.17 Sampling distribution of $\bar{y}$

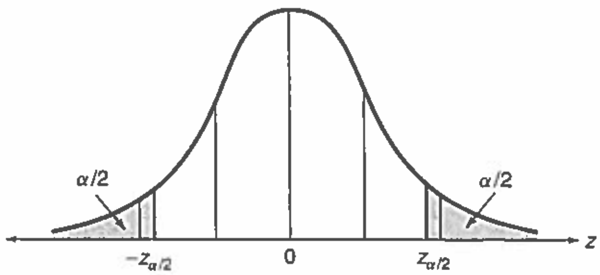


Figure 1.18 Locating $z_{\alpha/2}$ on the standard normal curve

$\bar{y}$ is an unbiased estimator of the population mean μ, and the probability that $\bar{y}$ will fall within $1.96\sigma_{\bar{y}} = 1.96\sigma/\sqrt{n}$ of the true value of μ is approximately .95.*

Since $\bar{y}$ will fall within $1.96\sigma_{\bar{y}}$ of μ approximately 95% of the time, it follows that the interval

$$\bar{y} - 1.96\sigma_{\bar{y}} \quad \text{to} \quad \bar{y} + 1.96\sigma_{\bar{y}}$$

will enclose μ approximately 95% of the time in repeated sampling. This interval is called a 95% **confidence interval**, and .95 is called the **confidence coefficient**.

Notice that μ is fixed and that the confidence interval changes from sample to sample. The probability that a confidence interval calculated using the formula

$$\bar{y} \pm 1.96\sigma_{\bar{y}}$$

will enclose μ is approximately .95. Thus, the confidence coefficient measures the confidence that we can place in a particular confidence interval.

Confidence intervals can be constructed using any desired confidence coefficient. For example, if we define $z_{\alpha/2}$ to be the value of a standard normal variable that places the area $\alpha/2$ in the right tail of the z distribution (see Figure 1.18), then a $100(1-\alpha)\%$ confidence interval for μ is given in the box.

Large-Sample $100(1-\alpha)\%$ Confidence Interval for μ

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

where $z_{\alpha/2}$ is the z-value with an area $\alpha/2$ to its right (see Figure 1.18) and $\sigma_{\bar{y}} = \sigma/\sqrt{n}$. The parameter σ is the standard deviation of the sampled population and n is the sample size. If σ is unknown, its value may be approximated by the sample standard deviation s. The approximation is valid for large samples (e.g., $n \geq 30$) only.

The confidence interval shown in the box is called a large-sample confidence interval because the sample size must be large enough to ensure approximate normality for the sampling distribution of $\bar{y}$. Also, and even more important, you will rarely, if ever, know the value of σ, so its value must be estimated using the sample standard deviation s. This approximation for σ will be adequate only when $n \geq 30$.

Typical confidence coefficients and corresponding values of $z_{\alpha/2}$ are shown in Table 1.8.

* Additionally, $\bar{y}$ has the smallest standard error among all unbiased estimators of μ. Consequently, we say that $\bar{y}$ is the **minimum variance unbiased estimator** (MVUE) for μ.

Table 1.8 Commonly used values of $z_{\alpha/2}$

Confidence Coefficient $(1-\alpha)$	α	$\alpha/2$	$z_{\alpha/2}$
.90	.10	.05	1.645
.95	.05	.025	1.96
.99	.01	.005	2.576

Example 1.11

Psychologists have found that twins, in their early years, tend to have lower intelligence quotients and pick up language more slowly than nontwins (*Wisconsin Twin Research Newsletter*, Winter 2004). The slower intellectual growth of twins may be caused by benign parental neglect. Suppose we want to investigate this phenomenon. A random sample of $n = 50$ sets of $2\frac{1}{2}$-year-old twin boys is selected, and the total parental attention time given to each pair during 1 week is recorded. The data (in hours) are listed in Table 1.9. Estimate μ, the mean attention time given to all $2\frac{1}{2}$-year-old twin boys by their parents, using a 99% confidence interval. Interpret the interval in terms of the problem.

ATTENTIMES

Table 1.9 Attention time for a random sample of $n = 50$ sets of twins

20.7	14.0	16.7	20.7	22.5	48.2	12.1	7.7	2.9	22.2
23.5	20.3	6.4	34.0	1.3	44.5	39.6	23.8	35.6	20.0
10.9	43.1	7.1	14.3	46.0	21.9	23.4	17.5	29.4	9.6
44.1	36.4	13.8	0.8	24.3	1.1	9.3	19.3	3.4	14.6
15.7	32.5	46.6	19.1	10.6	36.9	6.7	27.9	5.4	14.0

Solution

The general form of the 99% confidence interval for a population mean is

$$\bar{y} \pm z_{\alpha/2}\sigma_{\bar{y}} = \bar{y} \pm z_{.005}\sigma_{\bar{y}}.$$
$$= \bar{y} \pm 2.576\left(\frac{\sigma}{\sqrt{n}}\right)$$

A SAS printout showing descriptive statistics for the sample of $n = 50$ attention times is displayed in Figure 1.19. The values of $\bar{y}$ and s, shaded on the printout, are $\bar{y} = 20.85$ and $s = 13.41$. Thus, for the 50 twins sampled, the 99% confidence interval is

$$20.85 \pm 2.576\left(\frac{\sigma}{\sqrt{50}}\right)$$

We do not know the value of σ (the standard deviation of the weekly attention time given to $2\frac{1}{2}$-year-old twin boys by their parents), so we use our best approximation, the sample standard deviation s. (Since the sample size, $n = 50$, is large, the approximation is valid.) Then the 99% confidence interval is

$$20.85 \pm 2.576\left(\frac{13.41}{\sqrt{50}}\right) = 20.85 \pm 4.89$$

or, from 15.96 to 25.74. That is, we can be 99% confident that the true mean weekly attention given to $2\frac{1}{2}$-year-old twin boys by their parents falls between 15.96 and

Figure 1.19 SAS descriptive statistics for $n = 50$ sample attention times

```
Sample Statistics for ATTIME

      N          Mean        Std. Dev.       Std. Error
   -----------------------------------------------------
     50         20.85         13.41             1.90

Hypothesis Test

   Null hypothesis:     Mean of ATTIME =  0
   Alternative:         Mean of ATTIME ^= 0

   With a specified known standard deviation of 13.41

               Z Statistic     Prob > Z
               -----------     --------
                 10.993        <.0001

99% Confidence Interval for the Mean

               Lower Limit    Upper Limit
               -----------    -----------
                  15.96          25.73
```

25.74 hours. [Note: This interval is also shown at the bottom of the SAS printout, Figure 1.19.] ■

The large-sample method for making inferences about a population mean μ assumes that either σ is known or the sample size is large enough ($n \geq 30$) for the sample standard deviation s to be used as a good approximation to σ. The technique for finding a $100(1-\alpha)\%$ confidence interval for a population mean μ for small sample sizes requires that the sampled population have a normal probability distribution. The formula, which is similar to the one for a large-sample confidence interval for μ, is

$$\bar{y} \pm t_{\alpha/2} s_{\bar{y}}$$

where $s_{\bar{y}} = s/\sqrt{n}$ is the estimated standard error of $\bar{y}$. The quantity $t_{\alpha/2}$ is directly analogous to the standard normal value $z_{\alpha/2}$ used in finding a large-sample confidence interval for μ except that it is an upper-tail t-value obtained from a student's t distribution. Thus, $t_{\alpha/2}$ is an upper-tail t-value such that an area $\alpha/2$ lies to its right.

Like the standardized normal (z) distribution, a Student's t distribution is symmetric about the value $t = 0$, but it is more variable than a z distribution. The variability depends on the number of **degrees of freedom, df**, which in turn depends on the number of measurements available for estimating σ^2. The smaller the number of degrees of freedom, the greater will be the spread of the t distribution. For this application of a Student's t distribution, df $= n - 1$.[†] As the sample size increases (and df increases), the Student's t distribution looks more and more like a z distribution, and for $n \geq 30$, the two distributions will be nearly identical. A Student's t distribution based on df $= 4$ and a standard normal distribution are shown in Figure 1.20. Note the corresponding values of $z_{.025}$ and $t_{.025}$.

The upper-tail values of the Student's t distribution are given in Table 2 of Appendix D. An abbreviated version of the t table is presented in Table 1.10. To find the upper-tail t-value based on 4 df that places .025 in the upper tail of the t distribution, we look in the row of the table corresponding to df $= 4$ and the column corresponding to $t_{.025}$. The t-value is 2.776 and is shown in Figure 1.17.

The process of finding a small-sample confidence interval for μ is given in the next box.

† Think of df as the amount of information in the sample size n for estimating μ. We lose 1 df for estimating μ; hence df $= n - 1$.

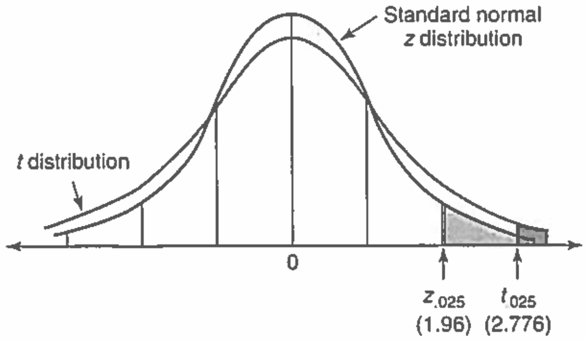
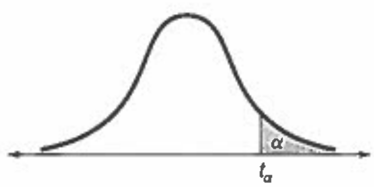


Figure 1.20 The $t_{.025}$ value in a t distribution with 4 df and the corresponding $z_{.025}$ value

Table 1.10 Reproduction of a portion of Table 2 of Appendix D



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

Small-Sample Confidence Interval for μ

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

where $s_{\bar{y}} = s/\sqrt{n}$ and $t_{\alpha/2}$ is a t-value based on $(n-1)$ degrees of freedom, such that the probability that $t > t_{\alpha/2}$ is $\alpha/2$.

Assumptions: The relative frequency distribution of the sampled population is approximately normal.

Example 1.12

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. Researchers (*Geothermics*, August 2002) 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:

SILICA

229	255	280	203	229

Estimate the mean amount of silicon dioxide present in the five antiscalant solutions. Use a 95% 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 provided in the MINITAB printout, Figure 1.21.

For a confidence coefficient of $1 - \alpha = .95$, we have $\alpha = .05$ and $\alpha/2 = .025$. 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 five 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

$$\bar{y} \pm t_{\alpha/2}(s_{\bar{y}}) = \bar{y} \pm t_{.025}\left(\frac{s}{\sqrt{n}}\right)$$
$$= 239.2 \pm t_{.025}\left(\frac{29.3}{\sqrt{5}}\right)$$

where $t_{.025}$ is the value corresponding to an upper-tail area of .025 in the Student's t distribution based on $(n - 1) = 4$ degrees of freedom. From Table 2 of Appendix D, the required t-value (shaded in Table 1.10) is $t_{.025} = 2.776$. Substituting this value yields

$$239.2 \pm t_{.025}\left(\frac{29.3}{\sqrt{5}}\right) = 239.2 \pm (2.776)\left(\frac{29.3}{\sqrt{5}}\right)$$
$$= 239.2 \pm 36.4$$

or 202.8–275.6 ppm.

Thus, if the distribution of silicon dioxide amounts is approximately normal, then we can be 95% confident that the interval (202.8–275.6) encloses μ, the true mean amount of silicon dioxide present in an antiscalant solution. Remember, the 95% confidence level implies that if we were to employ our interval estimator on repeated occasions, 95% of the intervals constructed would capture μ.

Figure 1.21 MINITAB descriptive statistics and confidence interval for Example 1.12

One-Sample T: PPM

Variable	N	Mean	StDev	SE Mean	95.0% CI
PPM	5	239.2	29.3	13.1	(202.8, 275.6)

The 95% confidence interval can also be obtained with statistical software. This interval is shaded on the MINITAB printout, Figure 1.21. You can see that the computer-generated interval is identical to our calculated one. ∎

Example 1.13 Suppose you want to reduce the width of the confidence interval obtained in Example 1.12. Specifically, you want to estimate the mean silicon dioxide content of an aliquot of brine correct to within 10 ppm with confidence coefficient approximately equal to .95. How many aliquots of brine would you have to include in your sample?

Solution

We will interpret the phrase, "correct to within 10 ppm ... equal to .95" to mean that we want half the width of a 95% confidence interval for μ to equal 10 ppm. That is, we want

$$t_{.025}\left(\frac{s}{\sqrt{n}}\right) = 10$$

To solve this equation for n, we need approximate values for $t_{.025}$ and s. Since we know from Example 1.12 that the confidence interval was wider than desired for $n = 5$, it is clear that our sample size must be larger than 5. Consequently, $t_{.025}$ will be very close to 2, and this value will provide a good approximation to $t_{.025}$. A good measure of the data variation is given by the standard deviation computed in Example 1.12. We substitute $t_{.025} \approx 2$ and $s \approx 29.3$ into the equation and solve for n:

$$t_{.025}\left(\frac{s}{\sqrt{n}}\right) = 10$$

$$2\left(\frac{29.3}{\sqrt{n}}\right) = 10$$

$$\sqrt{n} = 5.86$$

$$n = 34.3 \quad \text{or approximately } n = 34$$

Remember that this sample size is an approximate solution because we approximated the value of $t_{.025}$ and the value of s that might be computed from the prospective data. Nevertheless, $n = 34$ will be reasonably close to the sample size needed to estimate the mean silicon dioxide content correct to within 10 ppm. ∎

Important Note: Theoretically, the small-sample t procedure presented here requires that the sample data come from a population that is normally distributed. (See the assumption in the box, p. 37.) However, statisticians have found the one-sample t procedure to be **robust**, that is, to yield valid results even when the data are nonnormal, as long as the population is not highly skewed.

1.8 Exercises

1.40 Simulating a sampling distribution. The next table (p. 40) contains 50 random samples of random digits, $y = 0, 1, 2, 3, \ldots, 9$, where the probabilities corresponding to the values of y are given by the formula $p(y) = \frac{1}{10}$. Each sample contains $n = 6$ measurements.

(a) Use the 300 random digits to construct a relative frequency distribution for the data. This relative frequency distribution should approximate $p(y)$.

(b) Calculate the mean of the 300 digits. This will give an accurate estimate of μ (the mean of the population) and should be very near to $E(y)$, which is 4.5.

(c) Calculate s^2 for the 300 digits. This should be close to the variance of y, $\sigma^2 = 8.25$.

(d) Calculate $\bar{y}$ for each of the 50 samples. Construct a relative frequency distribution for the sample means to see how close they lie to the mean of $\mu = 4.5$. Calculate the mean and standard deviation of the 50 means.

EX1_40

SAMPLE	SAMPLE	SAMPLE	SAMPLE
8, 1, 8, 0, 6, 6	7, 6, 7, 0, 4, 3	4, 4, 5, 2, 6, 6	0, 8, 4, 7, 6, 9
7, 2, 1, 7, 2, 9	1, 0, 5, 9, 9, 6	2, 9, 3, 7, 1, 3	5, 6, 9, 4, 4, 2
7, 4, 5, 7, 7, 1	2, 4, 4, 7, 5, 6	5, 1, 9, 6, 9, 2	4, 2, 3, 7, 6, 3
8, 3, 6, 1, 8, 1	4, 6, 6, 5, 5, 6	8, 5, 1, 2, 3, 4	1, 2, 0, 6, 3, 3
0, 9, 8, 6, 2, 9	1, 5, 0, 6, 6, 5	2, 4, 5, 3, 4, 8	1, 1, 9, 0, 3, 2
0, 6, 8, 8, 3, 5	3, 3, 0, 4, 9, 6	1, 5, 6, 7, 8, 2	7, 8, 9, 2, 7, 0
7, 9, 5, 7, 7, 9	9, 3, 0, 7, 4, 1	3, 3, 8, 6, 0, 1	1, 1, 5, 0, 5, 1
7, 7, 6, 4, 4, 7	5, 3, 6, 4, 2, 0	3, 1, 4, 4, 9, 0	7, 7, 8, 7, 7, 6
1, 6, 5, 6, 4, 2	7, 1, 5, 0, 5, 8	9, 7, 7, 9, 8, 1	4, 9, 3, 7, 3, 9
9, 8, 6, 8, 6, 0	4, 4, 6, 2, 6, 2	6, 9, 2, 9, 8, 7	5, 5, 1, 1, 4, 0
3, 1, 6, 0, 0, 9	3, 1, 8, 8, 2, 1	6, 6, 8, 9, 6, 0	4, 2, 5, 7, 7, 9
0, 6, 8, 5, 2, 8	8, 9, 0, 6, 1, 7	3, 3, 4, 6, 7, 0	8, 3, 0, 6, 9, 7
8, 2, 4, 9, 4, 6	1, 3, 7, 3, 4, 3		

1.41 Effect of n on the standard deviation. Refer to Exercise 1.40. To see the effect of sample size on the standard deviation of the sampling distribution of a statistic, combine pairs of samples (moving down the columns of the table) to obtain 25 samples of $n = 12$ measurements. Calculate the mean for each sample.

(a) Construct a relative frequency distribution for the 25 means. Compare this with the distribution prepared for Exercise 1.40 that is based on samples of $n = 6$ digits.

(b) Calculate the mean and standard deviation of the 25 means. Compare the standard deviation of this sampling distribution with the standard deviation of the sampling distribution in Exercise 1.40. What relationship would you expect to exist between the two standard deviations?

1.42 Using Table 2. Let t_0 be a particular value of t. Use Table 2 of Appendix D to find t_0 values such that the following statements are true:

(a) $P(t \geq t_0) = .025$ where $n = 10$
(b) $P(t \geq t_0) = .01$ where $n = 5$
(c) $P(t \leq t_0) = .005$ where $n = 20$
(d) $P(t \leq t_0) = .05$ where $n = 12$

1.43 Critical part failures in NASCAR vehicles. *The Sport Journal* (Winter 2007) published an analysis of critical part failures at NASCAR races. The researchers found that the time y (in hours) until the first critical part failure has a highly skewed distribution with $\mu = .10$ and $\sigma = .10$. Now, consider a random sample of $n = 50$ NASCAR races and let $\bar{y}$ represent the sample mean time until the first critical part failure.

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

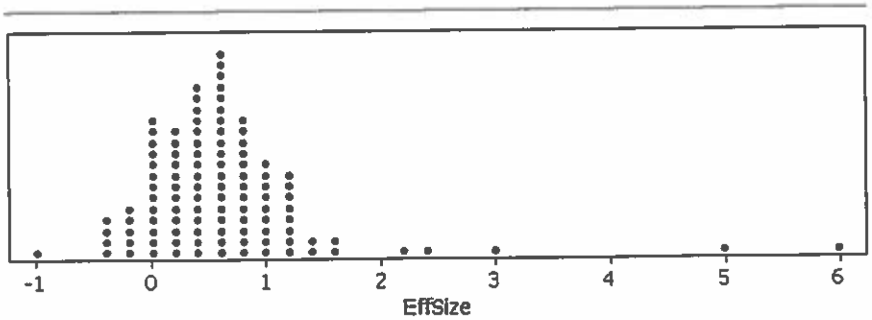
(b) Although y has a highly skewed distribution, the sampling distribution of $\bar{y}$ is approximately normal. Why?

(c) Find the probability that the sample mean time until the first critical part failure exceeds .13 hour.

PERAGGR

1.44 Personality and aggressive behavior. How does personality impact aggressive behavior? A team of university psychologists conducted a review of studies that examined the relationship between personality and aggressive behavior (*Psychological Bulletin*, Vol. 132, 2006). One variable of interest to the researchers was the difference between the aggressive behavior level of individuals in the study who scored high on a personality test and those who scored low on the test. This variable, standardized to be between −7 and 7, was called "effect size." (A large positive effect size indicates that those who score high on the personality test are more aggressive than those who score low.) The researchers collected the effect sizes for a sample of $n = 109$ studies published in

Minitab Output for Exercise 1.44




```
Variable    N    Mean    StDev   SE Mean         95% CI
EffSize   109  0.6477   0.8906   0.0853   (0.4786, 0.8167)
```

psychological journals. This data are saved in the PERAGGR file. A dot plot and summary statistics for effect size are shown in the accompanying MINITAB printouts (bottom of p. 40). Of interest to the researchers is the true mean effect size μ for all psychological studies of personality and aggressive behavior.

(a) Identify the parameter of interest to the researchers.
(b) Examine the dot plot. Does effect size have a normal distribution? Explain why your answer is irrelevant to the subsequent analysis.
(c) Locate and interpret a 95% confidence interval for μ on the accompanying printout.
(d) If the true mean effect size exceeds 0, then the researchers will conclude that in the population, those who score high on a personality test are more aggressive than those who score low. Can the researchers draw this conclusion? Explain.

1.45 **Chicken pecking experiment.** Animal behaviorists have discovered that the more domestic chickens peck at objects placed in their environment, the healthier the chickens seem to be. White string has been found to be a particularly attractive pecking stimulus. In one experiment, 72 chickens were exposed to a string stimulus. Instead of white string, blue-colored string was used. The number of pecks each chicken took at the blue string over a specified time interval was recorded. Summary statistics for the 72 chickens were $\bar{y} = 1.13$ pecks, $s = 2.21$ pecks (*Applied Animal Behaviour Science*, October 2000).

(a) Estimate the population mean number of pecks made by chickens pecking at blue string using a 99% confidence interval. Interpret the result.
(b) Previous research has shown that $\mu = 7.5$ pecks if chickens are exposed to white string. Based on the results, part a, is there evidence that chickens are more apt to peck at white string than blue string? Explain.

1.46 **Impact of cooking on air particles.** A group of Harvard University School of Public Health researchers studied the impact of cooking on the size of indoor air particles (*Environmental Science and Technology*, September 1, 2000). The decay rate (measured as μ m/hour) for fine particles produced from oven cooking or toasting was recorded on six randomly selected days. These six measurements are shown in the table.

(a) Find and interpret a 95% confidence interval for the true average decay rate of fine particles produced from oven cooking or toasting.
(b) Explain what the phrase "95% confident" implies in the interpretation of part a.
(c) What must be true about the distribution of the population of decay rates for the inference to be valid?

DECAY

.95	.83	1.20	.89	1.45	1.12

Source: Abt, E., et al. "Relative contribution of outdoor and indoor particle sources to indoor concentrations," *Environmental Science and Technology*, Vol. 34, No. 17, Sept. 1, 2000 (Table 3). Reprinted with permission from *Environmental Science.*

1.47 **Accounting and Machiavellianism.** Refer to the *Behavioral Research in Accounting* (January 2008) study of Machiavellian traits in accountants, Exercise 1.6 (p. 4). Recall that *Machiavellian* describes negative character traits that include manipulation, cunning, duplicity, deception, and bad faith. A Machiavellian ("Mach") rating score was determined for each in a sample of accounting alumni of a large southwestern university. Scores range from a low of 40 to a high of 160, with the theoretical neutral Mach rating score of 100. The 122 purchasing managers in the sample had a mean Mach rating score of 99.6, with a standard deviation of 12.6.

(a) From the sample, estimate the true mean Mach rating score of all purchasing managers.
(b) Form a 95% confidence interval for the estimate, part b.
(c) Give a practical interpretation of the interval, part c.
(d) A director of purchasing at a major firm claims that the true mean Mach rating score of all purchasing managers is 85. Is there evidence to dispute this claim?

1.48 **Wearout of used display panels.** Researchers presented a study of the wearout failure time of used colored display panels purchased by an outlet store (*Mathematical Sciences Colloquium*, December 2001). Prior to acquisition, the panels had been used for about one-third of their expected lifetimes. The failure times (in years) for a sample of 50 used panels are reproduced in the next table, followed by an SPSS printout of the analysis of the data.

(a) Locate a 95% confidence interval for the true mean failure time of used colored display panels on the printout.
(b) Give a practical interpretation of the interval, part a.
(c) In repeated sampling of the population of used colored display panels, where a 95% confidence interval for the mean failure time is computed for each sample, what proportion of all the confidence intervals generated will capture the true mean failure time?

PANELFAIL

0.01	1.21	1.71	2.30	2.96	0.19	1.22	1.75	2.30	2.98	0.51
1.24	1.77	2.41	3.19	0.57	1.48	1.79	2.44	3.25	0.70	1.54
1.88	2.57	3.31	0.73	1.59	1.90	2.61	1.19	0.75	1.61	1.93
2.62	3.50	0.75	1.61	2.01	2.72	3.50	1.11	1.62	2.16	2.76
3.50	1.16	1.62	2.18	2.84	3.50					

Source: Irony, T. Z., Lauretto, M., Pereira, C., and Stern, J. M. "A Weibull wearout test: Full Bayesian approach," paper presented at *Mathematical Sciences Colloquium*, Binghamton University, Binghamton, U.K., December 2001.

Descriptives

			Statistic	Std. Error
FAILTIME	Mean		1.9350	.13133
	95% Confidence Interval for Mean	Lower Bound	1.6711	
		Upper Bound	2.1989	
	5% Trimmed Mean		1.9454	
	Median		1.8350	
	Variance		.862	
	Std. Deviation		.92865	
	Minimum		.01	
	Maximum		3.50	
	Range		3.49	
	Interquartile Range		1.43	
	Skewness		-.008	.337
	Kurtosis		-.755	.662

1.49 Studies on treating Alzheimer's disease. Alzheimer's disease (AD) is a progressive disease of the brain. Much research has been conducted on how to treat AD. The journal *eCAM* (November 2006) published an article that critiqued the quality of the methodology used in studies on AD treatment. For each in a sample of 13 studies, the quality of the methodology was measured using the Wong Scale, with scores ranging from 9 (low quality) to 27 (high quality). The data are shown in the table below. Estimate the mean quality, μ, of all studies on the treatment of Alzheimer's disease with a 99% confidence interval. Interpret the result.

TREATAD

22	21	18	19	20	15	19	20	15	20	17	20	21

Source: Chiappelli, F., et al. "Evidence-based research in complementary and alternative medicine III: Treatment of patients with Alzheimer's disease," *eCAM*, Vol. 3, No. 4, Nov. 2006 (Table 1).

1.50 Reproduction of bacteria-infected spider mites. Zoologists in Japan investigated the reproductive traits of spider mites with a bacteria infection (*Heredity*, January 2007). Male and female pairs of infected spider mites were mated in a laboratory, and the number of eggs produced by each female recorded. Summary statistics for several samples are provided in the accompanying table. Note that in some samples, one or both infected spider mites were treated with an antibiotic prior to mating.

(a) For each female/male pair type, construct and interpret a 90% confidence interval for the population mean number of eggs produced by the female spider mite.
(b) Identify the female/male pair type that appears to produce the highest mean number of eggs.

FEMALE/ MALE PAIRS	SAMPLE SIZE	MEAN # OF EGGS	STANDARD DEVIATION
Both untreated	29	20.9	3.34
Male treated	23	20.3	3.50
Female treated	18	22.9	4.37
Both treated	21	18.6	2.11

Source: Reprinted by permission from Macmillan Publishers Ltd: *Heredity* (Gotoh, T., Noda, H., and Ito, S. "Cardinium symbionts cause cytoplasmic incompatibility in spider mites," Vol. 98, No. 1, Jan. 2007, Table 2). Copyright © 2007.

1.9 Testing a Hypothesis About a Population Mean

The procedure involved in testing a hypothesis about a population parameter can be illustrated with the procedure for a test concerning a population mean μ.

A statistical test of a hypothesis is composed of several elements, as listed in the box.

Elements of a Statistical Test of Hypothesis

1. **Null Hypothesis** (denoted H_0): This is the hypothesis that is postulated to be true.
2. **Alternative Hypothesis** (denoted H_a): This hypothesis is counter to the null hypothesis and is usually the hypothesis that the researcher wants to support.
3. **Test Statistic:** Calculated from the sample data, this statistic functions as a decision-maker.
4. **Level of significance** (denoted α): This is the probability of a *Type I error* (i.e., the probability of rejecting H_0 given that H_0 is true).
5. **Rejection Region:** Values of the test statistic that lead the researcher to reject H_0 and accept H_a.
6. ***p*-Value:** Also called the *observed significance level*, this is the probability of observing a value of the test statistic at least as contradictory to the null hypothesis as the observed test statistic value, assuming the null hypothesis is true.
7. **Conclusion:** The decision to reject or accept H_0 based on the value of the test statistic, α, the rejection region, and/or the p-value.

The test statistic for testing the null hypothesis that a population mean μ equals some specific value, say, μ_0, is the sample mean $\bar{y}$ or the standardized normal variable

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}} \quad \text{where } \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$

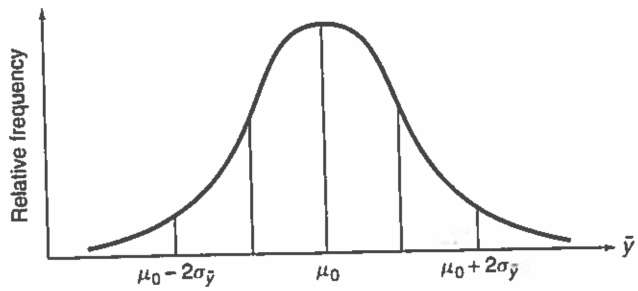
The logic used to decide whether sample data *disagree* with this hypothesis can be seen in the sampling distribution of $\bar{y}$ shown in Figure 1.22. If the population mean μ is equal to μ_0 (i.e., if the null hypothesis is true), then the mean $\bar{y}$ calculated from a sample should fall, with high probability, within $2\sigma_{\bar{y}}$ of μ_0. If $\bar{y}$ falls too far away from μ_0, or if the standardized distance

$$z = \frac{\bar{y} - \mu_0}{\sigma_{\bar{y}}}$$

is too large, we conclude that the data disagree with our hypothesis, and we reject the null hypothesis.

The reliability of a statistical test is measured by the probability of making an incorrect decision. This probability, or *level of significance*, is (as stated in the box) α = P(Type I error) = P(Reject H_0 given H_0 is true). Prior to conducting the test, the researcher selects a value of α (e.g., $\alpha = .05$). This value is then used to find the appropriate rejection region.

Figure 1.22 The sampling distribution of $\bar{y}$ for $\mu = \mu_0$



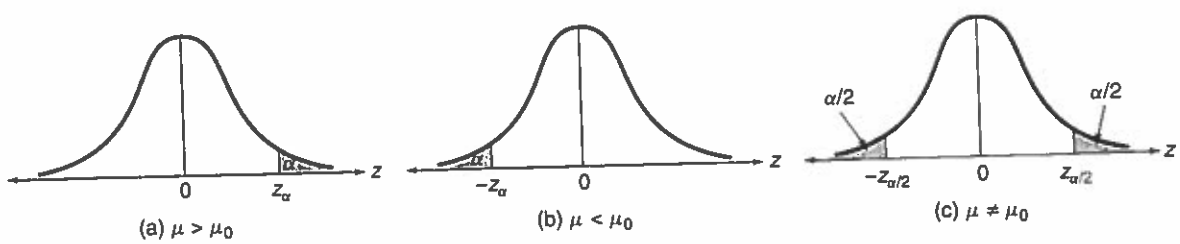


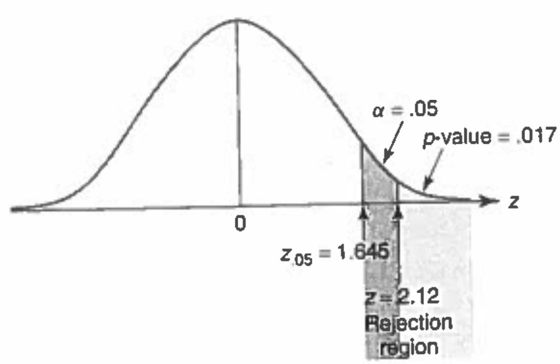
Figure 1.23 Location of the rejection region for various alternative hypotheses

For example, if we want to test the null hypothesis H_0: $\mu = \mu_0$ against the alternative hypothesis H_a: $\mu > \mu_0$, we locate the boundary of the rejection region in the upper tail of the z distribution, as shown in Figure 1.23(a), at the point z_α. Note that α is the tail probability in Figure 1.23(a). We will reject H_0 if $z > z_\alpha$. Similarly, to test H_0: $\mu = \mu_0$ against H_a: $\mu < \mu_0$, we place the rejection region in the lower tail of the z distribution, as shown in Figure 1.23(b). These tests are called **one-tailed** (or **one-sided**) **statistical tests.** To detect either $\mu < \mu_0$ or $\mu > \mu_0$, that is, to test H_a: $\mu \neq \mu_0$, we split α equally between the two tails of the z distribution and reject the null hypothesis if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$, as shown in Figure 1.23(c). This test is called a **two-tailed** (or **two-sided**) **statistical test.**

As an alternative to the rejection region approach, many researchers utilize the *p-value* to conduct the test. As stated above, the p-value is the probability of observing a value of the test statistic at least as contradictory to H_0 as the observed value of the test statistic, *assuming H_0 is true.* For example, if the test statistic for testing H_a: $\mu > \mu_0$ is $z = 2.12$, then the p-value is

$$p\text{-value} = P(z > 2.12) = .0170 \text{ (obtained from Table 1, Appendix D)}$$

Figure 1.24 Testing H_a: $\mu > \mu_0$ using a p-value



The researcher makes decisions about H_0 and H_a by comparing the p-value of the test to the selected value of α. If $\alpha > p$-value, we reject H_0 in favor of H_a. Figure 1.24 shows the p-value and rejection region at $\alpha = .05$ for testing H_a: $\mu > \mu_0$. Note that since $\alpha = .05$ exceeds p-value $= .0170$, we reject H_0. Also, the test statistic, $z = 2.12$, falls into the rejection region. Consequently, the two decision-making rules are equivalent. However, since statistical software automatically produces the p-value for a test of hypothesis, many researchers prefer the p-value approach.

The z test, summarized in the next box, is called a *large-sample test* because we will rarely know σ and hence will need a sample size that is large enough so that the sample standard deviation s will provide a good approximation to σ. Normally, we recommend that the sample size be $n \geq 30$.

Large-Sample ($n \geq 30$) Test of Hypothesis About μ

Test statistic: $z = (\bar{y} - \mu_0)/\sigma_{\bar{y}} \approx (\bar{y} - \mu_0)/(s/\sqrt{n})$

	ONE-TAILED TESTS		TWO-TAILED TEST
	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
	H_a: $\mu < \mu_0$	H_a: $\mu > \mu_0$	H_a: $\mu \neq \mu_0$
Rejection region:	$z < -z_\alpha$	$z > z_\alpha$	$\|z\| > z_{\alpha/2}$
p-*value*:	$P(z < z_c)$	$P(z > z_c)$	$2P(z > z_c)$ if z_c is positve $2P(z < z_c)$ if z_c is negative

Decision: Reject H_0 if $\alpha > p$-value, or if test statistic falls in rejection region

where $P(z > z_\alpha) = \alpha$, $P(z > z_{\alpha/2}) = \alpha/2$, z_c = calculated value of the test statistic, and α = P(Type I error) = P(Reject H_0| H_0 true).

We illustrate with an example.

Example 1.14

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 flourished. (Assume that the unearthed bones were all from the same unknown species.) The length-to-width ratios of the bones were measured and are listed in Table 1.11. Do these data present sufficient evidence to indicate that the mean ratio of all bones of this species differs from 8.5? Use $\alpha = .05$.

Solution

Since we wish to determine whether $\mu \neq 8.5$, the elements of the test are

H_0: $\mu = 8.5$

H_a: $\mu \neq 8.5$

$$\textit{Test statistic}: z = \frac{\bar{y} - 8.5}{\sigma_{\bar{y}}} = \frac{\bar{y} - 8.5}{\sigma/\sqrt{n}} \approx \frac{\bar{y} - 8.5}{s/\sqrt{n}}$$

Rejection region: $|z| > 1.96$ for $\alpha = .05$

The data in Table 1.11 were analyzed using SPSS. The SPSS printout is displayed in Figure 1.25.

BONES

Table 1.11 Length-to-width ratios of a sample of humerus bones

10.73	9.57	6.66	9.89
8.89	9.29	9.35	8.17
9.07	9.94	8.86	8.93
9.20	8.07	9.93	8.80
10.33	8.37	8.91	10.02
9.98	6.85	11.77	8.38
9.84	8.52	10.48	11.67
9.59	8.87	10.39	8.30
8.48	6.23	9.39	9.17
8.71	9.41	9.17	12.00
			9.38

Figure 1.25 SPSS printout for Example 1.14

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
LWRATIO	41	9.2576	1.20357	.18797

One-Sample Test

	Test Value = 8.5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
LWRATIO	4.030	40	.000	.7576	.3777	1.1375

Substituting the sample statistics $\bar{y} = 9.26$ and $s = 1.20$ (shown at the top of the SPSS printout) into the test statistic, we have

$$z \approx \frac{\bar{y} - 8.5}{s/\sqrt{n}} = \frac{9.26 - 8.5}{1.20/\sqrt{41}} = 4.03$$

For this two-tailed test, we also find

$$p\text{-value} = 2P(z > 4.03) \approx 0$$

Both the test statistic and p-value are shown (highlighted) at the bottom of the SPSS printout. Since the test statistic exceeds the critical value of 1.96 (or, since $\alpha = .05$ exceeds the p-value), we can reject H_0 at $\alpha = .05$. The sample data provide sufficient evidence to conclude that the true mean length-to-width ratio of all humerus bones of this species differs from 8.5.

The *practical* implications of the result obtained in Example 1.14 remain to be seen. Perhaps the animal discovered at the archeological site is of some species other than A. Alternatively, the unearthed humeri may have larger than normal length-to-width ratios because they are the bones of specimens having unusual feeding habits for 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, knowledge of the subject matter and the phenomenon under investigation. ■

Note: Earlier, we discussed the use of $\alpha = P(\text{Type I error})$ as a measure of reliability for the statistical test. A second type of error could be made if we accepted the null hypothesis when, in fact, the alternative hypothesis is true (a **Type II error**). As a general rule, you should never "accept" the null hypothesis unless you know the probability of making a Type II error. Since this probability (denoted by the symbol β) is often unknown, it is a common practice to defer judgment if a test statistic falls in the nonrejection region. That is, we say "fail to reject H_0" rather than "accept H_0."

A small-sample test of the null hypothesis $\mu = \mu_0$ using a Student's t statistic is based on the assumption that the sample was randomly selected from a population with a normal relative frequency distribution. The test is conducted in exactly the same manner as the large-sample z test except that we use

$$t = \frac{\bar{y} - \mu_0}{s_{\bar{y}}} = \frac{\bar{y} - \mu_0}{s/\sqrt{n}}$$

as the test statistic and we locate the rejection region in the tail(s) of a Student's t distribution with $\text{df} = n - 1$. We summarize the technique for conducting a small-sample test of hypothesis about a population mean in the box.

Small-Sample Test of Hypothesis About μ

Test statistic: $t = (\bar{y} - \mu_0)/(s/\sqrt{n})$

	ONE-TAILED TESTS		TWO-TAILED TEST
	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
	H_a: $\mu < \mu_0$	H_a: $\mu > \mu_0$	H_a: $\mu \neq \mu_0$
Rejection region:	$t < -t_\alpha$	$t > t_\alpha$	$\lvert t\rvert > t_{\alpha/2}$
p-*value*:	$P(t < t_c)$	$P(t > t_c)$	$2P(t > t_c)$ if t_c is positve $2P(t < t_c)$ if t_c is negative

Decision: Reject H_0 if $\alpha > p$-value, or if test statistic falls in rejection region

where $P(t > t_\alpha) = \alpha$, $P(t > t_{\alpha/2}) = \alpha/2$, $t_c =$ calculated value of the test statistic, and $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ true})$.

Assumption: The population from which the random sample is drawn is approximately normal.

Example 1.15

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 lowered the maximum allowable level of benzene in the workplace from 10 parts per million (ppm) to 1 ppm. 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 data in Table 1.12.

BENZENE

Table 1.12 Benzene content for 20 air samples

0.5	0.9	4.5	3.4	1.0
2.7	1.1	1.9	0.0	0.0
4.2	2.1	0.0	2.0	3.4
3.4	2.5	0.9	5.1	2.4

Is the steel manufacturing plant in violation of the changed 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

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: $t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$

Assumptions: 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 1.26)

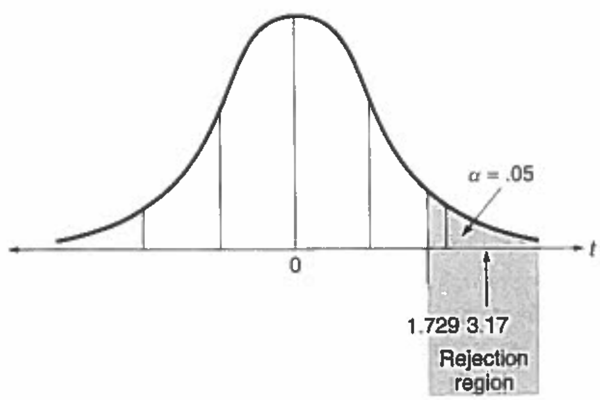


Figure 1.26 Rejection region for Example 1.15

The SAS printout, Figure 1.27, gives summary statistics for the sample data. Substituting $\bar{y} = 2.1$ and $s = 1.55$ into the test statistic formula, we obtain:

$$t = \frac{\bar{y} - 1}{s/\sqrt{n}} = \frac{2.1 - 1}{1.55/\sqrt{20}} = 3.17$$

This test statistic value, as well as the p-value of the test, is highlighted at the bottom of the SAS printout. Since the calculated t falls in the rejection region (or, since $\alpha = .05$ exceeds p-value $= .0025$), OSHA concludes that $\mu > 1$ ppm and the plant

Figure 1.27 SAS output for testing benzene mean

```
                        One Sample t-test for a Mean

Sample Statistics for benzene

     N          Mean          Std. Dev.       Std. Error
   --------------------------------------------------------
    20          2.10            1.55            0.35

Hypothesis Test

   Null hypothesis:    Mean of benzene <= 1
   Alternative:        Mean of benzene >  1

          t Statistic      Df      Prob > t
          ---------------------------------
             3.169         19       0.0025
```

is in violation of the revised government standards. The reliability associated with this inference is $\alpha = .05$. This implies that if the testing procedure were applied repeatedly to random samples of data collected at the plant, OSHA would falsely reject H_0 for only 5% of the tests. Consequently, OSHA is highly confident (95% confident) that the plant is violating the new standards. ■

1.9 Exercises

1.51 Key terms. Define each of the following:

(a) H_0 (b) H_a (c) Type I error
(d) Type II error (e) α (f) β
(g) p-value

1.52 Key questions. In hypothesis testing,

(a) who or what determines the size of the rejection region?
(b) does rejecting H_0 prove that the research hypothesis is correct?

1.53 Rejection regions. For each of the following rejection regions, sketch the sampling distribution for z, indicate the location of the rejection region, and give the value of α:

(a) $z > 1.96$ (b) $z > 1.645$ (c) $z > 2.576$
(d) $z < -1.29$ (e) $|z| > 1.645$ (f) $|z| > 2.576$

1.54 Play Golf America program. The Professional Golf Association (PGA) and *Golf Digest* have developed the Play Golf America program, in which teaching professionals at participating golf clubs provide a free 10-minute lesson to new customers. According to *Golf Digest* (July 2008), golf facilities that participate in the program gain, on average, \$2,400 in green fees, lessons, or equipment expenditures. A teaching professional at a golf club believes that the average gain in green fees, lessons, or equipment expenditures for participating golf facilities exceeds \$2,400.

(a) In order to support the claim made by the teaching professional, what null and alternative hypothesis should you test?
(b) Suppose you select $\alpha = .05$. Interpret this value in the words of the problem.
(c) For $\alpha = .05$, specify the rejection region of a large-sample test.

1.55 Mercury poisoning of wading birds. According to a University of Florida wildlife ecology and conservation researcher, the average level of mercury uptake in wading birds in the Everglades is declining (*UF News*, December 15, 2000). Last year, the average level was 15 parts per million.

(a) Give the null and alternative hypotheses for testing whether the average level this year is less than 15 ppm.
(b) Describe a Type I error for this test.
(c) Describe a Type II error for this test.

1.56 Crab-spiders hiding on flowers. *Behavioral Ecology* (January 2005) published the results of an experiment on crab-spiders' use of camouflage to hide from predators (e.g., birds) on flowers. Researchers at the French Museum of Natural History collected a sample of 10 adult female crab-spiders, each sitting on the yellow central part of a daisy, and measured the chromatic contrast of each spider to the flower. The data (where higher values indicate a greater contrast, and, presumably, an easier detection by predators) are shown in the next table. The researchers discovered that a contrast of 70 or greater allows birds to see the spider. Of interest is whether or not true mean chromatic contrast of crab-spiders on daisies is less than 70.

(a) Define the parameter of interest, μ.

(b) Setup the null and alternative hypothesis of interest.
(c) Find $\bar{y}$ and s for the sample data, then use these values to compute the test statistic.
(d) Give the rejection region for $\alpha = .10$.
(e) Find the p-value for the test.
(f) State the appropriate conclusion in the words of the problem.

CRABSPIDER

57	75	116	37	96	61	56	2	43	32

Source: Data adapted from Thery, M., et al. "Specific color sensitivities of prey and predator explain camouflage in different visual systems," *Behavioral Ecology*, Vol. 16, No. 1, Jan. 2005 (Table 1).

1.57 Social interaction of mental patients. The *Community Mental Health Journal* (August 2000) presented the results of a survey of over 6,000 clients of the Department of Mental Health and Addiction Services (DMHAS) in Connecticut. One of the many variables measured for each mental health patient was frequency of social interaction (on a 5-point scale, where 1 = very infrequently, 3 = occasionally, and 5 = very frequently). The 6,681 clients who were evaluated had a mean social interaction score of 2.95 with a standard deviation of 1.10.

(a) Conduct a hypothesis test (at $\alpha = .01$) to determine if the true mean social interaction score of all Connecticut mental health patients differs from 3.
(b) Examine the results of the study from a practical view, then discuss why "statistical significance" does not always imply "practical significance."
(c) Because the variable of interest is measured on a 5-point scale, it is unlikely that the population of ratings will be normally distributed. Consequently, some analysts may perceive the test, part a, to be invalid and search for alternative methods of analysis. Defend or refute this position.

1.58 Heart rate during laughter. Laughter is often called "the best medicine," since studies have shown that laughter can reduce muscle tension and increase oxygenation of the blood. In the *International Journal of Obesity* (January 2007), researchers at Vanderbilt University investigated the physiological changes that accompany laughter. Ninety subjects (18–34 years old) watched film clips designed to evoke laughter. During the laughing period, the researchers measured the heart rate (beats per minute) of each subject with the following summary results: $\bar{y} = 73.5$, $s = 6$. It is well known that the mean resting heart rate of adults is 71 beats/minute. At $\alpha = .05$, is there sufficient evidence to indicate that the true mean heart rate during laughter exceeds 71 beats/minute?

1.59 Feminizing faces study. Research published in *Nature* (August 27, 1998) revealed that people are more attracted to "feminized" faces, regardless of gender. In one experiment, 50 human subjects viewed both a Japanese female and Caucasian male face on a computer. Using special computer graphics, each subject could morph the faces (by making them more feminine or more masculine) until they attained the "most attractive" face. The level of feminization y (measured as a percentage) was measured.

(a) For the Japanese female face, $\bar{y} = 10.2\%$ and $s = 31.3\%$. The researchers used this sample information to test the null hypothesis of a mean level of feminization equal to 0%. Verify that the test statistic is equal to 2.3.
(b) Refer to part a. The researchers reported the p-value of the test as $p \approx .02$. Verify and interpret this result.
(c) For the Caucasian male face, $\bar{y} = 15.0\%$ and $s = 25.1\%$. The researchers reported the test statistic (for the test of the null hypothesis stated in part a) as 4.23 with an associated p-value of approximately 0. Verify and interpret these results.

1.60 Analyzing remote-sensing data to identify type of land cover. Geographers use remote sensing data from satellite pictures to identify urban land-cover as either grassland, commercial, or residential. In *Geographical Analysis* (October 2006), researchers from Arizona State, Florida State, and Louisiana State Universities collaborated on a new method for analyzing remote sensing data. A satellite photograph of an urban area was divided into 4 × 4 meter areas (called pixels). Of interest is a numerical measure of the distribution of gaps or hole sizes in the pixel, called *lacunarity*. The mean and standard deviation of the lacunarity measurements for a sample of 100 pixels randomly selected from a specific urban area are 225 and 20, respectively. It is known that the mean lacunarity measurement for all grassland pixels is 220. Do the data suggest that the area sampled is grassland? Test at $\alpha = .01$.

1.61 Falsifying candy counts. "Hot Tamales" are chewy, cinnamon-flavored candies. A bulk vending machine is known to dispense, on average, 15 Hot Tamales per bag. *Chance* (Fall 2000) published an article on a classroom project in which students were required to purchase bags of Hot Tamales from the machine and count the number of candies per bag. One student group claimed they purchased five bags that had the following candy counts: 25, 23, 21, 21, and 20. There was some question as to whether the students had fabricated the data. Use a hypothesis test to gain insight into whether or not the data collected by the students are fabricated. Use a level of significance that gives the benefit of the doubt to the students.

1.62 **Cooling method for gas turbines.** During periods of high electricity demand, especially during the hot summer months, the power output from a gas turbine engine can drop dramatically. One way to counter this drop in power is by cooling the inlet air to the gas turbine. An increasingly popular cooling method uses high pressure inlet fogging. The performance of a sample of 67 gas turbines augmented with high pressure inlet fogging was investigated in the *Journal of Engineering for Gas Turbines and Power* (January 2005). One measure of performance is heat rate (kilojoules per kilowatt per hour). Heat rates for the 67 gas turbines, saved in the GASTURBINE file, are listed in the table. Suppose that a standard gas turbine has, on average, a heat rate of 10,000 kJ/kWh.

(a) Conduct a test to determine if the mean heat rate of gas turbines augmented with high pressure inlet fogging exceeds 10,000 kJ/kWh. Use $\alpha = .05$.
(b) Identify a Type I error for this study, then a Type II error.

GASTURBINE

14622	13196	11948	11289	11964	10526	10387	10592	10460	10086
14628	13396	11726	11252	12449	11030	10787	10603	10144	11674
11510	10946	10508	10604	10270	10529	10360	14796	12913	12270
11842	10656	11360	11136	10814	13523	11289	11183	10951	9722
10481	9812	9669	9643	9115	9115	11588	10888	9738	9295
9421	9105	10233	10186	9918	9209	9532	9933	9152	9295
16243	14628	12766	8714	9469	11948	12414			

1.10 Inferences About the Difference Between Two Population Means

The reasoning employed in constructing a confidence interval and performing a statistical test for comparing two population means is identical to that discussed in Sections 1.7 and 1.8. First, we present procedures that are based on the assumption that we have selected *independent* random samples from the two populations. The parameters and sample sizes for the two populations, the sample means, and the sample variances are shown in Table 1.13. The objective of the sampling is to make an inference about the difference $(\mu_1 - \mu_2)$ between the two population means.

Because the sampling distribution of the difference between the sample means $(\bar{y}_1 - \bar{y}_2)$ is approximately normal for large samples, the large-sample techniques are

Table 1.13 Two-sample notation

	Population	
	1	2
Sample size	n_1	n_2
Population mean	μ_1	μ_2
Population variance	σ_1^2	σ_2^2
Sample mean	$\bar{y}_1$	$\bar{y}_2$
Sample variance	s_1^2	s_2^2

based on the standardized normal z statistic. Since the variances of the populations, σ_1^2 and σ_2^2, will rarely be known, we will estimate their values using s_1^2 and s_2^2.

To employ these large-sample techniques, we recommend that both sample sizes be large (i.e., each at least 30). The large-sample confidence interval and test are summarized in the boxes.

Large-Sample Confidence Interval for $(\mu_1 - \mu_2)$: Independent Samples

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

Assumptions: The two samples are randomly and independently selected from the two populations. The sample sizes, n_1 and n_2, are large enough so that $\bar{y}_1$ and $\bar{y}_2$ each have approximately normal sampling distributions and so that s_1^2 and s_2^2 provide good approximations to σ_1^2 and σ_2^2. This will be true if $n_1 \geq 30$ and $n_2 \geq 30$.

Large-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Independent Samples

Test statistic: $$z = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sigma_{(\bar{y}_1-\bar{y}_2)}} = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

	ONE-TAILED TESTS		TWO-TAILED TEST
	$H_0: \mu_1 - \mu_2 = D_0$	$H_0: \mu_1 - \mu_2 = D_0$	$H_0: \mu_1 - \mu_2 = D_0$
	$H_a: \mu_1 - \mu_2 < D_0$	$H_a: \mu_1 - \mu_2 > D_0$	$H_a: \mu_1 - \mu_2 \neq D_0$
Rejection region:	$z < -z_\alpha$	$z > z_\alpha$	$\lvert z\rvert > z_{\alpha/2}$
p-*value*:	$P(z < z_c)$	$P(z > z_c)$	$2P(z > z_c)$ if z_c is positve $2P(z < z_c)$ if z_c is negative

Decision: Reject H_0 if $\alpha > p$-value, or, if the test statistic falls in rejection region

where D_0 = hypothesized difference between means, $P(z > z_\alpha) = \alpha$, $P(z > z_{\alpha/2}) = \alpha/2$, z_c = calculated value of the test statistic, and α = P(Type I error) = P(Reject $H_0 | H_0$ true).

Assumptions: Same as for the previous large-sample confidence interval

Example 1.16

A dietitian has developed a diet that is low in fats, carbohydrates, and cholesterol. Although the diet was initially intended to be used by people with heart disease, the dietitian wishes to examine the effect this diet has on the weights of obese people. Two random samples of 100 obese people each are selected, and one group of 100 is placed on the low-fat diet. The other 100 are placed on a diet that contains approximately the same quantity of food but is not as low in fats, carbohydrates, and cholesterol. For each person, the amount of weight lost (or gained) in a 3-week period is recorded. The data, saved in the DIETSTUDY file, are listed in Table 1.14. Form a 95% confidence interval for the difference between the population mean weight losses for the two diets. Interpret the result.

*The symbol $\sigma_{(\bar{y}_1-\bar{y}_2)}$ is used to denote the standard error of the distribution of $(\bar{y}_1 - \bar{y}_2)$.

DIETSTUDY

Table 1.14 Diet study data, example 1.16

Weight Losses for Low-fat Diet

8	10	10	12	9	3	11	7	9	2
21	8	9	2	2	20	14	11	15	6
13	8	10	12	1	7	10	13	14	4
8	12	8	10	11	19	0	9	10	4
11	7	14	12	11	12	4	12	9	2
4	3	3	5	9	9	4	3	5	12
3	12	7	13	11	11	13	12	18	9
6	14	14	18	10	11	7	9	7	2
16	16	11	11	3	15	9	5	2	6
5	11	14	11	6	9	4	17	20	10

Weight Losses for Regular Diet

6	6	5	5	2	6	10	3	9	11
14	4	10	13	3	8	8	13	9	3
4	12	6	11	12	9	8	5	8	7
6	2	6	8	5	7	16	18	6	8
13	1	9	8	12	10	6	1	0	13
11	2	8	16	14	4	6	5	12	9
11	6	3	9	9	14	2	10	4	13
8	1	1	4	9	4	1	1	5	6
14	0	7	12	9	5	9	12	7	9
8	9	8	10	5	8	0	3	4	8

Solution

Let μ_1 represent the mean of the conceptual population of weight losses for all obese people who could be placed on the low-fat diet. Let μ_2 be similarly defined for the regular diet. We wish to form a confidence interval for $(\mu_1 - \mu_2)$.

Summary statistics for the diet data are displayed in the SPSS printout, Figure 1.28. Note that $\bar{y}_1 = 9.31$, $\bar{y}_2 = 7.40$, $s_1 = 4.67$, and $s_2 = 4.04$. Using these values and noting that $\alpha = .05$ and $z_{.025} = 1.96$, we find that the 95% confidence interval is:

$$(\bar{y}_1 - \bar{y}_2) \pm z_{.025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \approx$$

$$(9.31 - 7.40) \pm 1.96\sqrt{\frac{(4.67)^2}{100} + \frac{(4.04)^2}{100}} = 1.91 \pm (1.96)(.62) = 1.91 \pm 1.22$$

or (.69, 3.13). This result is also given (highlighted) on the SPSS printout. Using this estimation procedure over and over again for different samples, we know that approximately 95% of the confidence intervals formed in this manner will enclose the difference in population means $(\mu_1 - \mu_2)$. Therefore, we are highly confident that

Group Statistics

	DIET	N	Mean	Std. Deviation	Std. Error Mean
WTLOSS	LOWFAT	100	9.31	4.668	.467
	REGULAR	100	7.40	4.035	.404

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
WTLOSS	Equal variances assumed	1.367	.244	3.095	198	.002	1.910	.617	.693	3.127
	Equal variances not assumed			3.095	193.940	.002	1.910	.617	.693	3.127

Figure 1.28 SPSS analysis for diet study, Example 1.16

the mean weight loss for the low-fat diet is between .69 and 3.13 pounds more than the mean weight loss for the other diet. With this information, the dietitian better understands the potential of the low-fat diet as a weight-reducing diet. ■

The small-sample statistical techniques used to compare μ_1 and μ_2 with independent samples are based on the assumptions that both populations have normal probability distributions and that the variation within the two populations is of the same magnitude (i.e., $\sigma_1^2 = \sigma_2^2$). When these assumptions are approximately satisfied, we can employ a Student's t statistic to find a confidence interval and test a hypothesis concerning $(\mu_1 - \mu_2)$. The techniques are summarized in the following boxes.

Small-Sample Confidence Interval for $(\mu_1 - \mu_2)$: Independent Samples

$$(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

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

is a "pooled" estimate of the common population variance and $t_{\alpha/2}$ is based on $(n_1 + n_2 - 2)$ df.

Assumptions:

1. Both sampled populations have relative frequency distributions that are approximately normal.
2. The population variances are equal.
3. The samples are randomly and independently selected from the populations.

Example 1.17

Suppose you wish to compare a new method of teaching reading to "slow learners" to the current standard method. You decide to base this comparison on the results of a reading test given at the end of a learning period of 6 months. Of a random sample of 22 slow learners, 10 are taught by the new method and 12 are taught by the standard method. All 22 children are taught by qualified instructors under similar conditions for a 6-month period. The results of the reading test at the end of this period are given in Table 1.15.

Small-Sample Test of Hypothesis About $(\mu_1 - \mu_2)$: Independent Samples

Test statistic: $t = \dfrac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

	ONE-TAILED TESTS		TWO-TAILED TEST
	H_0: $\mu_1 - \mu_2 = D_0$	H_0: $\mu_1 - \mu_2 = D_0$	H_0: $\mu_1 - \mu_2 = D_0$
	H_a: $\mu_1 - \mu_2 < D_0$	H_a: $\mu_1 - \mu_2 > D_0$	H_a: $\mu_1 - \mu_2 \neq D_0$
Rejection region:	$t < -t_\alpha$	$z > t_\alpha$	$\lvert t\rvert > z_{\alpha/2}$
p-*value*:	$P(t < t_c)$	$P(z > t_c)$	$2P(t > t_c)$ if t_c is positve $2P(t < t_c)$ if t_c is negative

Decision: Reject H_0 if $\alpha > p$-value, or, if test statistic falls in rejection region

where D_0 = hypothesized difference between means, $P(t > t_\alpha) = \alpha$, $P(t > t_{\alpha/2}) = \alpha/2$, t_c = calculated value of the test statistic, and α = P(Type I error) = P(Reject $H_0 | H_0$ true).

Assumptions: Same as for the previous small-sample confidence interval.

READING

Table 1.15 Reading test scores for slow learners

New Method				Standard Method			
80	80	79	81	79	62	70	68
76	66	71	76	73	76	86	73
70	85			72	68	75	66

(a) Use the data in the table to test whether the true mean test scores differ for the new method and the standard method. Use $\alpha = .05$.

(b) What assumptions must be made in order that the estimate be valid?

Solution

(a) For this experiment, let μ_1 and μ_2 represent the mean reading test scores of slow learners taught with the new and standard methods, respectively. Then, we want to test the following hypothesis:

H_0: $(\mu_1 - \mu_2) = 0$ (i.e., no difference in mean reading scores)

H_a: $(\mu_1 - \mu_2) \neq 0$ (i.e., $\mu_1 \neq \mu_2$)

To compute the test statistic, we need to obtain summary statistics (e.g., $\bar{y}$ and s) on reading test scores for each method. The data of Table 1.15 was entered into a computer, and SAS was used to obtain these descriptive statistics. The SAS printout appears in Figure 1.29. Note that $\bar{y}_1 = 76.4$, $s_1 = 5.8348$, $\bar{y}_2 = 72.333$, and $s_2 = 6.3437$

Figure 1.29 SAS output for Example 1.17

```
                Two Sample t-test for the Means of SCORE within METHOD

Sample Statistics

    Group          N       Mean     Std. Dev.   Std. Error
    -------------------------------------------------------
    NEW           10       76.4      5.8348       1.8451
    STD           12   72.33333      6.3437       1.8313

Hypothesis Test

    Null hypothesis:     Mean 1 - Mean 2 =  0
    Alternative:         Mean 1 - Mean 2 ^= 0

    If Variances Are     t statistic      Df        Pr > t
    -------------------------------------------------------
    Equal                   1.552         20        0.1364
    Not Equal               1.564      19.77        0.1336

95% Confidence Interval for the Difference between Two Means

               Lower Limit    Upper Limit
               -----------    -----------
                  -1.40           9.53
```

Next, we calculate the pooled estimate of variance:

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

$$= \frac{(10 - 1)(5.8348)^2 + (12 - 1)(6.3437)^2}{10 + 12 - 2} = 37.45$$

where s_p^2 is based on $(n_1 + n_2 - 2) = (10 + 12 - 2) = 20$ degrees of freedom.

Now, we compute the test statistic:

$$t = \frac{(\bar{y}_1 - \bar{y}_2) - D_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(76.4 - 72.33) - 0}{\sqrt{37.45\left(\frac{1}{10} + \frac{1}{12}\right)}} = 1.55$$

The rejection region for this two-tailed test at $\alpha = .05$, based on 20 degrees of freedom, is

$$|t| > t_{.025} = 2.086 \qquad \text{(See Figure 1.30)}$$

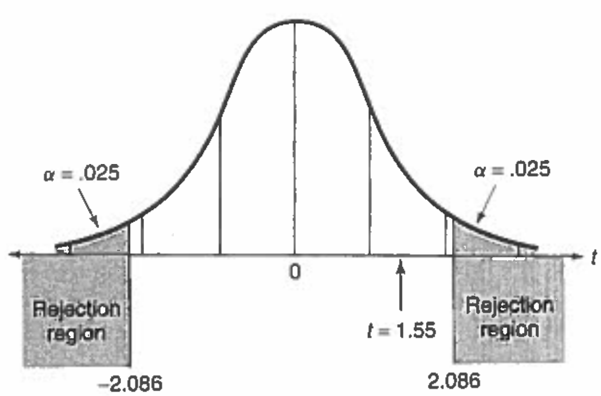


Figure 1.30 Rejection region for Example 1.17

Since the computed value of t does not fall in the rejection region, we fail to reject H_0. There is insufficient evidence (at $\alpha = .05$) of a difference between the true mean test scores for the two reading methods.

This conclusion can also be obtained by using the p-value approach. Both the test statistic ($t = 1.552$) and p-value (.1364) are highlighted on the SAS printout, Figure 1.29. Since $\alpha = .05$ is less than the p-value, we fail to reject H_0.

(b) To properly use the small-sample confidence interval, the following assumptions must be satisfied:

(1) The samples are randomly and independently selected from the populations of slow learners taught by the new method and the standard method.

(2) The test scores are normally distributed for both teaching methods.

(3) The variance of the test scores are the same for the two populations, that is, $\sigma_1^2 = \sigma_2^2$.

The two-sample t statistic is a powerful tool for comparing population means when the necessary assumptions are satisfied. It has also been found that the two-sample t procedure is more robust against nonnormal data than the one-sample method. And, when the sample sizes are equal, the assumption of equal population variances can be relaxed. That is, when $n_1 = n_2$, σ_1^2 and σ_2^2 can be quite different and the test statistic will still have (approximately) a Student's t distribution. ■

In Example 1.17, suppose it is possible to measure the slow learners' "reading IQs" *before* they are subjected to a teaching method. Eight pairs of slow learners with similar reading IQs are found, and one member of each pair is randomly assigned to the standard teaching method while the other is assigned to the new method. The data are given in Table 1.16. Do the data support the hypothesis that the population mean reading test score for slow learners taught by the new method is greater than the mean reading test score for those taught by the standard method?

Now, we want to test

$$H_0: (\mu_1 - \mu_2) = 0$$
$$H_a: (\mu_1 - \mu_2) > 0$$

It appears that we could conduct this test using the t statistic for two independent samples, as in Example 1.17. However, *the independent samples* t*-test is not a valid procedure to use with this set of data.* Why?

The t-test is inappropriate because the assumption of independent samples is invalid. We have randomly chosen *pairs of test scores*, and thus, once we have chosen the sample for the new method, we have *not* independently chosen the sample for the standard method. The dependence between observations within pairs can be seen by examining the pairs of test scores, which tend to rise and fall together as we go from pair to pair. This pattern provides strong visual evidence of a violation of the assumption of independence required for the two-sample t-test used in Example 1.17.

We now consider a valid method of analyzing the data of Table 1.16. In Table 1.17 we add the column of differences between the test scores of the pairs of slow learners. We can regard these differences in test scores as a random sample of differences for all pairs (matched on reading IQ) of slow learners, past and present. Then we can use this sample to make inferences about the mean of the population of differences, μ_d, which is equal to the difference $(\mu_1 - \mu_2)$. That is, the mean of the population (and sample) of differences equals the difference between the population (and sample) means. Thus, our test becomes

$$H_0: \mu_d = 0 \quad (\mu_1 - \mu_2 = 0)$$
$$H_a: \mu_d > 0 \quad (\mu_1 - \mu_2 > 0)$$

PAIREDSCORES

Table 1.16 Reading test scores for eight pairs of slow learners

Pair	New Method (1)	Standard Method (2)
1	77	72
2	74	68
3	82	76
4	73	68
5	87	84
6	69	68
7	66	61
8	80	76

Table 1.17 Differences in reading test scores

Pair	New Method	Standard Method	Difference (New Method–Standard Method)
1	77	72	5
2	74	68	6
3	82	76	6
4	73	68	5
5	87	84	3
6	69	68	1
7	66	61	5
8	80	76	4

The test statistic is a one-sample t (Section 1.9), since we are now analyzing a single sample of differences for small n:

$$\textit{Test statistic: } t = \frac{\bar{y}_d - 0}{s_d/\sqrt{n_d}}$$

where $\bar{y}_d$ = Sample mean difference
s_d = Sample standard deviation of differences
n_d = Number of differences = Number of pairs

Assumptions: The population of differences in test scores is approximately normally distributed. The sample differences are randomly selected from the population differences. [*Note*: We do not need to make the assumption that $\sigma_1^2 = \sigma_2^2$.]

Rejection region: At significance level $\alpha = .05$, we will reject H_0: if $t > t_{.05}$, where $t_{.05}$ is based on $(n_d - 1)$ degrees of freedom.

Referring to Table 2 in Appendix D, we find the t-value corresponding to $\alpha = .05$ and $n_d - 1 = 8 - 1 = 7$ df to be $t_{.05} = 1.895$. Then we will reject the null hypothesis if $t > 1.895$ (see Figure 1.31). Note that the number of degrees of freedom decreases from $n_1 + n_2 - 2 = 14$ to 7 when we use the paired difference experiment rather than the two independent random samples design.

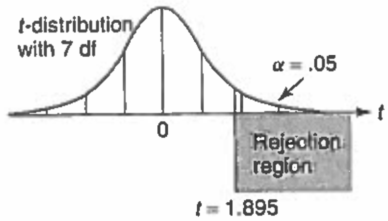


Figure 1.31 Rejection region for analysis of data in Table 1.17

Summary statistics for the $n = 8$ differences are shown in the MINITAB printout, Figure 1.32. Note that $\bar{y}_d = 4.375$ and $s_d = 1.685$. Substituting these values into the formula for the test statistic, we have

$$t = \frac{\bar{y}_d - 0}{s_d/\sqrt{n_d}} = \frac{4.375}{1.685/\sqrt{8}} = 7.34$$

Because this value of t falls in the rejection region, we conclude (at $\alpha = .05$) that the population mean test score for slow learners taught by the new method exceeds the population mean score for those taught by the standard method. We can reach the same conclusion by noting that the p-value of the test, highlighted in Figure 1.32, is much smaller than $\alpha = .05$.

This kind of experiment, in which observations are paired and the differences are analyzed, is called a **paired difference experiment**. The hypothesis-testing procedures and the method of forming confidence intervals for the difference between two means using a paired difference experiment are summarized in the next two boxes for both large and small n.

Figure 1.32 MINITAB paired difference analysis of data in Table 1.17

```
Paired T for NEW - STANDARD

             N    Mean   StDev  SE Mean
NEW          8   76.00    6.93     2.45
STANDARD     8   71.63    7.01     2.48
Difference   8   4.375   1.685    0.596

95% lower bound for mean difference: 3.246
T-Test of mean difference = 0 (vs > 0): T-Value = 7.34  P-Value = 0.000
```

Paired Difference Confidence Interval for $\mu_d = \mu_1 - \mu_2$

Large Sample

$$\bar{y}_d \pm z_{\alpha/2}\frac{\sigma_d}{\sqrt{n_d}} \approx \bar{y}_d \pm z_{\alpha/2}\frac{s_d}{\sqrt{n_d}}$$

Assumption: Sample differences are randomly selected from the population.

Small Sample

$$\bar{y}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n_d}}$$

where $t_{\alpha/2}$ is based on $(n_d - 1)$ degrees of freedom

Assumptions:

1. Population of differences has a normal distribution.
2. Sample differences are randomly selected from the population.

Paired Difference Test of Hypothesis for $\mu_d = \mu_1 - \mu_2$

	ONE-TAILED TESTS		TWO-TAILED TEST
	H_0: $\mu_d = D_0$	H_0: $\mu_d = D_0$	H_0: $\mu_d = D_0$
	H_a: $\mu_d < D_0$	H_a: $\mu_d > D_0$	H_a: $\mu_d \neq D_0$

Large Sample

$$\textit{Test statistic: } z = \frac{\bar{y}_d - D_0}{\sigma_d/\sqrt{n_d}} \approx \frac{\bar{y}_d - D_0}{s_d/\sqrt{n_d}}$$

Rejection Region:	$z < -z_\alpha$	$z > z_d$	$\lvert z\rvert > z_{\alpha/2}$
p-*value*:	$P(z < z_c)$	$P(z > z_c)$	$2P(z > z_c)$ if z_c positive $2P(z < z_c)$ if z_c negative

Assumption: The differences are randomly selected from the population of differences.

Small Sample

$$\textit{Test statistic: } t = \frac{\bar{y}_d - D_0}{s_d/\sqrt{n_d}}$$

Rejection region:	$t < -t_\alpha$	$t > t_\alpha$	$\lvert t\rvert > t_{\alpha/2}$
p-*value*:	$P(t < t_c)$	$P(t > t_c)$	$2P(t > t_c)$ if t_c is positive $2P(t < t_c)$ if t_c is negative

Assumptions:

1. The relative frequency distribution of the population of differences is normal.
2. The differences are randomly selected from the population of differences.

1.10 Exercises

1.63 Describe the sampling distribution of $(\bar{y}_1 - \bar{y}_2)$.

1.64 To use the t statistic to test for differences between the means of two populations based on independent samples, what assumptions must be made about the two sampled populations? What assumptions must be made about the two samples?

1.65 **How do you choose to argue?** Educators frequently lament weaknesses in students' oral and written arguments. In *Thinking and Reasoning* (October 2006), researchers at Columbia University conducted a series of studies to assess the cognitive skills required for successful arguments. One study focused on whether students would choose to argue by weakening the opposing position or by strengthening the favored position. (Example: You are told you would do better at basketball than soccer, but you like soccer. An argument that weakens the opposing position is "You need to be tall to play basketball." An argument that strengthens the favored position is "With practice, I can become really good at soccer.") A sample of 52 graduate students in psychology was equally divided into two groups. Group 1 was presented with 10 items, where the argument always attempts to strengthen the favored position. Group 2 was presented with the same 10 items, but where the argument always attempts to weaken the nonfavored position. Each student then rated the 10 arguments on a 5-point scale from very weak (1) to very strong (5). The variable of interest was the sum of the 10 item scores, called total rating. Summary statistics for the data are shown in the table. Use the methodology of this chapter to compare the mean total ratings for the two groups, at $\alpha = .05$. Give a practical interpretation of the results in the words of the problem.

	GROUP 1 (SUPPORT FAVORED POSITION)	GROUP 2 (WEAKEN OPPOSING POSITION)
Sample size:	26	26
Mean:	28.6	24.9
Standard deviation:	12.5	12.2

Source: Kuhn, D., and Udell, W. "Coordinating own and other perspectives in argument," *Thinking and Reasoning*, October 2006.

1.66 Eating disorder study. The "fear of negative evaluation" (FNE) scores for 11 female students known to suffer from the eating disorder bulimia and 14 female students with normal eating habits, first presented in Exercise 1.19 (p. 16), are reproduced below. (Recall that the higher the score, the greater the fear of negative evaluation.)

(a) Find a 95% confidence interval for the difference between the population means of the FNE scores for bulimic and normal female students. Interpret the result.

(b) What assumptions are required for the interval of part a to be statistically valid? Are these assumptions reasonably satisfied? Explain.

1.67 Does rudeness really matter in the workplace? Studies have established that rudeness in the workplace can lead to retaliatory and counterproductive behavior. However, there has been little research on how rude behaviors influence a victim's task performance. Such a study was conducted and the results published in the *Academy of Management Journal* (October 2007). College students enrolled in a management course were randomly assigned to one of two experimental conditions: rudeness condition (45 students) and control group (53 students). Each student was asked to write down as many uses for a brick as possible in 5 minutes. For those students in the rudeness condition, the facilitator displayed rudeness by berating the students in general for being irresponsible and unprofessional (due to a late-arriving confederate). No comments were made about the late-arriving confederate for students in the control group. The number of different uses for a brick was recorded for each of the 98 students and the data saved in the RUDE file, shown below. Conduct a statistical analysis (at $\alpha = .01$) to determine if the true mean performance level for students in the rudeness condition is lower than the true mean performance level for students in the control group.

EVOS

1.68 Impact study of an oil spill. The *Journal of Agricultural, Biological, and Environmental Statistics* (September 2000) reported on an impact study of a tanker oil spill on the seabird population in Alaska. For each of 96 shoreline locations (called transects), the number of seabirds found, the length (in kilometers) of the transect, and whether or not the transect was in an oiled area were recorded. (The data are saved in the EVOS file.) Observed seabird density is defined as the observed count

BULIMIA

Bulimic students	21	13	10	20	25	19	16	21	24	13	14			
Normal students	13	6	16	13	8	19	23	18	11	19	7	10	15	20

RUDE

Control Group:

1 24 5 16 21 7 20 1 9 20 19 10 23 16 0 4 9 13 17 13 0 2 12 11 7
1 19 9 12 18 5 21 30 15 4 2 12 11 10 13 11 3 6 10 13 16 12 28 19 12 20 3 11

Rudeness Condition:

4 11 18 11 9 6 5 11 9 12 7 5 7 3 11 1 9 11 10 7 8 9 10 7
11 4 13 5 4 7 8 3 8 15 9 16 10 0 7 15 13 9 2 13 10

divided by the length of the transect. A comparison of the mean densities of oiled and unoiled transects is displayed in the following MINITAB printout. Use this information to make an inference about the difference in the population mean seabird densities of oiled and unoiled transects.

```
Two-Sample T-Test and CI: density, oil

Two-sample T for density

oil      N     Mean    StDev   SE Mean
no      36     3.27     6.70       1.1
yes     60     3.50     5.97      0.77

Difference = mu (no ) - mu (yes)
Estimate for difference:  -0.22
95% CI for difference: (-2.93, 2.49)
T-Test of difference = 0 (vs not =): T-Value = -0.16  P-Value = 0.871  DF = 67
```

1.69 Family involvement in homework. Teachers Involve Parents in Schoolwork (TIPS) is an interactive homework process designed to improve the quality of homework assignments for elementary, middle, and high school students. TIPS homework assignments require students to conduct interactions with family partners (parents, guardians, etc.) while completing the homework. Frances Van Voorhis (Johns Hopkins University) conducted a study to investigate the effects of TIPS in science, mathematics, and language arts homework assignments (April 2001). Each in a sample of 128 middle school students was assigned to complete TIPS homework assignments, while 98 students in a second sample were assigned traditional, noninteractive homework assignments (called ATIPS). At the end of the study, all students reported on the level of family involvement in their homework on a 4-point scale (0 = never, 1 = rarely, 2 = sometimes, 3 = frequently, 4 = always). Three scores were recorded for each student: one for science homework, one for math homework, and one for language arts homework. The data for the study are saved in the HWSTUDY file. (The first five and last five observations in the data set are listed in the next table.)

(a) Conduct an analysis to compare the mean level of family involvement in science homework assignments of TIPS and ATIPS students. Use $\alpha = .05$. Make a practical conclusion.

(b) Repeat part a for mathematics homework assignments.

(c) Repeat part a for language arts homework assignments.

(d) What assumptions are necessary for the inferences of parts a–c to be valid? Are they reasonably satisfied?

HWSTUDY (First and last five observations)

HOMEWORK CONDITION	SCIENCE	MATH	LANGUAGE
ATIPS	1	0	0
ATIPS	0	1	1
ATIPS	0	1	0
ATIPS	1	2	0
ATIPS	1	1	2
TIPS	2	3	2
TIPS	1	4	2
TIPS	2	4	2
TIPS	4	0	3
TIPS	2	0	1

Source: Van Voorhis, F. L. "Teachers' use of interactive homework and its effects on family involvement and science achievement of middle grade students." Paper presented at the annual meeting of the American Educational Research Association, Seattle, April 2001.

1.70 Comparing voltage readings. Refer to the Harris Corporation/University of Florida study to determine whether a manufacturing process performed at a remote location could be established locally, Exercise 1.22 (p. 17). Test devices (pilots) were setup 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 below. Descriptive statistics are displayed in the SAS printout. [*Note*: Larger voltage readings are better than smaller voltage readings.]

VOLTAGE

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.

(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?

1.71 Laughter among deaf signers. The *Journal of Deaf Studies and Deaf Education* (Fall 2006) published an article on vocalized laughter among deaf

SAS output for Exercise 1.70

```
------------------------------ location=NEW ------------------------------
                         The MEANS Procedure
                   Analysis Variable : voltage
 N          Mean          Std Dev          Minimum          Maximum
 30     9.4223333       0.4788757        8.5100000       10.1200000

------------------------------ location=OLD ------------------------------
                   Analysis Variable : voltage
 N          Mean          Std Dev          Minimum          Maximum
 30     9.8036667       0.5409155        8.0500000       10.5500000
```

users of American sign language (ASL). In videotaped ASL conversations among deaf participants, 28 laughed at least once. The researchers wanted to know if they laughed more as speakers (while signing) or as audience members (while listening). For each of the 28 deaf participants, the number of laugh episodes as a speaker and the number of laugh episodes as an audience member was determined. One goal of the research was to compare the mean number of laugh episodes of speakers and audience members.

(a) Explain why the data should be analyzed as a paired difference experiment.
(b) Identify the study's target parameter.
(c) The study yielded a sample mean of 3.4 laughter episodes for speakers and a sample mean of 1.3 laughter episodes as an audience. Is this sufficient evidence to conclude that the population means are different? Explain.
(d) A paired difference t-test resulted in $t = 3.14$ and p-value $< .01$. Interpret the results in the words of the problem.

1.72 **Impact of red light cameras on car crashes.** To combat red light–running crashes—the phenomenon of a motorist entering an intersection after the traffic signal turns red and causing a crash—many states are adopting photo-red enforcement programs. In these programs, red light cameras installed at dangerous intersections photograph the license plates of vehicles that run the red light. How effective are photo-red enforcement programs in reducing red light–running crash incidents at intersections? The Virginia Department of Transportation (VDOT) conducted a comprehensive study of its newly adopted photo-red enforcement program and published the results in a June 2007 report. In one portion of the study, the VDOT provided crash data both before and after installation of red light cameras at several intersections. The data (measured as the number of crashes caused by red light running per intersection per year) for 13 intersections in Fairfax County, Virginia, are given in the table. Analyze the data for the VDOT. What do you conclude?

REDLIGHT

INTERSECTION	BEFORE CAMERA	AFTER CAMERA
1	3.60	1.36
2	0.27	0
3	0.29	0
4	4.55	1.79
5	2.60	2.04
6	2.29	3.14
7	2.40	2.72
8	0.73	0.24
9	3.15	1.57
10	3.21	0.43
11	0.88	0.28
12	1.35	1.09
13	7.35	4.92

Source: Virginia Transportation Research Council, "Research Report: The Impact of Red Light Cameras (Photo-Red Enforcement) on Crashes in Virginia," June 2007.

1.73 **Light-to-dark transition of genes.** *Synechocystis*, a type of cyanobacterium that can grow and survive in a wide range of conditions, is used by scientists to model DNA behavior. In the *Journal of Bacteriology* (July 2002), scientists isolated genes of the bacterium responsible for photosynthesis and respiration and investigated the sensitivity of the genes to light. Each gene sample was grown to midexponential phase in a growth incubator in "full light." The lights were extinguished and growth measured after 24 hours in the dark ("full dark"). The lights were then turned back on for 90 minutes ("transient light") followed immediately by an additional 90 minutes in the dark ("transient dark"). Standardized growth measurements in each light/dark condition were obtained for 103 genes. The complete data set is saved in the GENEDARK file. Data for the first 10 genes are shown in the accompanying table.

(a) Treat the data for the first 10 genes as a random sample collected from the population of 103 genes and test the hypothesis of no difference between the mean standardized growth of genes in the full-dark condition and genes in the transient light condition. Use $\alpha = .01$.

(b) Use a statistical software package to compute the mean difference in standardized growth of the 103 genes in the full-dark condition and the transient-light condition. Did the test, part a, detect this difference?

(c) Repeat parts a and b for a comparison of the mean standardized growth of genes in the full-dark condition and genes in the transient-dark condition.

GENEDARK (*first 10 observations shown*)

GENE ID	FULL-DARK	TR-LIGHT	TR-DARK
SLR2067	−0.00562	1.40989	−1.28569
SLR1986	−0.68372	1.83097	−0.68723
SSR3383	−0.25468	−0.79794	−0.39719
SLL0928	−0.18712	−1.20901	−1.18618
SLR0335	−0.20620	1.71404	−0.73029
SLR1459	−0.53477	2.14156	−0.33174
SLL1326	−0.06291	1.03623	0.30392
SLR1329	−0.85178	−0.21490	0.44545
SLL1327	0.63588	1.42608	−0.13664
SLL1325	−0.69866	1.93104	−0.24820

Source: Gill, R. T., et al. "Genome-wide dynamic transcriptional profiling of the light to dark transition in *Synechocystis Sp.* PCC6803," *Journal of Bacteriology*, Vol. 184, No. 13, July 2002.

(d) Repeat parts a and b for a comparison of the mean standardized growth of genes in the transient-light condition and genes in the transient-dark condition.

1.11 Comparing Two Population Variances

Suppose you want to use the two-sample t statistic to compare the mean productivity of two paper mills. However, you are concerned that the assumption of equal variances of the productivity for the two plants may be unrealistic. It would be helpful to have a statistical procedure to check the validity of this assumption.

The common statistical procedure for comparing population variances σ_1^2 and σ_2^2 is to make an inference about the ratio, σ_1^2/σ_2^2, using the ratio of the sample variances, s_1^2/s_2^2. Thus, we will attempt to support the research hypothesis that the ratio σ_1^2/σ_2^2 differs from 1 (i.e., the variances are unequal) by testing the null hypothesis that the ratio equals 1 (i.e., the variances are equal).

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \ (\sigma_1^2 = \sigma_2^2)$$

$$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \ (\sigma_1^2 \neq \sigma_2^2)$$

We will use the test statistic

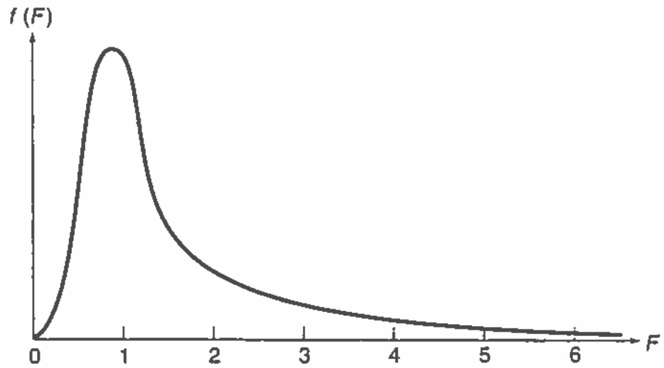
$$F = \frac{s_1^2}{s_2^2}$$

To establish a rejection region for the test statistic, we need to know how s_1^2/s_2^2 is distributed in repeated sampling. That is, we need to know the sampling distribution of s_1^2/s_2^2. As you will subsequently see, the sampling distribution of s_1^2/s_2^2 depends on two of the assumptions already required for the t-test, as follows:

1. The two sampled populations are normally distributed.
2. The samples are randomly and independently selected from their respective populations.

When these assumptions are satisfied and when the null hypothesis is true (i.e., $\sigma_1^2 = \sigma_2^2$), the sampling distribution of s_1^2/s_2^2 is an ***F* distribution** with $(n_1 - 1)$ df and $(n_2 - 1)$ df, respectively. The shape of the F distribution depends on the degrees of freedom associated with s_1^2 and s_2^2, that is, $(n_1 - 1)$ and $(n_2 - 1)$. An F distribution with 7 and 9 df is shown in Figure 1.33. As you can see, the distribution is skewed to the right.

Figure 1.33 An F distribution with 7 and 9 df



When the population variances are unequal, we expect the ratio of the sample variances, $F = s_1^2/s_2^2$, to be either very large or very small. Therefore, we will need to find F-values corresponding to the tail areas of the F distribution to establish the rejection region for our test of hypothesis. The upper-tail F-values can be found in Tables 3, 4, 5, and 6 of Appendix D. Table 4 is partially reproduced in Table 1.18. It gives F-values that correspond to $\alpha = .05$ upper-tail areas for different degrees of freedom. The columns of the tables correspond to various degrees of freedom for the numerator sample variance s_1^2, whereas the rows correspond to the degrees of freedom for the denominator sample variance s_2^2.

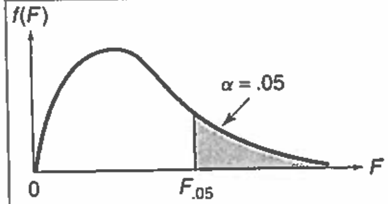
Thus, if the numerator degrees of freedom is 7 and the denominator degrees of freedom is 9, we look in the seventh column and ninth row to find $F_{.05} = 3.29$. As shown in Figure 1.34, $\alpha = .05$ is the tail area to the right of 3.29 in the F distribution with 7 and 9 df. That is, if $\sigma_1^2 = \sigma_2^2$, the probability that the F statistic will exceed 3.29 is $\alpha = .05$.

Suppose we want to compare the variability in production for two paper mills and we have obtained the following results:

Sample 1	Sample 2
$n_1 = 13$ days	$n_2 = 18$ days
$\bar{y}_1 = 26.3$ production units	$\bar{y}_2 = 19.7$ production units
$s_1 = 8.2$ production units	$s_2 = 4.7$ production units

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 3, 4, 5, and 6. To accomplish this, **we will always place the larger sample variance in the numerator of the *F*-test**. This doubles 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. In effect, we make the test two-tailed by putting the larger variance in the numerator rather than establishing rejection regions in both tails.

Table 1.18 Reproduction of part of Table 4 of Appendix D: $\alpha = .05$



v_2 \ v_1	Numerator Degress of Freedom 1	2	3	4	5	6	7	8	9
Denominator Degress of Freedom 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	3.25	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

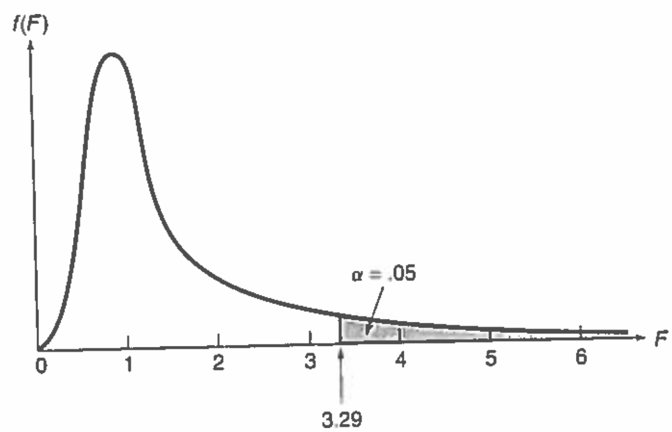


Figure 1.34 An F distribution for 7 and 9 df: $\alpha = .05$

Thus, for our production example, we have a numerator s_1^2 with df $= n_1 - 1 = 12$ and a denominator s_2^2 with df $= n_2 - 1 = 17$. Therefore, the test statistic will be

$$F = \frac{\text{Larger sample variance}}{\text{Smaller sample variance}} = \frac{s_1^2}{s_2^2}$$

and we will reject H_0: $\sigma_1^2 = \sigma_2^2$ for $\alpha = .10$ if the calculated value of F exceeds the tabulated value:

$$F_{.05} = 2.38 \text{ (see Figure 1.34)}$$

Now, what do the data tell us? We calculate

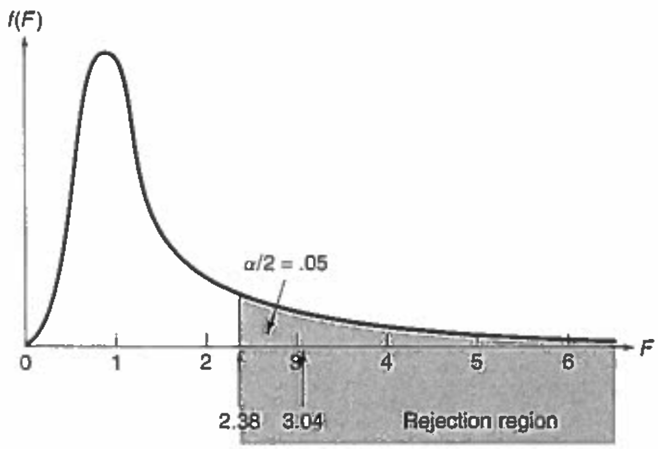
$$F = \frac{s_1^2}{s_2^2} = \frac{(8.2)^2}{(4.7)^2} = 3.04$$

and compare it to the rejection region shown in Figure 1.35. Since the calculated F-value, 3.04, falls in the rejection region, the data provide sufficient evidence to indicate that the population variances differ. Consequently, we would be reluctant to use the two-sample t statistic to compare the population means, since the assumption of equal population variances is apparently untrue.

What would you have concluded 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 (accepting H_0 when H_a is true) without knowing the probability of this error (the probability of accepting H_0: $\sigma_1^2 = \sigma_2^2$ when it is false). Since we will not consider the calculation of β for specific alternatives in this text, 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 $\sigma_1^2 = \sigma_2^2$.**

The F-test for equal population variances is summarized in the box on p. 68.

Figure 1.35 Rejection region for production example F distribution



Example 1.18

In Example 1.17 we used the two-sample t statistic to compare the mean reading scores of two groups of slow learners who had been taught to read using two different methods. The data are repeated in Table 1.19 for convenience. The use of the t statistic was based on the assumption that the population variances of the test scores were equal for the two methods. Conduct a test of hypothesis to check this assumption at $\alpha = .10$.

READING

Table 1.19 Reading test scores for slow learners

New Method				Standard Method			
80	80	79	81	79	62	70	68
76	66	71	76	73	76	86	73
70	85			72	68	75	66

F-Test for Equal Population Variances: Independent Samples*

	ONE-TAILED TESTS		TWO-TAILED TEST
	H_0: $\sigma_1^2 = \sigma_2^2$	H_0: $\sigma_1^2 = \sigma_2^2$	H_0: $\sigma_1^2 = \sigma_2^2$
	H_a: $\sigma_1^2 < \sigma_2^2$	H_a: $\sigma_1^2 > \sigma_2^2$	H_a: $\sigma_1^2 \neq \sigma_2^2$
Test statistic:	$F = s_2^2/s_1^2$	$F = s_1^2/s_2^2$	$F = \dfrac{\text{Larger sample variance}}{\text{Smaller sample variance}}$
Rejection region:	$F > F_\alpha$	$F > F_\alpha$	$F > F_{\alpha/2}$
Numerator df (ν_1):	$n_2 - 1$	$n_1 - 1$	$n-1$ for larger variance
Denominator df (ν_2):	$n_1 - 1$	$n_2 - 1$	$n-1$ for smaller variance
p-value:	$P(F > F_c)$	$P(F > F_c)$	$P(F^* < 1/F_c) + P(F > F_c)$

Decision: Reject H_0 if $\alpha > p$-value, or, if test statistic falls in rejection region

where F is based on ν_1 numerator df and ν_2 denominator df; F^* is based on ν_2 numerator df and ν_1 denominator df; F_c = calculated value of the test statistic; and α = P(Type I error) = P(Reject $H_0 | H_0$ true).

Assumptions:

1. Both sampled populations are normally distributed.
2. The samples are random and independent.

Figure 1.36 SAS F-test for the data in Table 1.19

Two Sample Test for Variances of SCORE within METHOD

Sample Statistics

METHOD Group	N	Mean	Std. Dev.	Variance
1:STD	12	72.33333	6.3437	40.24242
2:NEW	10	76.4	5.8348	34.04444

Hypothesis Test

```
Null hypothesis:     Variance 1 / Variance 2 =  1
Alternative:         Variance 1 / Variance 2 ^= 1
```

F	- Degrees of Freedom - Numer.	Denom.	Pr > F
1.18	11	9	0.8148

Solution

We want to test

$$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1 \text{ (i.e., } \sigma_1^2 = \sigma_2^2)$$

$$H_a: \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \text{ (i.e., } \sigma_1^2 \neq \sigma_2^2)$$

*Although a test of a hypothesis of equality of variances is the most common application of the F-test, it can also be used to test a hypothesis that the ratio between the population variances is equal to some specified value, H_0: $\sigma_1^2/\sigma_2^2 = k$. The test is conducted in exactly the same way as specified in the box, except that we use the test statistic

$$F = \left(\frac{s_1^2}{s_2^2}\right)\left(\frac{1}{k}\right)$$

The data were entered into SAS, and the SAS printout shown in Figure 1.36 was obtained. Both the test statistic, $F = 1.18$, and two-tailed p-value, .8148, are highlighted on the printout. Since $\alpha = .10$ is less than the p-value, we do not reject the null hypothesis that the population variances of the reading test scores are equal. ■

The previous examples demonstrate how to conduct a two-tailed F-test when the alternative hypothesis is H_a: $\sigma_1^2 \neq \sigma_2^2$. One-tailed tests for determining whether one population variance is larger than another population variance (i.e., H_a: $\sigma_1^2 > \sigma_2^2$) are conducted similarly. However, the α value no longer needs to be doubled since the area of rejection lies only in the upper (or lower) tail area of the F distribution. The procedure for conducting an upper-tailed F-test is outlined in the previous box. Whenever you conduct a one-tailed F-test, be sure to write H_a in the form of an upper-tailed test. This can be accomplished by numbering the populations so that the variance hypothesized to be larger in H_a is associated with population 1 and the hypothesized smaller variance is associated with population 2.

Important: As a final comment, we note that (unlike the small-sample t procedure for means) the F-test for comparing variances is not very robust against nonnormal data. Consequently, with nonnormal data it is difficult to determine whether a significant F-value implies that the population variances differ or is simply due to the fact that the populations are not normally distributed.

1.11 Exercises

1.74 Use Tables 3, 4, 5, and 6 of Appendix D to find F_α for α, numerator df, and denominator df equal to:

(a) .05, 8, 7 (b) .01, 15, 20
(c) .025, 12, 5 (d) .01, 5, 25
(e) .10, 5, 10 (f) .05, 20, 9

1.75 **Is honey a cough remedy?** Refer to the *Archives of Pediatrics and Adolescent Medicine* (December 2007) study of honey as a children's cough remedy, Exercise 1.21 (p. 17). The data (cough improvement scores) for the 33 children in the DM dosage group and the 35 children in the honey dosage group are reproduced in the table below. The researchers want to know if the variability in coughing improvement scores differs for the two groups. Conduct the appropriate analysis, using $\alpha = .10$.

1.76 **How do you choose to argue?** Refer to the *Thinking and Reasoning* (October 2006) study of the cognitive skills required for successful arguments, Exercise 1.65 (p. 60). Recall that 52 psychology graduate students were equally divided into two groups. Group 1 was presented with arguments that always attempted to strengthen the favored position. Group 2 was presented with arguments that always attempted to weaken the nonfavored position. Summary statistics for the student ratings of the arguments are reproduced in the table. In Exercise 1.65 you compared the mean ratings for the two groups with a small-sample t-test, assuming equal variances. Determine the validity of this assumption at $\alpha = .05$.

	GROUP 1 (SUPPORT FAVORED POSITION)	GROUP 2 (WEAKEN OPPOSING POSITION)
Sample size:	26	26
Mean:	28.6	24.9
Standard deviation:	12.5	12.2

Source: Kuhn, D., and Udell, W. "Coordinating won and other perspectives in argument," *Thinking and Reasoning*, October 2006.

HONEYCOUGH

Honey Dosage: 12 11 15 11 10 13 10 4 15 16 9 14 10 6 10 8 11 12 12 8
12 9 11 15 10 15 9 13 8 12 10 8 9 5 12

DM Dosage: 4 6 9 4 7 7 7 9 12 10 11 6 3 4 9 12 7 6 8 12 12 4 12
13 7 10 13 9 4 4 10 15 9

Source: Paul, I. M., et al. "Effect of honey, dextromethorphan, and no treatment on nocturnal cough and sleep quality for coughing children and their parents," *Archives of Pediatrics and Adolescent Medicine*, Vol. 161, No. 12, Dec. 2007 (data simulated).

1.77 **Eating disorder study.** Refer to Exercise 1.66 (p. 61). The "fear of negative evaluation" (FNE) scores for the 11 bulimic females and 14 females with normal eating habits are reproduced in the table. The confidence interval you constructed in Exercise 1.66 requires that the variance of the FNE scores of bulimic females is equal to the variance of the FNE scores of normal females. Conduct a test (at $\alpha = .05$) to determine the validity of this assumption.

BULIMIA

Bulimic:	21 13 10 20 25 19 16 21 24 13 14
Normal:	13 6 16 13 8 19 23 18 11 19 7 10 15 20

Source: Randles, R. H. "On neutral responses (zeros) in the sign test and ties in the Wilcoxon-Mann-Whitney test," *American Statistician*, Vol. 55, No. 2, May 2001 (Figure 3).

1.78 **Human inspection errors.** Tests of product quality using human inspectors can lead to serious inspection error problems. To evaluate the performance of inspectors in a new company, a quality manager had a sample of 12 novice inspectors evaluate 200 finished products. The same 200 items were evaluated by 12 experienced inspectors. The quality of each item—whether defective or nondefective—was known to the manager. The next table lists the number of inspection errors (classifying a defective item as nondefective or vice versa) made by each inspector. A SAS printout comparing the two types of inspectors is shown in the next column.

(a) Prior to conducting this experiment, the manager believed the variance in inspection errors was lower for experienced inspectors than for novice inspectors. Do the sample data support her belief? Test using $\alpha = .05$.

(b) What is the appropriate p-value of the test you conducted in part a?

INSPECT

NOVICE INSPECTORS				EXPERIENCED INSPECTORS			
30	35	26	40	31	15	25	19
36	20	45	31	28	17	19	18
33	29	21	48	24	10	20	21

```
              Two Sample Test for Variances of ERRORS within INSPECT

Sample Statistics

     INSPECT
     Group          N      Mean      Std. Dev.    Variance
     ------------------------------------------------------
     1NOVICE       12   32.83333      8.6427     74.69697
     2EXPER        12   20.58333      5.7439     32.99242

Hypothesis Test

     Null hypothesis:     Variance 1 / Variance 2 <= 1
     Alternative:         Variance 1 / Variance 2 >  1

                   - Degrees of Freedom -
          F          Numer.     Denom.           Pr > F
     ---------------------------------------------------
        2.26           11         11             0.0955
```

1.79 **Variation in wet sampler readings.** 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 were 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 vs. three wet samplers). Test using $\alpha = .05$.

Quick Summary/Guides

KEY IDEAS

Types of statistical applications

1. descriptive
2. inferential

Descriptive statistics

1. Identify **population** or **sample** (collection of **experimental units**)
2. Identify **variable(s)**
3. Collect **data**
4. **Describe** data

Inferential statistics

1. Identify **population** (collection of *all* **experimental units**)
2. Identify **variable(s)**
3. Collect **sample** data (*subset* of population)
4. **Inference** about population based on sample
5. **Measure of reliability** for inference

Types of data

1. qualitative (categorical in nature)
2. quantitative (numerical in nature)

Graphs for qualitative data

1. pie chart
2. bar graph

Graphs for quantitative data

1. stem-and-leaf display
2. histogram

Measure of central tendency

mean (or **average**)

Measures of variation

1. range
2. variance
3. standard deviation

Percentage of measurements within 2 standard deviations of the mean

1. *any data set*: at least 3/4 (Tchebysheff's Theorem)
2. ***normal** (mound-shaped) **distribution***: 95%

Key Formulas

	Sample	Population
Mean:	$\bar{y} = (\Sigma y_i)/n$	μ
Variance:	$s^2 = \dfrac{\Sigma(y_i - \bar{y})^2}{n-1}$	σ^2
Std. Dev.	$s = \sqrt{s^2}$	σ

Properties of the sampling distribution of $\bar{y}$

1. $E(\bar{y}) = \mu$
2. $\text{Var}(\bar{y}) = \sigma^2/n$

Central Limit Theorem

For large n, the sampling distribution of $\bar{y}$ is approximately normal.

Formulation of Confidence Intervals for a Population Parameter θ and Test Statistics for H_0: $\theta = \theta_0$, where $\theta = \mu$ or $(\mu_1 - \mu_2)$

SAMPLE SIZE	CONFIDENCE INTERVAL	TEST STATISTIC
Large	$\hat{\theta} \pm z_{\alpha/2} s_{\hat{\theta}}$	$z = \dfrac{\hat{\theta} - \theta_0}{s_{\hat{\theta}}}$
Small	$\hat{\theta} \pm t_{\alpha/2} s_{\hat{\theta}}$	$t = \dfrac{\hat{\theta} - \theta_0}{s_{\hat{\theta}}}$

Note: The test statistic for testing H_0: $\sigma_1^2/\sigma_2^2 = 1$ is $F = s_1^2/s_2^2$ (see the box on page 68).

Population Parameters and Corresponding Estimators and Standard Errors

PARAMETER (θ)	ESTIMATOR ($\hat{\theta}$)	STANDARD ERROR ($\sigma_{\hat{\theta}}$)	ESTIMATE OF STANDARD ERROR ($s_{\hat{\theta}}$)
μ Mean (average)	$\bar{y}$	$\dfrac{\sigma}{\sqrt{n}}$	$\dfrac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$ Difference between means (averages), independent samples	$\bar{y}_1 - \bar{y}_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$, $n_1 \geq 30$, $n_2 \geq 30$ $\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$, either $n_1 < 30$ or $n_2 < 30$ where $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
$\mu_d = \mu_1 - \mu_2$, Difference between means, paired samples	$\bar{y}_d$	$\sigma_d/\sqrt{n}$	$s_d/\sqrt{n}$
$\dfrac{\sigma_1^2}{\sigma_2^2}$ Ratio of variances	$\dfrac{s_1^2}{s_2^2}$	(not necessary)	(not necessary)

Supplementary Exercises

1.80. Using Tchebysheff's theorem. Tchebysheff's theorem states that at least $1 - (1/K^2)$ of a set of measurements will lie within K standard deviations of the mean of the data set. Use Tchebysheff's theorem to find the fraction of a set of measurements that will lie within:

(a) 2 standard deviations of the mean ($K = 2$)
(b) 3 standard deviations of the mean
(c) 1.5 standard deviations of the mean

1.81. Computing descriptive statistics. For each of the following data sets, compute $\bar{y}$, s^2, and s.

(a) 11, 2, 2, 1, 9 (b) 22, 9, 21, 15
(c) 1, 0, 1, 10, 11, 11, 0 (d) 4, 4, 4, 4

1.82. Normal probabilities. Use Table 1 of Appendix D to find each of the following:

(a) $P(z \geq 2)$ (b) $P(z \leq -2)$ (c) $P(z \geq -1.96)$
(d) $P(z \geq 0)$ (e) $P(z \leq -.5)$ (f) $P(z \leq -1.96)$

1.83. Finding z-scores. Suppose the random variable y has mean $\mu = 30$ and standard deviation $\sigma = 5$. How many standard deviations away from the mean of y is each of the following y values?

(a) $y = 10$ (b) $y = 32.5$ (c) $y = 30$ (d) $y = 60$

1.84. Deep-hole drilling for oil. "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). Some important variables in the drilling process are described here. Identify the data type for each variable.

(a) Chip discharge rate (number of chips discarded per minute)
(b) Drilling depth (millimeters)
(c) Oil velocity (millimeters per second)
(d) Type of drilling (single-edge, BTA, or ejector)
(e) Quality of hole surface

1.85. Finding misplaced items. Are men or women more adept at remembering where they leave misplaced items (like car keys)? According to University of Florida psychology professor Robin West, women show greater competence in actually finding these objects (*Explore*, Fall 1998). Approximately 300 men and women from Gainesville, Florida, participated in a study in which each person placed 20 common objects in a 12-room "virtual" house represented on a computer screen. Thirty minutes later, the subjects were asked to recall where they put each of the objects. For each object, a recall variable was measured as "yes" or "no."

(a) Identify the population of interest to the psychology professor.
(b) Identify the sample.
(c) Does the study involve descriptive or inferential statistics? Explain.
(d) Are the variables measured in the study quantitative or qualitative?

1.86. Salient roles of older adults. In *Psychology and Aging* (December 2000), University of Michigan School of Public Health researchers studied the roles that older adults feel are the most important to them in late life. The accompanying table summarizes the most salient roles identified by each in a national sample of 1,102 adults, 65 years or older.

MOST SALIENT ROLE	NUMBER
Spouse	424
Parent	269
Grandparent	148
Other relative	59
Friend	73
Homemaker	59
Provider	34
Volunteer, club, church member	36
Total	1,102

Source: Krause, N., and Shaw, B.A. "Role-specific feelings of control and mortality," *Psychology and Aging*, Vol. 15, No. 4, Table 2, Copyright © 2000, American Psychological Association, reprinted with permission.

(a) Describe the qualitative variable summarized in the table. Give the categories associated with the variable.
(b) Are the numbers in the table frequencies or relative frequencies?
(c) Display the information in the table in a bar graph.
(d) Which role is identified by the highest percentage of older adults? Interpret the relative frequency associated with this role.

1.87. Ancient Greek pottery. Archaeologists excavating the ancient Greek settlement at Phylakopi classified the pottery found in trenches (*Chance*, Fall 2000). The next table describes the collection of 837 pottery pieces uncovered in a particular layer at the excavation site. Construct and interpret a graph that will aid the archaeologists in understanding the distribution of the pottery types found at the site.

Table for Exercise 1.87

POT CATEGORY	NUMBER FOUND
Burnished	133
Monochrome	460
Slipped	55
Painted in curvilinear decoration	14
Painted in geometric decoration	165
Painted in naturalistic decoration	4
Cycladic white clay	4
Conical cup clay	2
Total	837

Source: Berg, I., and Bliedon, S. "The pots of Phylakopi: Applying statistical techniques to archaeology," *Chance*, Vol. 13, No. 4, Fall 2000.

OILSPILL

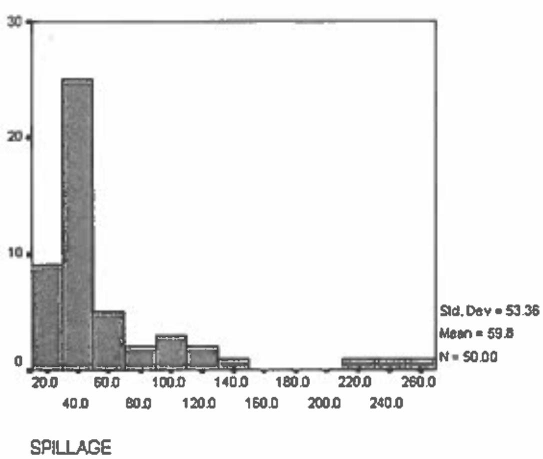
1.88. Tanker oil spills. Owing to several major ocean oil spills by tank vessels, Congress passed the 1990 Oil Pollution Act, which requires all tankers to be designed with thicker hulls. Further improvements in the structural design of a tank vessel have been proposed since then, each with the objective of reducing the likelihood of an oil spill and decreasing the amount of outflow in the event of a hull puncture. To aid in this development, *Marine Technology* (January 1995) reported on the spillage amount and cause of puncture for 50 recent major oil spills from tankers and carriers. The data are saved in the OILSPILL file.

(a) Use a graphical method to describe the cause of oil spillage for the 50 tankers.

(b) Does the graph, part a, suggest that any one cause is more likely to occur than any other? How is this information of value to the design engineers?

(c) Now consider the data on spillage amounts (in thousands of metric tons) for 50 major oil spills. An SPSS histogram for the 50 spillage amounts is shown below. Interpret the histogram.

(d) Descriptive statistics for the 50 spillage amounts are also shown on the SPSS histogram. Use this information to form an interval that can be used to predict the spillage amount for the next major oil spill.



1.89. Eclipses of Saturnian satellites. Saturn has five satellites that rotate around the planet. *Astronomy* (August 1995) lists 19 different events involving eclipses or occults of Saturnian satellites during the month of August. For each event, the percent of light lost by the eclipsed or occulted satellite at midevent is recorded in the table.

SATURN

DATE	EVENT	LIGHT LOSS (%)
Aug. 2	Eclipse	65
4	Eclipse	61
5	Occult	1
6	Eclipse	56
8	Eclipse	46
8	Occult	2
9	Occult	9
11	Occult	5
12	Occult	39
14	Occult	1
14	Eclipse	100
15	Occult	5
15	Occult	4
16	Occult	13
20	Occult	11
23	Occult	3
23	Occult	20
25	Occult	20
28	Occult	12

Source: Astronomy magazine, Aug. 1995, p. 60.

(a) Construct a stem-and-leaf display for light loss percentage of the 19 events.

(b) Locate on the stem-and-leaf plot, part a, the light losses associated with eclipses of Saturnian satellites. (Circle the light losses on the plot.)

(c) Based on the marked stem-and-leaf display, part b, make an inference about which event type (eclipse or occult) is more likely to lead to a greater light loss.

1.90. Comparing voltage readings. Refer to the data on process voltage readings at two locations, Exercise 1.70. Use the SAS printout for Exercise 1.70 (p. 63) and the rule of thumb to compare the voltage reading distributions for the two locations.

1.91. Normal probabilities. The random variable y has a normal distribution with $\mu = 80$ and $\sigma = 10$. Find the following probabilities:

(a) $P(y \leq 75)$ (b) $P(y \geq 90)$ (c) $P(60 \leq y \leq 70)$
(d) $P(y \geq 75)$ (e) $P(y = 75)$ (f) $P(y \leq 105)$

1.92. Fluid loss in crab spiders. A group of University of Virginia biologists studied nectivory (nectar drinking) in crab spiders to determine if adult males were feeding on nectar to prevent fluid loss (*Animal Behavior*, June 1995). Nine male spiders were weighed and then placed on the flowers of Queen Anne's lace. One hour later, the spiders were removed and reweighed. The evaporative fluid loss (in milligrams) of each of the nine male spiders is given in the table.

SPIDERS

MALE SPIDER	FLUID LOSS
A	.018
B	.020
C	.017
D	.024
E	.020
F	.024
G	.003
H	.001
I	.009

Source: Reprinted from *Animal Behaviour*, Vol. 49, Issue 6, Simon D. Pollard, Mike W. Beck, and Gary N. Dodson, "Why do male crab spiders drink nectar?" p. 1445 (Table II), Copyright © 1995, with permission from Elsevier.

(a) Summarize the fluid losses of male crab spiders with a stem-and-leaf display.
(b) Of the nine spiders, only three drank any nectar from the flowers of Queen Anne's lace. These three spiders are identified as G, H, and I in the table. Locate and circle these three fluid losses on the stem-and-leaf display. Does the pattern depicted in the graph give you any insight into whether feeding on flower nectar reduces evaporative fluid loss for male crab spiders? Explain.

MTBE

1.93. Groundwater contamination in wells. Refer to the *Environmental Science and Technology* (January 2005) study of the factors related to MTBE contamination in 223 New Hampshire wells, Exercise 1.17 (p. 12). The data are saved in the MTBE file. Two of the many quantitative variables measured for each well are the pH level (standard units) and the MTBE level (micrograms per liter).

(a) Construct a histogram for the pH levels of the sampled wells. From the histogram, estimate the proportion of wells with pH values less than 7.0.
(b) For those wells with detectible levels of MTBE, construct a histogram for the MTBE values. From the histogram, estimate the proportion of contaminated wells with MTBE values that exceed 5 micrograms per liter.
(c) Find the mean and standard deviation for the pH levels of the sampled wells, and construct the interval $\bar{y} \pm 2s$. Estimate the percentage of wells with pH levels that fall within the interval. What rule did you apply to obtain the estimate? Explain.
(d) Find the mean and standard deviation for the MTBE levels of the sampled wells and construct the interval $\bar{y} \pm 2s$. Estimate the percentage of wells with MTBE levels that fall within the interval. What rule did you apply to obtain the estimate? Explain.

1.94. Dental Anxiety Scale. Psychology students at Wittenberg University completed the Dental Anxiety Scale questionnaire (*Psychological Reports*, August 1997). Scores on the scale range from 0 (no anxiety) to 20 (extreme anxiety). The mean score was 11 and the standard deviation was 3.5. Assume that the distribution of all scores on the Dental Anxiety Scale is normal with $\mu = 11$ and $\sigma = 3.5$.

(a) Suppose you score a 16 on the Dental Anxiety Scale. Find the z-value for this score.
(b) Find the probability that someone scores between a 10 and a 15 on the Dental Anxiety Scale.
(c) Find the probability that someone scores above a 17 on the Dental Anxiety Scale.

1.95. Improving SAT scores. Refer to the *Chance* (Winter 2001) study of students who paid a private tutor to help them improve their Standardized Admission Test (SAT) scores, Exercise 1.31 (p. 25). The table summarizing the changes in both the SAT-Mathematics and SAT-Verbal scores for these students is reproduced here. Assume that both distributions of SAT score changes are approximately normal.

	SAT-MATH	SAT-VERBAL
Mean change in score	19	7
Standard deviation of score changes	65	49

(a) What is the probability that a student increases his or her score on the SAT-Math test by at least 50 points?
(b) What is the probability that a student increases his or her score on the SAT-Verbal test by at least 50 points?

1.96. Fitness of cardiac patients. The physical fitness of a patient is often measured by the patient's maximum oxygen uptake (recorded in milliliters per kilogram, ml/kg). The mean maximum oxygen uptake for cardiac patients who regularly participate in sports or exercise programs was found to be 24.1 with a standard deviation of 6.30 (*Adapted Physical Activity Quarterly*, October 1997). Assume this distribution is approximately normal.

(a) What is the probability that a cardiac patient who regularly participates in sports has a maximum oxygen uptake of at least 20 ml/kg?
(b) What is the probability that a cardiac patient who regularly exercises has a maximum oxygen uptake of 10.5 ml/kg or lower?
(c) Consider a cardiac patient with a maximum oxygen uptake of 10.5. Is it likely that this patient participates regularly in sports or exercise programs? Explain.

1.97. Susceptibility to hypnosis. The Computer-Assisted Hypnosis Scale (CAHS) is designed to measure a person's susceptibility to hypnosis. In computer-assisted hypnosis, the computer serves as a facilitator of hypnosis by using digitized speech processing coupled with interactive involvement with the hypnotic subject. CAHS scores range from 0 (no susceptibility) to 12 (extremely high susceptibility). A study in *Psychological Assessment* (March 1995) reported a mean CAHS score of 4.59 and a standard deviation of 2.95 for University of Tennessee undergraduates. Assume that $\mu = 4.29$ and $\sigma = 2.95$ for this population. Suppose a psychologist uses CAHS to test a random sample of 50 subjects.

(a) Would you expect to observe a sample mean CAHS score of $\bar{y} = 6$ or higher? Explain.
(b) Suppose the psychologist actually observes $\bar{y} = 6.2$. Based on your answer to part a, make an inference about the population from which the sample was selected.

1.98. Data on postmortem intervals. In Exercise 1.20 (p. 17) you learned that postmortem interval (PMI) is the elapsed time between death and the performance of an autopsy on the cadaver. *Brain and Language* (June 1995) reported on the PMIs of 22 randomly selected human brain specimens obtained at autopsy. The data are reproduced in the table below.

BRAINPMI

5.5	14.5	6.0	5.5	5.3	5.8	11.0	6.1
7.0	14.5	10.4	4.6	4.3	7.2	10.5	6.5
3.3	7.0	4.1	6.2	10.4	4.9		

Source: Reprinted from *Brain and Language,* Vol. 49, Issue 3, T. L. Hayes and D. A. Lewis, "Anatomical Specialization of the Anterior Motor Speech Area: Hemispheric Differences in Magnopyramidal Neurons," p. 292 (Table 1), Copyright © 1995, with permission of Elsevier.

(a) Construct a 95% confidence interval for the true mean PMI of human brain specimens obtained at autopsy.
(b) Interpret the interval, part a.
(c) What assumption is required for the interval, part a, to be valid? Is this assumption satisfied? Explain.
(d) What is meant by the phrase "95% confidence"?

1.99. Pigeon diet study. The *Australian Journal of Zoology* (Vol. 43, 1995) reported on a study of the diets and water requirements of spinifex pigeons. Sixteen pigeons were captured in the desert and the crop (i.e., stomach) contents of each examined. The accompanying table reports the weight (in grams) of dry seed in the crop of each pigeon. Use the SAS printout below to find a 99% confidence interval for the average weight of dry seeds in the crops of spinifex pigeons inhabiting the Western Australian desert. Interpret the result.

PIGEONS

.457	3.751	.238	2.967	2.509	1.384	1.454	.818
.335	1.436	1.603	1.309	.201	.530	2.144	.834

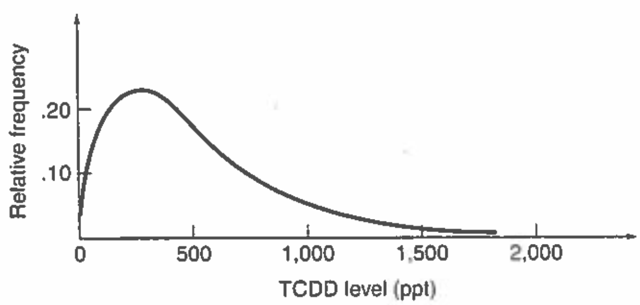
Source: Table 2 from Williams, J. B., Bradshaw, D., and Schmidt, L. "Field metabolism and water requirements of spinifex pigeons *(Geophaps plumifera)* in Western Australia. "*Australian Journal of Zoology,* Vol. 43, no. 1, 1995, p. 7. Reprinted by permission of CSIRO Publishing. http://www.publish.csiro.au/nid/90/paper/ZO9950001.htm.

```
Sample Statistics for SEEDWT

     N          Mean        Std. Dev.       Std. Error
---------------------------------------------------------
    16          1.37          1.03             0.26

Hypothesis Test

   Null hypothesis:     Mean of SEEDWT =  0
   Alternative:         Mean of SEEDWT ^= 0

           t Statistic      Df       Prob > t
           ----------------------------------
              5.312         15        <.0001

99 % Confidence Interval for the Mean

      Lower Limit:             0.61
      Upper Limit:             2.13
```

1.100. Psychological stress in victims of violence. Interpersonal violence (e.g., rape) generally leads to psychological stress for the victim. *Clinical Psychology Review* (Vol. 15, 1995) reported on the results of all recently published studies of the relationship between interpersonal violence and psychological stress. The distribution of the time elapsed between the violent incident and the initial sign of stress has a mean of 5.1 years and a standard deviation of 6.1 years. Consider a random sample of $n = 150$ victims of interpersonal violence. Let $\bar{y}$ represent the mean time elapsed between the violent act and the first sign of stress for the sampled victims.

(a) Give the mean and standard deviation of the sampling distribution of $\bar{y}$.
(b) Will the sampling distribution of $\bar{y}$ be approximately normal? Explain.
(c) Find $P(\bar{y} > 5.5)$.
(d) Find $P(4 < \bar{y} < 5)$.

1.101. Workers exposed to dioxin. The Occupational Safety and Health Administration (OSHA) conducted a study to evaluate the level of exposure of workers to the dioxin 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.

1.102. Alkalinity levels in river water. The mean alkalinity level of water specimens collected from the Han River in Seoul, Korea, is 50 milligrams per liter (*Environmental Science and Engineering*, September 1, 2000). Consider a random sample of 100 water specimens collected from a tributary of the Han River. Suppose the mean and standard deviation of the alkalinity levels for the sample are $\bar{y} = 67.8$ mpl and $s = 14.4$ mpl. Is there sufficient evidence (at $\alpha = .01$) to indicate that the population mean alkalinity level of water in the tributary exceeds 50 mpl?

1.103. Temperature of molten iron. The Cleveland Casting Plant produces iron automotive castings for Ford Motor Company. When the process is stable, the target pouring temperature of the molten iron is 2,550 degrees (*Quality Engineering*, Vol. 7, 1995). The pouring temperatures (in degrees Fahrenheit) for a random sample of 10 crankshafts produced at the plant are listed in the table below. Conduct a test to determine whether the true mean pouring temperature differs from the target setting. Test using $\alpha = .01$.

IRONTEMP

2,543	2,541	2,544	2,620	2,560	2,559	2,562
2,553	2,552	2,553				

Source: Price, B., and Barth, B. "A structural model relating process inputs and final product characteristics," *Quality Engineering*, Vol. 7, No. 4, 1995, p. 696 (Table 2).

1.104. Mating habits of snails. *Genetical Research* (June 1995) published a study of the mating habits of hermaphroditic snails. The mating habits of the snails were identified as either self-fertilizing or cross-fertilizing. The effective population sizes of the two groups were compared. The data for the study are summarized in the table. Geneticists are interested in comparing the variation in population size of the two types of mating systems. Conduct this analysis for the researcher. Interpret the result.

	EFFECTIVE POPULATION SIZE		
SNAIL MATING SYSTEM	SAMPLE SIZE	MEAN	STANDARD DEVIATION
Cross-fertilizing	17	4,894	1,932
Self-fertilizing	5	4,133	1,890

Source: Jarne, P. "Mating system, bottlenecks, and genetic polymorphism in hermaphroditic animals." *Genetics Research*, Vol. 65, No. 3, June 1995, p. 197 (Table 4). 2009 © Cambridge Journals, reproduced with permission.

1.105. Heights of children who repeat grades. Are children who repeat a grade in elementary school shorter on average than their peers? To answer this question, researchers compared the heights of Australian schoolchildren who repeated a grade to those who did not (*Archives of Disease in Childhood*, April 2000). All height measurements were standardized using z-scores. A summary of the results, by gender, is shown in the table on p. 77.

(a) Conduct a test of hypothesis to determine whether the average height of Australian boys who repeated a grade is less than the average height of boys who never repeated. Use $\alpha = .05$.
(b) Repeat part a for Australian girls.
(c) Summarize the results of the hypothesis tests in the words of the problem.

Summary table for Exercise 1.105

	NEVER REPEATED	REPEATED A GRADE
Boys	$n = 1{,}349$ $\bar{x} = .30$ $s = .97$	$n = 86$ $\bar{x} = -.04$ $s = 1.17$
Girls	$n = 1{,}366$ $\bar{x} = .22$ $s = 1.04$	$n = 43$ $\bar{x} = .26$ $s = .94$

Source: Reproduced from *Archives of Disease in Childhood,* "Does height influence progression through primary school grades?" Melissa Wake, David Coghlan, and Kylie Hesketh, Vol. 82, Issue 4, April 2000 (Table 3), with permission from BMJ Publishing Group Ltd.

1.106. College students attitudes toward parents. Researchers at the University of South Alabama compared the attitudes of male college students toward their fathers with their attitudes toward their mothers (*Journal of Genetic Psychology,* March 1998). Each of a sample of 13 males was asked to complete the following statement about each of their parents: My relationship with my father (mother) can best be described as: (1) awful, (2) poor, (3) average, (4) good, or (5) great. The following data were obtained:

FMATTITUDES

STUDENT	ATTITUDE TOWARD FATHER	ATTITUDE TOWARD MOTHER
1	2	3
2	5	5
3	4	3
4	4	5
5	3	4
6	5	4
7	4	5
8	2	4
9	4	5
10	5	4
11	4	5
12	5	4
13	3	3

Source: Adapted from Vitulli, W. F., and Richardson, D. K. "College student's attitudes toward relationships with parents: A five-year comparative analysis," *Journal of Genetic Psychology,* Vol. 159, No. 1 (March 1998), pp. 45–52.

(a) Specify the appropriate hypotheses for testing whether male students' attitudes toward their fathers differ from their attitudes toward their mothers, on average.

(b) Conduct the test of part a at $\alpha = .05$. Interpret the results in the context of the problem.

1.107. Mathematics and gender. On average, do males outperform females in mathematics? To answerthis question, psychologists at the University of Minnesota compared the scores of male and female eighth-grade students who took a basic skills mathematics achievement test (*American Educational Research Journal,* Fall 1998). One form of the test consisted of 68 multiple-choice questions. A summary of the test scores is displayed in the table.

	MALES	FEMALES
Sample size	1,764	1,739
Mean	48.9	48.4
Standard deviation	12.96	11.85

Source: Bielinski, J., and Davison, M. L. "Gender differences by item difficulty interactions in multiple-choice mathematics items," *American Educational Research Journal,* Vol. 35, No. 3, Fall 1998, p. 464 (Table 1). Reprinted by Permission of SAGE Publications.

(a) Is there evidence of a difference between the true mean mathematics test scores of male and female eighth-graders?

(b) Use a 90% confidence interval to estimate the true difference in mean test scores between males and females. Does the confidence interval support the result of the test you conducted in part a?

(c) What assumptions about the distributions of the populations of test scores are necessary to ensure the validity of the inferences you made in parts a and b?

(d) What is the observed significance level of the test you conducted in part a?

(e) The researchers hypothesized that the distribution of test scores for males is more variable than the distribution for females. Test this claim at $\alpha = .05$.

1.108. Visual search study. When searching for an item (e.g., a roadside traffic sign, a lost earring, or a tumor in a mammogram), common sense dictates that you will not reexamine items previously rejected. However, researchers at Harvard Medical School found that a visual search has no memory (*Nature,* August 6, 1998). In their experiment, nine subjects searched for the letter "T" mixed among several letters "L." Each subject conducted the search under two conditions: random and static. In the random condition, the location of the letters were changed every 111 milliseconds; in the static condition, the location of the letters remained unchanged. In each trial, the reaction time (i.e., the amount of time it took the subject to locate the target letter) was recorded in milliseconds.

(a) One goal of the research is to compare the mean reaction times of subjects in the two experimental conditions. Explain why the data should be analyzed as a paired-difference experiment.

(b) If a visual search has no memory, then the main reaction times in the two conditions will not differ. Specify H_0 and H_a for testing the "no memory" theory.

(c) The test statistic was calculated as $t = 1.52$ with p-value $= .15$. Make the appropriate conclusion.

MILK

1.109. **Detection of rigged school milk prices.** Each year, the state of Kentucky invites bids from dairies to supply half-pint containers of fluid milk products for its school districts. In several school districts in northern Kentucky (called the "tri-county" market), two suppliers—Meyer Dairy and Trauth Dairy—were accused of price-fixing, that is, conspiring to allocate the districts so that the winning bidder was predetermined and the price per pint was set above the competitive price. These two dairies were the only two bidders on the milk contracts in the "tri-county market" between 1983 and 1991. (In contrast, a large number of different dairies won the milk contracts for school districts in the remainder of the northern Kentucky market—called the "surrounding" market.) Did Meyer and Trauth conspire to rig their bids in the tri-county market? If so, economic theory states that the mean winning price in the rigged tri-county market will be higher than the mean winning price in the competitive surrounding market. Data on all bids received from the dairies competing for the milk contracts between 1983 and 1991 are saved in the MILK file.

(a) A MINITAB printout of the comparison of mean bid prices for whole white milk for the two Kentucky milk markets is shown below. Is there support for the claim that the dairies in the tri-county market participated in collusive practices? Explain in detail.

(b) In competitive sealed bid markets, vendors do not share information about their bids. Consequently, more dispersion or variability among the bids is typically observed than in collusive markets, where vendors communicate about their bids and have a tendency to submit bids in close proximity to one another in an attempt to make the bidding appear competitive. If collusion exists in the tri-county milk market, the variation in winning bid prices in the surrounding ("competitive") market will be significantly larger than the corresponding variation in the tri-county ("rigged") market. A MINITAB analysis of the whole white milk data in the MILK file yielded the

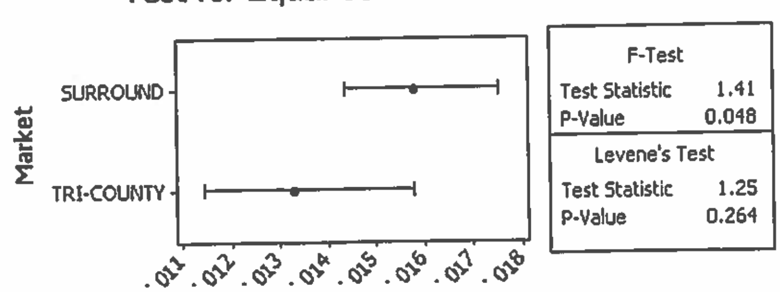
MINITAB Output for Exercise 1.109(a)

Two-Sample T-Test and CI: WWBID, Market

```
Two-sample T for WWBID

Market        N    Mean   StDev  SE Mean
SURROUND    254  0.1331  0.0158  0.00099
TRI-COUNTY  100  0.1431  0.0133   0.0013


Difference = mu (SURROUND) - mu (TRI-COUNTY)
Estimate for difference:  -0.009970
95% upper bound for difference:  -0.007232
T-Test of difference = 0 (vs <): T-Value = -6.02  P-Value = 0.000  DF = 213
```

MINITAB Output for Exercise 1.109(b)

Test for Equal Variances for WWBID



printout at the bottom of p. 78. Is there evidence that the bid price variance for the surrounding market exceeds the bid price variance for the tri-county market?

1.110. **Alzheimer's and homophone spelling.** A *homophone* is a word whose pronunciation is the same as that of another word having a different meaning and spelling (e.g., *nun* and *none*, *doe* and *dough*, etc.). *Brain and Language* (April 1995) reported on a study of homophone spelling in patients with Alzheimer's disease. Twenty Alzheimer's patients were asked to spell 24 homophone pairs given in random order, then the number of homophone confusions (e.g., spelling *doe* given the context, *bake bread dough*) was recorded for each patient. One year later, the same test was given to the same patients. The data for the study are provided in the table. The researchers posed the following question: "Do Alzheimer's patients show a significant increase in mean homophone confusion errors over time?" Perform an analysis of the data to answer the researchers' question. Use the relevant information in the SAS printout. What assumptions are necessary for the procedure used to be valid? Are they satisfied?

```
                    Two Sample Paired t-test for the Means of TIME1 and TIME2

Sample Statistics

     Group          N       Mean     Std. Dev.    Std. Error
     -------------------------------------------------------
     TIME1         20       4.15      3.4985        0.7823
     TIME2         20        5.8      4.2128         0.942

Hypothesis Test

     Null hypothesis:    Mean of (TIME1 - TIME2) => 0
     Alternative:        Mean of (TIME1 - TIME2) <  0

          t Statistic      Df      Prob > t
          ---------------------------------
            -2.306         19       0.0163

95% Confidence Interval for the Difference between Two Paired Means

                    Lower Limit    Upper Limit
                    -----------    -----------
                       -3.15          -0.15
```

HOMOPHONE

PATIENT	TIME 1	TIME 2
1	5	5
2	1	3
3	0	0
4	1	1
5	0	1
6	2	1
7	5	6
8	1	2
9	0	9
10	5	8
11	7	10
12	0	3
13	3	9
14	5	8
15	7	12
16	10	16
17	5	5
18	6	3
19	9	6
20	11	8

Source: Neils, J., Roeltgen, D. P., and Constantinidou, F. "Decline in homophone spelling associated with loss of semantic influence on spelling in Alzheimer's disease," *Brain and Language*, Vol. 49, No. 1, Apr. 1995, p. 36 (Table 3). Copyright © 1995, with permission from Elsevier.

References

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

Mcclave, J. T., and Sincich, T. *A First Course in Statistics*, 10th ed. Upper Saddle River, N.J.: Pearson Prentice Hall, 2009.

Mendenhall, W., Beaver, R. J., and Beaver, B. M. *Introduction to Probability and Statistics*, 12th ed. N. Scituate: Duxbury, 2006.

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*. San Francisco: Holden-Day, 1989.

Tukey, J. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley, 1977.