CHAPTER

9

Hypothesis Tests for One Population Mean

CHAPTER OUTLINE

CHAPTER OBJECTIVES

In Chapter 8, we examined methods for obtaining confidence intervals for one population mean. We know that a confidence interval for a population mean, μ, is based on a sample mean, $\bar{x}$. Now we show how that statistic can be used to make decisions about hypothesized values of a population mean.

For example, suppose that we want to decide whether the mean prison sentence, μ, of all people imprisoned last year for drug offenses exceeds the year 2000 mean of 75.5 months. To make that decision, we can take a random sample of people imprisoned last year for drug offenses, compute their sample mean sentence, $\bar{x}$, and then apply a statistical-inference technique called a *hypothesis test.*

In this chapter, we describe hypothesis tests for one population mean. In doing so, we consider three different procedures. The first two are called the *one-mean z-test* and the *one-mean t-test,* which are the hypothesis-test analogues of the one-mean z-interval and one-mean t-interval confidence-interval procedures, respectively, discussed in Chapter 8. The third is a nonparametric method called the *Wilcoxon signed-rank test,* which applies when the variable under consideration has a symmetric distribution.

We also examine two different approaches to hypothesis testing—namely, the critical-value approach and the *P*-value approach.

CASE STUDY

Gender and Sense of Direction

Many of you have been there, a classic scene: mom yelling at dad to turn left, while dad decides to do just the opposite. Well, who made the right call? More generally, who has a better sense of direction, women or men?

Dr. J. Sholl et al. considered these and related questions in the paper "The Relation of Sex and Sense of Direction to Spatial Orientation in an Unfamiliar Environment" (*Journal of Environmental Psychology*, Vol. 20, pp. 17–28).

In their study, the spatial orientation skills of 30 male students and 30 female students from Boston College were challenged in Houghton Garden Park, a wooded park near campus in Newton, Massachusetts. Before driving to the park, the participants were asked to rate their own sense of direction as either good or poor.

In the park, students were instructed to point to predesignated landmarks and also to the direction of south. Pointing was carried out by students moving a pointer attached to a 360° protractor; the angle of the pointing response was then recorded to the nearest degree. For the female students who had rated their sense of direction to be good, the following table displays the pointing errors (in degrees) when they attempted to point south.

14	122	128	109	12
91	8	78	31	36
27	68	20	69	18

Based on these data, can you conclude that, in general, women who consider themselves to have a good sense of direction really do better, on average, than they would by randomly guessing at the direction of south? To answer that question, you need to conduct a hypothesis test, which you will do after you study hypothesis testing in this chapter.

9.1 The Nature of Hypothesis Testing

We often use inferential statistics to make decisions or judgments about the value of a parameter, such as a population mean. For example, we might need to decide whether the mean weight, μ, of all bags of pretzels packaged by a particular company differs from the advertised weight of 454 grams (g), or we might want to determine whether the mean age, μ, of all cars in use has increased from the year 2000 mean of 9.0 years.

One of the most commonly used methods for making such decisions or judgments is to perform a *hypothesis test.* A **hypothesis** is a statement that something is true. For example, the statement "the mean weight of all bags of pretzels packaged differs from the advertised weight of 454 g" is a hypothesis.

Typically, a hypothesis test involves two hypotheses: the *null hypothesis* and the *alternative hypothesis* (or *research hypothesis*), which we define as follows.

DEFINITION 9.1

What Does It Mean?

Originally, the word *null* in *null hypothesis* stood for "no difference" or "the difference is null." Over the years, however, *null hypothesis* has come to mean simply a hypothesis to be tested.

Null and Alternative Hypotheses; Hypothesis Test

Null hypothesis: A hypothesis to be tested. We use the symbol H_0 to represent the null hypothesis.

Alternative hypothesis: A hypothesis to be considered as an alternative to the null hypothesis. We use the symbol H_a to represent the alternative hypothesis.

Hypothesis test: The problem in a hypothesis test is to decide whether the null hypothesis should be rejected in favor of the alternative hypothesis.

For instance, in the pretzel packaging example, the null hypothesis might be "the mean weight of all bags of pretzels packaged equals the advertised weight of 454 g," and the alternative hypothesis might be "the mean weight of all bags of pretzels packaged differs from the advertised weight of 454 g."

Choosing the Hypotheses

The first step in setting up a hypothesis test is to decide on the null hypothesis and the alternative hypothesis. The following are some guidelines for choosing these two hypotheses. Although the guidelines refer specifically to hypothesis tests for one population mean, μ, they apply to any hypothesis test concerning one parameter.

Null Hypothesis

In this book, the null hypothesis for a hypothesis test concerning a population mean, μ, always specifies a single value for that parameter. Hence we can express the null hypothesis as

$$H_0: \mu = \mu_0,$$

where μ_0 is some number.

Alternative Hypothesis

The choice of the alternative hypothesis depends on and should reflect the purpose of the hypothesis test. Three choices are possible for the alternative hypothesis.

- If the primary concern is deciding whether a population mean, μ, is *different from* a specified value μ_0, we express the alternative hypothesis as

  $$H_a: \mu \neq \mu_0.$$

  A hypothesis test whose alternative hypothesis has this form is called a **two-tailed test.**
- If the primary concern is deciding whether a population mean, μ, is *less than* a specified value μ_0, we express the alternative hypothesis as

  $$H_a: \mu < \mu_0.$$

  A hypothesis test whose alternative hypothesis has this form is called a **left-tailed test.**
- If the primary concern is deciding whether a population mean, μ, is *greater than* a specified value μ_0, we express the alternative hypothesis as

  $$H_a: \mu > \mu_0.$$

  A hypothesis test whose alternative hypothesis has this form is called a **right-tailed test.**

A hypothesis test is called a **one-tailed test** if it is either left tailed or right tailed.

EXAMPLE 9.1 Choosing the Null and Alternative Hypotheses

Quality Assurance A snack-food company produces a 454-g bag of pretzels. Although the actual net weights deviate slightly from 454 g and vary from one bag to another, the company insists that the mean net weight of the bags be 454 g.

As part of its program, the quality assurance department periodically performs a hypothesis test to decide whether the packaging machine is working properly, that is, to decide whether the mean net weight of all bags packaged is 454 g.

a. Determine the null hypothesis for the hypothesis test.
b. Determine the alternative hypothesis for the hypothesis test.
c. Classify the hypothesis test as two tailed, left tailed, or right tailed.

Solution Let μ denote the mean net weight of all bags packaged.

a. The null hypothesis is that the packaging machine is working properly, that is, that the mean net weight, μ, of all bags packaged *equals* 454 g. In symbols, H_0: $\mu = 454$ g.
b. The alternative hypothesis is that the packaging machine is not working properly, that is, that the mean net weight, μ, of all bags packaged is *different from* 454 g. In symbols, H_a: $\mu \neq 454$ g.
c. This hypothesis test is two tailed because a does-not-equal sign ($\neq$) appears in the alternative hypothesis.

EXAMPLE 9.2 Choosing the Null and Alternative Hypotheses

Prices of History Books The R. R. Bowker Company collects information on the retail prices of books and publishes the data in *The Bowker Annual Library and Book Trade Almanac*. In 2005, the mean retail price of history books was \$78.01. Suppose that we want to perform a hypothesis test to decide whether this year's mean retail price of history books has increased from the 2005 mean.

a. Determine the null hypothesis for the hypothesis test.
b. Determine the alternative hypothesis for the hypothesis test.
c. Classify the hypothesis test as two tailed, left tailed, or right tailed.

Solution Let μ denote this year's mean retail price of history books.

a. The null hypothesis is that this year's mean retail price of history books *equals* the 2005 mean of \$78.01; that is, H_0: $\mu = \$78.01$.
b. The alternative hypothesis is that this year's mean retail price of history books is *greater than* the 2005 mean of \$78.01; that is, H_a: $\mu > \$78.01$.
c. This hypothesis test is right tailed because a greater-than sign (>) appears in the alternative hypothesis.

EXAMPLE 9.3 Choosing the Null and Alternative Hypotheses

Poverty and Dietary Calcium Calcium is the most abundant mineral in the human body and has several important functions. Most body calcium is stored in the bones and teeth, where it functions to support their structure. Recommendations for calcium are provided in *Dietary Reference Intakes*, developed by the Institute of Medicine of the National Academy of Sciences. The recommended adequate intake (RAI) of calcium for adults (ages 19–50 years) is 1000 milligrams (mg) per day.

Suppose that we want to perform a hypothesis test to decide whether the average adult with an income below the poverty level gets less than the RAI of 1000 mg.

a. Determine the null hypothesis for the hypothesis test.
b. Determine the alternative hypothesis for the hypothesis test.
c. Classify the hypothesis test as two tailed, left tailed, or right tailed.

Solution Let μ denote the mean calcium intake (per day) of all adults with incomes below the poverty level.

a. The null hypothesis is that the mean calcium intake of all adults with incomes below the poverty level *equals* the RAI of 1000 mg per day; that is, H_0: $\mu = 1000$ mg.
b. The alternative hypothesis is that the mean calcium intake of all adults with incomes below the poverty level is *less than* the RAI of 1000 mg per day; that is, H_a: $\mu < 1000$ mg.
c. This hypothesis test is left tailed because a less-than sign (<) appears in the alternative hypothesis.

You try it!

Exercise 9.5 on page 364

The Logic of Hypothesis Testing

After we have chosen the null and alternative hypotheses, we must decide whether to reject the null hypothesis in favor of the alternative hypothesis. The procedure for deciding is roughly as follows.

Basic Logic of Hypothesis Testing

Take a random sample from the population. If the sample data are consistent with the null hypothesis, do not reject the null hypothesis; if the sample data are inconsistent with the null hypothesis and supportive of the alternative hypothesis, reject the null hypothesis in favor of the alternative hypothesis.

In practice, of course, we must have a precise criterion for deciding whether to reject the null hypothesis. We discuss such criteria in Sections 9.2 and 9.3. At this point, we simply note that a precise criterion involves a **test statistic,** a statistic calculated from the data that is used as a basis for deciding whether the null hypothesis should be rejected.

Type I and Type II Errors

Any decision we make based on a hypothesis test may be incorrect because we have used partial information obtained from a sample to draw conclusions about the entire population. There are two types of incorrect decisions—*Type I error* and *Type II error*, as indicated in Table 9.1 and Definition 9.2.

TABLE 9.1
Correct and incorrect decisions for a hypothesis test

		H_0 is:	
		True	False
Decision:	Do not reject H_0	Correct decision	Type II error
	Reject H_0	Type I error	Correct decision

DEFINITION 9.2

Type I and Type II Errors

Type I error: Rejecting the null hypothesis when it is in fact true.

Type II error: Not rejecting the null hypothesis when it is in fact false.

EXAMPLE 9.4 Type I and Type II Errors

Quality Assurance Consider again the pretzel-packaging hypothesis test. The null and alternative hypotheses are, respectively,

$$H_0: \mu = 454 \text{ g (the packaging machine is working properly)}$$
$$H_a: \mu \neq 454 \text{ g (the packaging machine is not working properly),}$$

where μ is the mean net weight of all bags of pretzels packaged. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis $\mu = 454$ g, that is, to the conclusion that $\mu \neq 454$ g. Classify that conclusion by error type or as a correct decision if

d. the mean net weight, μ, is in fact 454 g.
e. the mean net weight, μ, is in fact not 454 g.

Solution

a. A Type I error occurs when a true null hypothesis is rejected. In this case, a Type I error would occur if in fact $\mu = 454$ g but the results of the sampling lead to the conclusion that $\mu \neq 454$ g.

Interpretation A Type I error occurs if we conclude that the packaging machine is not working properly when in fact it is working properly.

b. A Type II error occurs when a false null hypothesis is not rejected. In this case, a Type II error would occur if in fact $\mu \neq 454$ g but the results of the sampling fail to lead to that conclusion.

Interpretation A Type II error occurs if we fail to conclude that the packaging machine is not working properly when in fact it is not working properly.

c. A correct decision can occur in either of two ways.

- A true null hypothesis is not rejected. That would happen if in fact $\mu = 454$ g and the results of the sampling do not lead to the rejection of that fact.
- A false null hypothesis is rejected. That would happen if in fact $\mu \neq 454$ g and the results of the sampling lead to that conclusion.

Interpretation A correct decision occurs if either we fail to conclude that the packaging machine is not working properly when in fact it is working properly, or we conclude that the packaging machine is not working properly when in fact it is not working properly.

d. If in fact $\mu = 454$ g, the null hypothesis is true. Consequently, by rejecting the null hypothesis $\mu = 454$ g, we have made a Type I error—we have rejected a true null hypothesis.

e. If in fact $\mu \neq 454$ g, the null hypothesis is false. Consequently, by rejecting the null hypothesis $\mu = 454$ g, we have made a correct decision—we have rejected a false null hypothesis.

You try it!

Exercise 9.21 on page 365

Probabilities of Type I and Type II Errors

Part of evaluating the effectiveness of a hypothesis test involves analyzing the chances of making an incorrect decision. A Type I error occurs if a true null hypothesis is rejected. The probability of that happening, the **Type I error probability,** commonly called the **significance level** of the hypothesis test, is denoted $\boldsymbol{\alpha}$ (the lowercase Greek letter alpha).

DEFINITION 9.3

Significance Level

The probability of making a Type I error, that is, of rejecting a true null hypothesis, is called the **significance level, $\boldsymbol{\alpha}$,** of a hypothesis test.

A Type II error occurs if a false null hypothesis is not rejected. The probability of that happening, the **Type II error probability,** is denoted $\boldsymbol{\beta}$ (the lowercase Greek letter beta). Calculation of Type II error probabilities is examined in Section 9.7.

Ideally, both Type I and Type II errors should have small probabilities. Then the chance of making an incorrect decision would be small, regardless of whether the null hypothesis is true or false. As we soon demonstrate, we can design a hypothesis test to have any specified significance level. So, for instance, if not rejecting a true null hypothesis is important, we should specify a small value for α. However, in making our choice for α, we must keep Key Fact 9.1 in mind.

KEY FACT 9.1

Relation between Type I and Type II Error Probabilities

For a fixed sample size, the smaller we specify the significance level, α, the larger will be the probability, β, of not rejecting a false null hypothesis.

Consequently, we must always assess the risks involved in committing both types of errors and use that assessment as a method for balancing the Type I and Type II error probabilities.

Possible Conclusions for a Hypothesis Test

The significance level, α, is the probability of making a Type I error, that is, of rejecting a true null hypothesis. Therefore, if the hypothesis test is conducted at a small significance level (e.g., $\alpha = 0.05$), the chance of rejecting a true null hypothesis will be small. In this text, we generally specify a small significance level. Thus, if we do reject the null hypothesis, we can be reasonably confident that the null hypothesis is false. In other words, if we do reject the null hypothesis, we conclude that the data provide sufficient evidence to support the alternative hypothesis.

However, we usually do not know the probability, β, of making a Type II error, that is, of not rejecting a false null hypothesis. Consequently, if we do not reject the null hypothesis, we simply reserve judgment about which hypothesis is true. In other words, if we do not reject the null hypothesis, we conclude only that the data do not provide sufficient evidence to support the alternative hypothesis; we do not conclude that the data provide sufficient evidence to support the null hypothesis.

KEY FACT 9.2

Possible Conclusions for a Hypothesis Test

Suppose that a hypothesis test is conducted at a small significance level.

- If the null hypothesis is rejected, we conclude that the data provide sufficient evidence to support the alternative hypothesis.
- If the null hypothesis is not rejected, we conclude that the data do not provide sufficient evidence to support the alternative hypothesis.

When the null hypothesis is rejected in a hypothesis test performed at the significance level α, we frequently express that fact with the phrase "the test results are **statistically significant** at the α level." Similarly, when the null hypothesis is not rejected in a hypothesis test performed at the significance level α, we often express that fact with the phrase "the test results are **not statistically significant** at the α level."

Exercises 9.1

Understanding the Concepts and Skills

9.1 Explain the meaning of the term *hypothesis* as used in inferential statistics.

9.2 What role does the decision criterion play in a hypothesis test?

9.3 Suppose that you want to perform a hypothesis test for a population mean μ.

a. Express the null hypothesis both in words and in symbolic form.

b. Express each of the three possible alternative hypotheses in words and in symbolic form.

9.4 Suppose that you are considering a hypothesis test for a population mean, μ. In each part, express the alternative hypothesis symbolically and identify the hypothesis test as two tailed, left tailed, or right tailed.

a. You want to decide whether the population mean is different from a specified value μ_0.

b. You want to decide whether the population mean is less than a specified value μ_0.

c. You want to decide whether the population mean is greater than a specified value μ_0.

In Exercises ***9.5–9.13****, hypothesis tests are proposed. For each hypothesis test,*

a. determine the null hypothesis.

b. determine the alternative hypothesis.

c. classify the hypothesis test as two tailed, left tailed, or right tailed.

9.5 Toxic Mushrooms? Cadmium, a heavy metal, is toxic to animals. Mushrooms, however, are able to absorb and accumulate cadmium at high concentrations. The Czech and Slovak governments have set a safety limit for cadmium in dry vegetables at 0.5 part per million (ppm). M. Melgar et al. measured the cadmium levels in a random sample of the edible mushroom *Boletus pinicola* and published the results in the paper "Influence of Some Factors in Toxicity and Accumulation of Cd from Edible Wild Macrofungi in NW Spain" (*Journal of Environmental Science and Health*, Vol. B33(4), pp. 439–455). A hypothesis test is to be performed to decide whether the mean cadmium level in *Boletus pinicola* mushrooms is greater than the government's recommended limit.

9.6 Agriculture Books. The R. R. Bowker Company collects information on the retail prices of books and publishes the data in *The Bowker Annual Library and Book Trade Almanac*. In 2005, the mean retail price of agriculture books was \$57.61. A hypothesis test is to be performed to decide whether this year's mean retail price of agriculture books has changed from the 2005 mean.

9.7 Iron Deficiency? Iron is essential to most life forms and to normal human physiology. It is an integral part of many proteins and enzymes that maintain good health. Recommendations for iron are provided in *Dietary Reference Intakes*, developed by the Institute of Medicine of the National Academy of Sciences. The recommended dietary allowance (RDA) of iron for adult females under the age of 51 years is 18 milligrams (mg) per day. A hypothesis test is to be performed to decide whether adult females under the age of 51 years are, on average, getting less than the RDA of 18 mg of iron.

9.8 Early-Onset Dementia. Dementia is the loss of the intellectual and social abilities severe enough to interfere with judgment, behavior, and daily functioning. Alzheimer's disease is the most common type of dementia. In the article "Living with Early Onset Dementia: Exploring the Experience and Developing Evidence-Based Guidelines for Practice" (*Alzheimer's Care Quarterly*, Vol. 5, Issue 2, pp. 111–122), P. Harris and J. Keady explored the experience and struggles of people diagnosed with dementia and their families. A hypothesis test is to be performed to decide whether the mean age at diagnosis of all people with early-onset dementia is less than 55 years old.

9.9 Serving Time. According to the Bureau of Crime Statistics and Research of Australia, as reported on *Lawlink*, the mean length of imprisonment for motor-vehicle-theft offenders in Australia is 16.7 months. You want to perform a hypothesis test to decide whether the mean length of imprisonment for motor-vehicle-theft offenders in Sydney differs from the national mean in Australia.

9.10 Worker Fatigue. A study by M. Chen et al. titled "Heat Stress Evaluation and Worker Fatigue in a Steel Plant" (*American Industrial Hygiene Association*, Vol. 64, pp. 352–359) assessed fatigue in steel-plant workers due to heat stress. Among other things, the researchers monitored the heart rates of a random sample of 29 casting workers. A hypothesis test is to be conducted to decide whether the mean post-work heart rate of casting workers exceeds the normal resting heart rate of 72 beats per minute (bpm).

9.11 Body Temperature. A study by researchers at the University of Maryland addressed the question of whether the mean body temperature of humans is 98.6°F. The results of the study by P. Mackowiak et al. appeared in the article "A Critical Appraisal of 98.6°F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich" (*Journal of the American Medical Association*, Vol. 268, pp. 1578–1580). Among other data, the researchers obtained the body temperatures of 93 healthy humans. Suppose that you want to use those data to decide whether the mean body temperature of healthy humans differs from 98.6°F.

9.12 Teacher Salaries. The Educational Resource Service publishes information about wages and salaries in the public schools system in *National Survey of Salaries and Wages in Public Schools*. The mean annual salary of (public) classroom teachers is \$49.0 thousand. A hypothesis test is to be performed to decide whether the mean annual salary of classroom teachers in Hawaii is greater than the national mean.

9.13 Cell Phones. The number of cell phone users has increased dramatically since 1987. According to the *Semi-annual Wireless Survey*, published by the Cellular Telecommunications & Internet Association, the mean local monthly bill for cell phone users in the United States was \$49.94 in 2007. A hypothesis test is to be performed to determine whether last year's mean local monthly bill for cell phone users has decreased from the 2007 mean of \$49.94.

9.14 Suppose that, in a hypothesis test, the null hypothesis is in fact true.
a. Is it possible to make a Type I error? Explain your answer.
b. Is it possible to make a Type II error? Explain your answer.

9.15 Suppose that, in a hypothesis test, the null hypothesis is in fact false.
a. Is it possible to make a Type I error? Explain your answer.
b. Is it possible to make a Type II error? Explain your answer.

9.16 What is the relation between the significance level of a hypothesis test and the probability of making a Type I error?

9.17 Answer true or false and explain your answer: If it is important not to reject a true null hypothesis, the hypothesis test should be performed at a small significance level.

9.18 Answer true or false and explain your answer: For a fixed sample size, decreasing the significance level of a hypothesis test results in an increase in the probability of making a Type II error.

9.19 Identify the two types of incorrect decisions in a hypothesis test. For each incorrect decision, what symbol is used to represent the probability of making that type of error?

9.20 Suppose that a hypothesis test is performed at a small significance level. State the appropriate conclusion in each case by referring to Key Fact 9.2.
a. The null hypothesis is rejected.
b. The null hypothesis is not rejected.

9.21 Toxic Mushrooms? Refer to Exercise 9.5. Explain what each of the following would mean.
a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to nonrejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean cadmium level in *Boletus pinicola* mushrooms
d. equals the safety limit of 0.5 ppm.
e. exceeds the safety limit of 0.5 ppm.

9.22 Agriculture Books. Refer to Exercise 9.6. Explain what each of the following would mean.
a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact this year's mean retail price of agriculture books
d. equals the 2005 mean of \$57.61.
e. differs from the 2005 mean of \$57.61.

9.23 Iron Deficiency? Refer to Exercise 9.7. Explain what each of the following would mean.
a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean iron intake of all adult females under the age of 51 years

d. equals the RDA of 18 mg per day.

e. is less than the RDA of 18 mg per day.

9.24 Early-Onset Dementia. Refer to Exercise 9.8. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to nonrejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean age at diagnosis of all people with early-onset dementia

d. is 55 years old.

e. is less than 55 years old.

9.25 Serving Time. Refer to Exercise 9.9. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to nonrejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean length of imprisonment for motor-vehicle-theft offenders in Sydney

d. equals the national mean of 16.7 months.

e. differs from the national mean of 16.7 months.

9.26 Worker Fatigue. Refer to Exercise 9.10. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean post-work heart rate of casting workers

d. equals the normal resting heart rate of 72 bpm.

e. exceeds the normal resting heart rate of 72 bpm.

9.27 Body Temperature. Refer to Exercise 9.11. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean body temperature of all healthy humans

d. is 98.6°F.

e. is not 98.6°F.

9.28 Teacher Salaries. Refer to Exercise 9.12. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to nonrejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact the mean salary of classroom teachers in Hawaii

d. equals the national mean of $49.0 thousand.

e. exceeds the national mean of $49.0 thousand.

9.29 Cell Phones. Refer to Exercise 9.13. Explain what each of the following would mean.

a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to nonrejection of the null hypothesis. Classify that conclusion by error type or as a correct decision if in fact last year's mean local monthly bill for cell phone users

d. equals the 2007 mean of $49.94.

e. is less than the 2007 mean of $49.94.

9.30 Approving Nuclear Reactors. Suppose that you are performing a statistical test to decide whether a nuclear reactor should be approved for use. Further suppose that failing to reject the null hypothesis corresponds to approval. What property would you want the Type II error probability, β, to have?

9.31 Guilty or Innocent? In the U.S. court system, a defendant is assumed innocent until proven guilty. Suppose that you regard a court trial as a hypothesis test with null and alternative hypotheses

$$H_0\text{: Defendant is innocent}$$
$$H_a\text{: Defendant is guilty.}$$

a. Explain the meaning of a Type I error.

b. Explain the meaning of a Type II error.

c. If you were the defendant, would you want α to be large or small? Explain your answer.

d. If you were the prosecuting attorney, would you want β to be large or small? Explain your answer.

e. What are the consequences to the court system if you make $\alpha = 0$? $\beta = 0$?

9.2 Critical-Value Approach to Hypothesis Testing[†]

With the critical-value approach to hypothesis testing, we choose a "cutoff point" (or cutoff points) based on the significance level of the hypothesis test. The criterion for deciding whether to reject the null hypothesis involves a comparison of the value of the test statistic to the cutoff point(s). Our next example introduces these ideas.

EXAMPLE 9.5 The Critical-Value Approach

Golf Driving Distances Jack tells Jean that his average drive of a golf ball is 275 yards. Jean is skeptical and asks for substantiation. To that end, Jack hits 25 drives. The results, in yards, are shown in Table 9.2.

[†] Those concentrating on the P-value approach to hypothesis testing can skip this section if so desired.

TABLE 9.2
Distances (yards) of 25 drives by Jack

266	254	248	249	297
261	293	261	266	279
222	212	282	281	265
240	284	253	274	243
272	279	261	273	295

The (sample) mean of Jack's 25 drives is only 264.4 yards. Jack still maintains that, on average, he drives a golf ball 275 yards and that his (relatively) poor performance can reasonably be attributed to chance.

At the 5% significance level, do the data provide sufficient evidence to conclude that Jack's mean driving distance is less than 275 yards? We use the following steps to answer the question.

a. State the null and alternative hypotheses.
b. Discuss the logic of this hypothesis test.
c. Obtain a precise criterion for deciding whether to reject the null hypothesis in favor of the alternative hypothesis.
d. Apply the criterion in part (c) to the sample data and state the conclusion.

For our analysis, we assume that Jack's driving distances are normally distributed (which can be shown to be reasonable) and that the population standard deviation of all such driving distances is 20 yards.†

Solution

a. Let μ denote the population mean of (all) Jack's driving distances. The null hypothesis is Jack's claim of an overall driving-distance average of 275 yards. The alternative hypothesis is Jean's suspicion that Jack's overall driving-distance average is less than 275 yards. Hence, the null and alternative hypotheses are, respectively,

$$H_0\text{: } \mu = 275 \text{ yards (Jack's claim)}$$
$$H_a\text{: } \mu < 275 \text{ yards (Jean's suspicion).}$$

Note that this hypothesis test is left tailed.

b. Basically, the logic of this hypothesis test is as follows: If the null hypothesis is true, then the mean distance, $\bar{x}$, of the sample of Jack's 25 drives should approximately equal 275 yards. We say "approximately equal" because we cannot expect a sample mean to equal exactly the population mean; some sampling error is anticipated. However, if the sample mean driving distance is "too much smaller" than 275 yards, we would be inclined to reject the null hypothesis in favor of the alternative hypothesis.

c. We use our knowledge of the sampling distribution of the sample mean and the specified significance level to decide how much smaller is "too much smaller." Assuming that the null hypothesis is true, Key Fact 7.4 on page 313 shows that, for samples of size 25, the sample mean driving distance, $\bar{x}$, is normally distributed with mean and standard deviation

$$\mu_{\bar{x}} = \mu = 275 \text{ yards} \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4 \text{ yards},$$

respectively. Thus, from Key Fact 6.2 on page 254, the standardized version of $\bar{x}$,

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 275}{4},$$

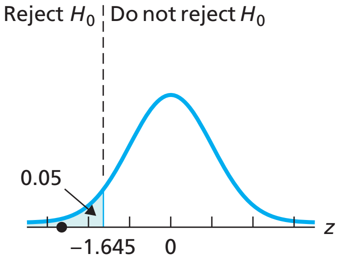
has the standard normal distribution. We use this variable, $z = (\bar{x} - 275)/4$, as our test statistic.

Because the hypothesis test is left tailed and we want a 5% significance level (i.e., $\alpha = 0.05$), we choose the cutoff point to be the z-score with area 0.05 to its left under the standard normal curve. From Table II, we find that z-score to be -1.645.

Consequently, "too much smaller" is a sample mean driving distance with a z-score of -1.645 or less. Figure 9.1 displays our criterion for deciding whether to reject the null hypothesis.

† We are assuming that the population standard deviation is known, for simplicity. The more usual case in which the population standard deviation is unknown is discussed in Section 9.5.

FIGURE 9.1
Criterion for deciding whether to reject the null hypothesis



d. Now we compute the value of the test statistic and compare it to our cutoff point of −1.645. As we noted, the sample mean driving distance of Jack's 25 drives is 264.4 yards. Hence, the value of the test statistic is

$$z = \frac{\bar{x} - 275}{4} = \frac{264.4 - 275}{4} = -2.65.$$

This value of z is marked with a dot in Fig. 9.1. We see that the value of the test statistic, −2.65, is less than the cutoff point of −1.645 and, hence, we reject H_0.

Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that Jack's mean driving distance is less than his claimed 275 yards.

Note: The curve in Fig. 9.1—which is the standard normal curve—is the normal curve for the test statistic $z = (\bar{x} - 275)/4$, provided that the null hypothesis is true. We see then from Fig. 9.1 that the probability of rejecting the null hypothesis if it is in fact true (i.e., the probability of making a Type I error) is 0.05. In other words, the significance level of the hypothesis test is indeed 0.05 (5%), as required.

Terminology of the Critical-Value Approach

Referring to the preceding example, we present some important terminology that is used with the critical-value approach to hypothesis testing. The set of values for the test statistic that leads us to reject the null hypothesis is called the **rejection region.** In this case, the rejection region consists of all z-scores that lie to the left of −1.645—that part of the horizontal axis under the shaded area in Fig. 9.1.

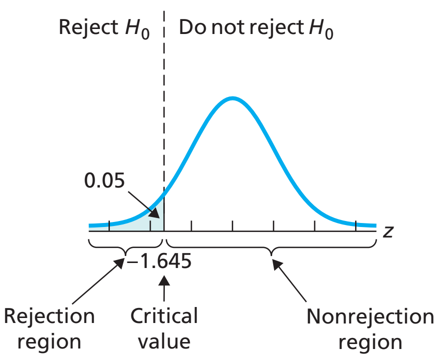
The set of values for the test statistic that leads us not to reject the null hypothesis is called the **nonrejection region.** Here, the nonrejection region consists of all z-scores that lie to the right of −1.645—that part of the horizontal axis under the unshaded area in Fig. 9.1.

The value of the test statistic that separates the rejection and nonrejection region (i.e., the cutoff point) is called the **critical value.** In this case, the critical value is $z = -1.645$.

We summarize the preceding discussion in Fig. 9.2, and, with that discussion in mind, we present Definition 9.4. Before doing so, however, we note the following:

- The rejection region pictured in Fig. 9.2 is typical of that for a left-tailed test. Soon we will discuss the form of the rejection regions for a two-tailed test and a right-tailed test.
- The terminology introduced so far in this section (and most of that which will be presented later) applies to any hypothesis test, not just to hypothesis tests for a population mean.

FIGURE 9.2
Rejection region, nonrejection region, and critical value for the golf-driving-distances hypothesis test



DEFINITION 9.4

Rejection Region, Nonrejection Region, and Critical Values

Rejection region: The set of values for the test statistic that leads to rejection of the null hypothesis.

Nonrejection region: The set of values for the test statistic that leads to nonrejection of the null hypothesis.

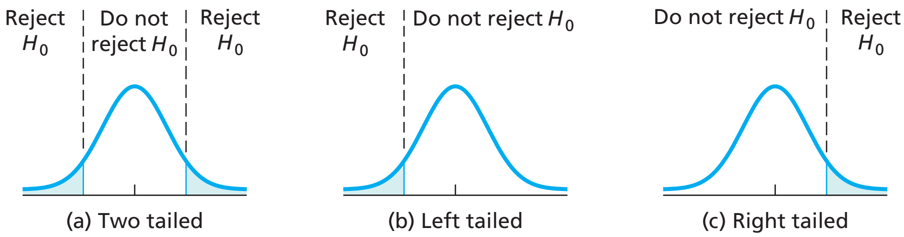
Critical value(s): The value or values of the test statistic that separate the rejection and nonrejection regions. A critical value is considered part of the rejection region.

What Does It Mean?

If the value of the test statistic falls in the rejection region, reject the null hypothesis; otherwise, do not reject the null hypothesis.

For a two-tailed test, as in Example 9.1 on page 360 (the pretzel-packaging illustration), the null hypothesis is rejected when the test statistic is either too small or too large. Thus the rejection region for such a test consists of two parts: one on the left and one on the right, as shown in Fig. 9.3(a).

FIGURE 9.3
Graphical display of rejection regions for two-tailed, left-tailed, and right-tailed tests



For a left-tailed test, as in Example 9.3 on page 361 (the calcium-intake illustration), the null hypothesis is rejected only when the test statistic is too small. Thus the rejection region for such a test consists of only one part, which is on the left, as shown in Fig. 9.3(b).

For a right-tailed test, as in Example 9.2 on page 361 (the history-book illustration), the null hypothesis is rejected only when the test statistic is too large. Thus the rejection region for such a test consists of only one part, which is on the right, as shown in Fig. 9.3(c).

Table 9.3 and Fig. 9.3 summarize our discussion. Figure 9.3 shows why the term *tailed* is used: The rejection region is in both tails for a two-tailed test, in the left tail for a left-tailed test, and in the right tail for a right-tailed test.

You try it!

Exercise 9.33 on page 372

TABLE 9.3
Rejection regions for two-tailed, left-tailed, and right-tailed tests

	Two-tailed test	Left-tailed test	Right-tailed test
Sign in H_a	$\neq$	$<$	$>$
Rejection region	Both sides	Left side	Right side

Obtaining Critical Values

Recall that the significance level of a hypothesis test is the probability of rejecting a true null hypothesis. With the critical-value approach, we reject the null hypothesis if and only if the test statistic falls in the rejection region. Therefore, we have Key Fact 9.3.

KEY FACT 9.3

Obtaining Critical Values

Suppose that a hypothesis test is to be performed at the significance level α. Then the critical value(s) must be chosen so that, if the null hypothesis is true, the probability is α that the test statistic will fall in the rejection region.

Obtaining Critical Values for a One-Mean z-Test

The first hypothesis-testing procedure that we discuss is called the **one-mean z-test.** This procedure is used to perform a hypothesis test for one population mean when the population standard deviation is known and the variable under consideration is normally distributed. Keep in mind, however, that because of the central limit theorem, the one-mean z-test will work reasonably well when the sample size is large, regardless of the distribution of the variable.

As you have seen, the null hypothesis for a hypothesis test concerning one population mean, μ, has the form H_0: $\mu = \mu_0$, where μ_0 is some number. Referring to part (c) of the solution to Example 9.5, we see that the test statistic for a one-mean z-test is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

which, by the way, tells you how many standard deviations the observed sample mean, $\bar{x}$, is from μ_0 (the value specified for the population mean in the null hypothesis).

The basis of the hypothesis-testing procedure is in Key Fact 7.4: If x is a normally distributed variable with mean μ and standard deviation σ, then, for samples of size n, the variable $\bar{x}$ is also normally distributed and has mean μ and standard deviation $\sigma/\sqrt{n}$. This fact and Key Fact 6.2 (page 254) applied to $\bar{x}$ imply that, if the null hypothesis is true, the test statistic z has the standard normal distribution.

Consequently, in view of Key Fact 9.3, for a specified significance level α, we need to choose the critical value(s) so that the area under the standard normal curve that lies above the rejection region equals α.

EXAMPLE 9.6 Obtaining the Critical Values for a One-Mean z-Test

Determine the critical value(s) for a one-mean z-test at the 5% significance level ($\alpha = 0.05$) if the test is

a. two tailed. **b.** left tailed. **c.** right tailed.

Solution Because $\alpha = 0.05$, we need to choose the critical value(s) so that the area under the standard normal curve that lies above the rejection region equals 0.05.

a. For a two-tailed test, the rejection region is on both the left and right. So the critical values are the two z-scores that divide the area under the standard normal curve into a middle 0.95 area and two outside areas of 0.025. In other words, the critical values are $\pm z_{0.025}$. From Table II in Appendix A, $\pm z_{0.025} = \pm 1.96$, as shown in Fig. 9.4(a).

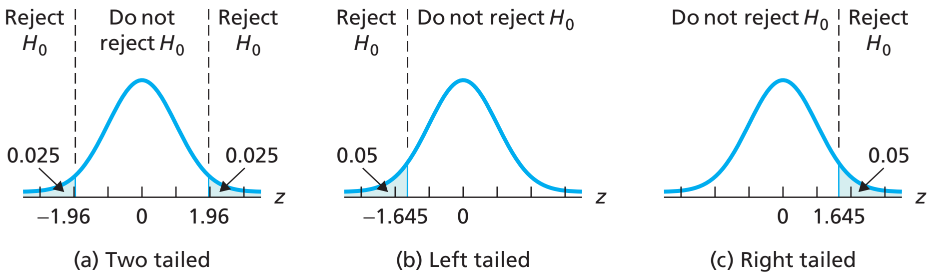


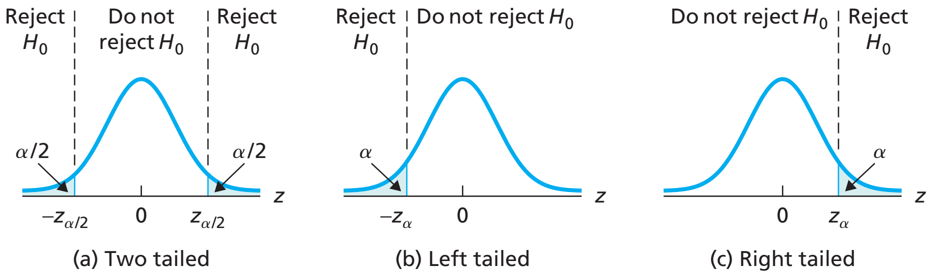
FIGURE 9.4 Critical value(s) for a hypothesis test at the 5% significance level if the test is (a) two tailed, (b) left tailed, or (c) right tailed

b. For a left-tailed test, the rejection region is on the left. So the critical value is the z-score with area 0.05 to its left under the standard normal curve, which is $-z_{0.05}$. From Table II, $-z_{0.05} = -1.645$, as shown in Fig. 9.4(b).

c. For a right-tailed test, the rejection region is on the right. So the critical value is the z-score with area 0.05 to its right under the standard normal curve, which is $z_{0.05}$. From Table II, $z_{0.05} = 1.645$, as shown in Fig. 9.4(c).

By reasoning as we did in the previous example, we can obtain the critical value(s) for any specified significance level α. As shown in Fig. 9.5, for a two-tailed test, the critical values are $\pm z_{\alpha/2}$; for a left-tailed test, the critical value is $-z_\alpha$; and for a right-tailed test, the critical value is z_α.

FIGURE 9.5
Critical value(s) for a hypothesis test at the significance level α if the test is (a) two tailed, (b) left tailed, or (c) right tailed



You try it!

Exercise 9.39 on page 372

The most commonly used significance levels are 0.10, 0.05, and 0.01. If we consider both one-tailed and two-tailed tests, these three significance levels give rise to five "tail areas." Using the standard-normal table, Table II, we obtained the value of z_α corresponding to each of those five tail areas as shown in Table 9.4.

Alternatively, we can find these five values of z_α at the bottom of the t-table, Table IV, where they are displayed to three decimal places. Can you explain the slight discrepancy between the values given for $z_{0.005}$ in the two tables?

TABLE 9.4
Some important values of z_α

$z_{0.10}$	$z_{0.05}$	$z_{0.025}$	$z_{0.01}$	$z_{0.005}$
1.28	1.645	1.96	2.33	2.575

Steps in the Critical-Value Approach to Hypothesis Testing

We have now covered all the concepts required for the critical-value approach to hypothesis testing. The general steps involved in that approach are presented in Table 9.5.

TABLE 9.5
General steps for the critical-value approach to hypothesis testing

CRITICAL-VALUE APPROACH TO HYPOTHESIS TESTING

Step 1	State the null and alternative hypotheses.
Step 2	Decide on the significance level, α.
Step 3	Compute the value of the test statistic.
Step 4	Determine the critical value(s).
Step 5	If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.
Step 6	Interpret the result of the hypothesis test.

Throughout the text, we present dedicated step-by-step procedures for specific hypothesis-testing procedures. Those using the critical-value approach, however, are all based on the steps shown in Table 9.5.

Exercises 9.2

Understanding the Concepts and Skills

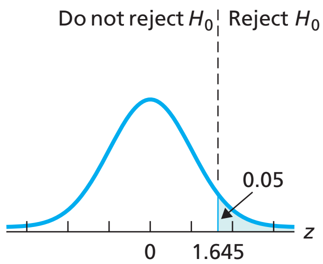
9.32 Explain in your own words the meaning of each of the following terms.

a. test statistic
b. rejection region
c. nonrejection region
d. critical values
e. significance level

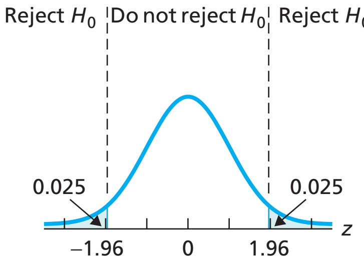
*Exercises **9.33–9.38** contain graphs portraying the decision criterion for a one-mean z-test. The curve in each graph is the normal curve for the test statistic under the assumption that the null hypothesis is true. For each exercise, determine the*

a. rejection region.
b. nonrejection region.
c. critical value(s).
d. significance level.
e. Construct a graph similar to that in Fig 9.2 on page 368 that depicts your results from parts (a)–(d).
f. Identify the hypothesis test as two tailed, left tailed, or right tailed.

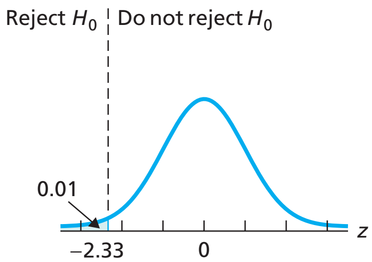
9.33



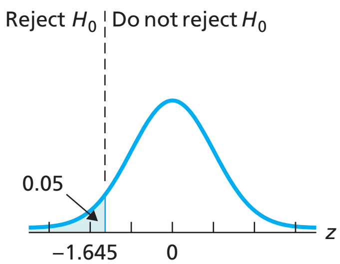
9.34



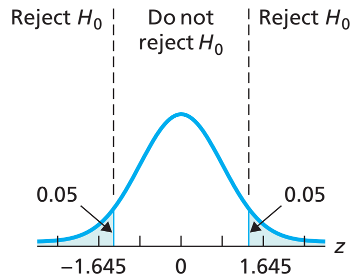
9.35



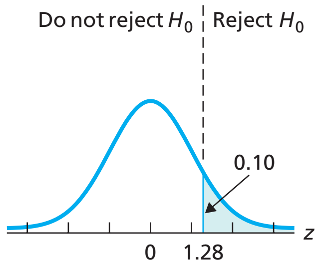
9.36



9.37



9.38



*In each of Exercises **9.39–9.44**, determine the critical value(s) for a one-mean z-test. For each exercise, draw a graph that illustrates your answer.*

9.39 A two-tailed test with $\alpha = 0.10$.

9.40 A right-tailed test with $\alpha = 0.05$.

9.41 A left-tailed test with $\alpha = 0.01$.

9.42 A left-tailed test with $\alpha = 0.05$.

9.43 A right-tailed test with $\alpha = 0.01$.

9.44 A two-tailed test with $\alpha = 0.05$.

9.3 *P*-Value Approach to Hypothesis Testing†

Roughly speaking, with the P-value approach to hypothesis testing, we first evaluate how likely observation of the value obtained for the test statistic would be if the null hypothesis is true. The criterion for deciding whether to reject the null hypothesis involves a comparison of that likelihood with the specified significance level of the hypothesis test. Our next example introduces these ideas.

EXAMPLE 9.7 The *P*-Value Approach

Golf Driving Distances Jack tells Jean that his average drive of a golf ball is 275 yards. Jean is skeptical and asks for substantiation. To that end, Jack hits 25 drives. The results, in yards, are shown in Table 9.6.

† Those concentrating on the critical-value approach to hypothesis testing can skip this section if so desired. Note, however, that this section is prerequisite to the (optional) technology materials that appear in The Technology Center sections.

TABLE 9.6
Distances (yards) of 25 drives by Jack

266	254	248	249	297
261	293	261	266	279
222	212	282	281	265
240	284	253	274	243
272	279	261	273	295

The (sample) mean of Jack's 25 drives is only 264.4 yards. Jack still maintains that, on average, he drives a golf ball 275 yards and that his (relatively) poor performance can reasonably be attributed to chance.

At the 5% significance level, do the data provide sufficient evidence to conclude that Jack's mean driving distance is less than 275 yards? We use the following steps to answer the question.

a. State the null and alternative hypotheses.
b. Discuss the logic of this hypothesis test.
c. Obtain a precise criterion for deciding whether to reject the null hypothesis in favor of the alternative hypothesis.
d. Apply the criterion in part (c) to the sample data and state the conclusion.

For our analysis, we assume that Jack's driving distances are normally distributed (which can be shown to be reasonable) and that the population standard deviation of all such driving distances is 20 yards.†

Solution

a. Let μ denote the population mean of (all) Jack's driving distances. The null hypothesis is Jack's claim of an overall driving-distance average of 275 yards. The alternative hypothesis is Jean's suspicion that Jack's overall driving-distance average is less than 275 yards. Hence, the null and alternative hypotheses are, respectively,

$$H_0\text{: } \mu = 275 \text{ yards (Jack's claim)}$$
$$H_a\text{: } \mu < 275 \text{ yards (Jean's suspicion).}$$

Note that this hypothesis test is left tailed.

b. Basically, the logic of this hypothesis test is as follows: If the null hypothesis is true, then the mean distance, $\bar{x}$, of the sample of Jack's 25 drives should approximately equal 275 yards. We say "approximately equal" because we cannot expect a sample mean to equal exactly the population mean; some sampling error is anticipated. However, if the sample mean driving distance is "too much smaller" than 275 yards, we would be inclined to reject the null hypothesis in favor of the alternative hypothesis.

c. We use our knowledge of the sampling distribution of the sample mean and the specified significance level to decide how much smaller is "too much smaller." Assuming that the null hypothesis is true, Key Fact 7.4 on page 313 shows that, for samples of size 25, the sample mean driving distance, $\bar{x}$, is normally distributed with mean and standard deviation

$$\mu_{\bar{x}} = \mu = 275 \text{ yards} \qquad \text{and} \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4 \text{ yards,}$$

respectively. Thus, from Key Fact 6.2 on page 254, the standardized version of $\bar{x}$,

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 275}{4},$$

has the standard normal distribution. We use this variable, $z = (\bar{x} - 275)/4$, as our test statistic.

Because the hypothesis test is left tailed, we compute the probability of observing a value of the test statistic z that is as small as or smaller than the value actually observed. This probability is called the P*-value* of the hypothesis test and is denoted by the letter P.

† We are assuming that the population standard deviation is known, for simplicity. The more usual case in which the population standard deviation is unknown is discussed in Section 9.5.

Our criterion for deciding whether to reject the null hypothesis is then as follows: If the P-value is less than or equal to the specified significance level, we reject the null hypothesis; otherwise, we do not reject the null hypothesis.

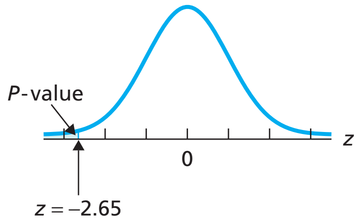
d. Now we obtain the P-value and compare it to the specified significance level of 0.05. As we have noted, the sample mean driving distance of Jack's 25 drives is 264.4 yards. Hence, the value of the test statistic is

$$z = \frac{\bar{x} - 275}{4} = \frac{264.4 - 275}{4} = -2.65.$$

Consequently, the P-value is the probability of observing a value of z of -2.65 or smaller if the null hypothesis is true. That probability equals the area under the standard normal curve to the left of -2.65, the shaded region in Fig. 9.6. From Table II, we find that area to be 0.0040. Because the P-value, 0.0040, is less than the specified significance level of 0.05, we reject H_0.

FIGURE 9.6

P-value for golf-driving-distances hypothesis test



Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that Jack's mean driving distance is less than his claimed 275 yards.

Note: The P-value will be less than or equal to 0.05 whenever the value of the test statistic z has area 0.05 or less to its left under the standard normal curve, which is exactly 5% of the time if the null hypothesis is true. Thus, we see that, by using the decision criterion "reject the null hypothesis if $P \leq 0.05$; otherwise, do not reject the null hypothesis," the probability of rejecting the null hypothesis if it is in fact true (i.e., the probability of making a Type I error) is 0.05. In other words, the significance level of the hypothesis test is indeed 0.05 (5%), as required.

Let us emphasize the meaning of the P-value, 0.0040, obtained in the preceding example. Specifically, if the null hypothesis is true, we would observe a value of the test statistic z of -2.65 or less only 4 times in 1000. In other words, if the null hypothesis is true, a random sample of 25 of Jack's drives would have a mean distance of 264.4 yards or less only 0.4% of the time. The sample data provide very strong evidence against the null hypothesis (Jack's claim) and in favor of the alternative hypothesis (Jean's suspicion).

Terminology of the *P*-Value Approach

We introduced the P-value in the context of the preceding example. More generally, we define the P-value as follows.

DEFINITION 9.5

What Does It Mean?

Small P-values provide evidence against the null hypothesis; larger P-values do not.

P-Value

The ***P*-value** of a hypothesis test is the probability of getting sample data at least as inconsistent with the null hypothesis (and supportive of the alternative hypothesis) as the sample data actually obtained.† We use the letter ***P*** to denote the P-value.

Note: The smaller (closer to 0) the P-value, the stronger is the evidence against the null hypothesis and, hence, in favor of the alternative hypothesis. Stated simply, *an outcome that would rarely occur if the null hypothesis were true provides evidence against the null hypothesis and, hence, in favor of the alternative hypothesis.*

† Alternatively, we can define the P-value to be the percentage of samples that are at least as inconsistent with the null hypothesis (and supportive of the alternative hypothesis) as the sample actually obtained.

As illustrated in the solution to part (c) of Example 9.7 (golf driving distances), with the P-value approach to hypothesis testing, we use the following criterion to decide whether to reject the null hypothesis.

KEY FACT 9.4

Decision Criterion for a Hypothesis Test Using the *P*-Value

If the P-value is less than or equal to the specified significance level, reject the null hypothesis; otherwise, do not reject the null hypothesis. In other words, if $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

The P-value of a hypothesis test is also referred to as the **observed significance level.** To understand why, suppose that the P-value of a hypothesis test is $P = 0.07$. Then, for instance, we see from Key Fact 9.4 that we can reject the null hypothesis at the 10% significance level (because $P \leq 0.10$), but we cannot reject the null hypothesis at the 5% significance level (because $P > 0.05$). In fact, here, the null hypothesis can be rejected at any significance level of at least 0.07 and cannot be rejected at any significance level less than 0.07.

More generally, we have the following fact.

KEY FACT 9.5

P-Value as the Observed Significance Level

The P-value of a hypothesis test equals the smallest significance level at which the null hypothesis can be rejected, that is, the smallest significance level for which the observed semple data results in rejection of H_0.

Determining *P*-Values

We defined the P-value of a hypothesis test in Definition 9.5. To actually determine a P-value, however, we rely on the value of the test statistic, as follows.

KEY FACT 9.6

Determining a *P*-Value

To determine the P-value of a hypothesis test, we assume that the null hypothesis is true and compute the probability of observing a value of the test statistic as extreme as or more extreme than that observed. By *extreme* we mean "far from what we would expect to observe if the null hypothesis is true."

Determining the *P*-Value for a One-Mean *z*-Test

The first hypothesis-testing procedure that we discuss is called the **one-mean *z*-test.** This procedure is used to perform a hypothesis test for one population mean when the population standard deviation is known and the variable under consideration is normally distributed. Keep in mind, however, that because of the central limit theorem, the one-mean z-test will work reasonably well when the sample size is large, regardless of the distribution of the variable.

As you have seen, the null hypothesis for a hypothesis test concerning one population mean, μ, has the form H_0: $\mu = \mu_0$, where μ_0 is some number. Referring to part (c) of the solution to Example 9.7, we see that the test statistic for a one-mean z-test is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

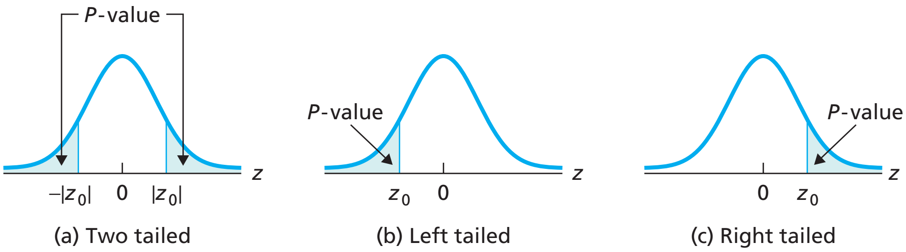
which, by the way, tells you how many standard deviations the observed sample mean, $\bar{x}$, is from μ_0 (the value specified for the population mean in the null hypothesis).

The basis of the hypothesis-testing procedure is in Key Fact 7.4: If $\bar{x}$ is a normally distributed variable with mean μ and standard deviation σ, then, for samples of size n, the variable $\bar{x}$ is also normally distributed and has mean μ and standard deviation $\sigma/\sqrt{n}$. This fact and Key Fact 6.2 (page 254) applied to $\bar{x}$ imply that, if the null hypothesis is true, the test statistic z has the standard normal distribution and hence that its probabilities equal areas under the standard normal curve.

Therefore, in view of Key Fact 9.6, if we let z_0 denote the observed value of the test statistic z, we determine the P-value as follows:

- *Two-tailed test:* The P-value equals the probability of observing a value of the test statistic z that is at least as large in magnitude as the value actually observed, which is the area under the standard normal curve that lies outside the interval from $-|z_0|$ to $|z_0|$, as illustrated in Fig. 9.7(a).
- *Left-tailed test:* The P-value equals the probability of observing a value of the test statistic z that is as small as or smaller than the value actually observed, which is the area under the standard normal curve that lies to the left of z_0, as illustrated in Fig. 9.7(b).
- *Right-tailed test:* The P-value equals the probability of observing a value of the test statistic z that is as large as or larger than the value actually observed, which is the area under the standard normal curve that lies to the right of z_0, as illustrated in Fig. 9.7(c).

FIGURE 9.7
P-value for a one-mean z-test if the test is (a) two tailed, (b) left tailed, or (c) right tailed



EXAMPLE 9.8 Determining the *P*-Value for a One-Mean *z*-Test

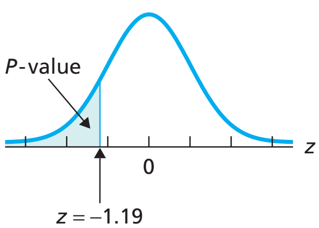
The value of the test statistic for a left-tailed one-mean z-test is $z = -1.19$.

a. Determine the P-value.

b. At the 5% significance level, do the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis?

Solution

FIGURE 9.8
Value of the test statistic and the P-value



a. Because the test is left tailed, the P-value is the probability of observing a value of z of -1.19 or less if the null hypothesis is true. That probability equals the area under the standard normal curve to the left of -1.19, the shaded area shown in Fig. 9.8, which, by Table II, is 0.1170. Therefore, $P = 0.1170$.

b. The specified significance level is 5%, that is, $\alpha = 0.05$. Hence, from part (a), we see that $P > \alpha$. Thus, by Key Fact 9.4, we do not reject the null hypothesis. At the 5% significance level, the data do not provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.

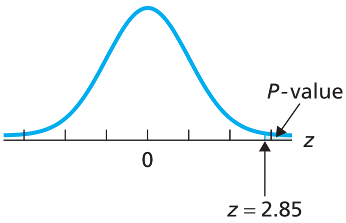
EXAMPLE 9.9 Determining the *P*-Value for a One-Mean *z*-Test

The value of the test statistic for a right-tailed one-mean z-test is $z = 2.85$.

a. Determine the P-value.
b. At the 1% significance level, do the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis?

Solution

FIGURE 9.9
Value of the test statistic and the P-value

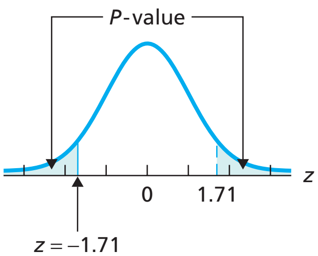


a. Because the test is right tailed, the P-value is the probability of observing a value of z of 2.85 or greater if the null hypothesis is true. That probability equals the area under the standard normal curve to the right of 2.85, the shaded area shown in Fig. 9.9, which, by Table II, is $1 - 0.9978 = 0.0022$. Therefore, $P = 0.0022$.

b. The specified significance level is 1%, that is, $\alpha = 0.01$. Hence, from part (a), we see that $P \leq \alpha$. Thus, by Key Fact 9.4, we reject the null hypothesis. At the 1% significance level, the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.

EXAMPLE 9.10 Determining the *P*-Value for a One-Mean *z*-Test

FIGURE 9.10
Value of the test statistic and the P-value



Exercise 9.55 on page 378

The value of the test statistic for a two-tailed one-mean z-test is $z = -1.71$.

a. Determine the P-value.
b. At the 5% significance level, do the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis?

Solution

a. Because the test is two tailed, the P-value is the probability of observing a value of z of 1.71 or greater in magnitude if the null hypothesis is true. That probability equals the area under the standard normal curve that lies either to the left of -1.71 or to the right of 1.71, the shaded area shown in Fig. 9.10, which, by Table II, is $2 \cdot 0.0436 = 0.0872$. Therefore, $P = 0.0872$.

b. The specified significance level is 5%, that is, $\alpha = 0.05$. Hence, from part (a), we see that $P > \alpha$. Thus, by Key Fact 9.4, we do not reject the null hypothesis. At the 5% significance level, the data do not provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.

Steps in the *P*-Value Approach to Hypothesis Testing

We have now covered all the concepts required for the P-value approach to hypothesis testing. The general steps involved in that approach are presented in Table 9.7.

TABLE 9.7
General steps for the *P*-value approach to hypothesis testing

***P*-VALUE APPROACH TO HYPOTHESIS TESTING**

Step 1	State the null and alternative hypotheses.
Step 2	Decide on the significance level, α.
Step 3	Compute the value of the test statistic.
Step 4	Determine the P-value, P.
Step 5	If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.
Step 6	Interpret the result of the hypothesis test.

Throughout the text, we present dedicated step-by-step procedures for specific hypothesis-testing procedures. Those using the P-value approach, however, are all based on the steps shown in Table 9.7.

Using the *P*-Value to Assess the Evidence Against the Null Hypothesis

Key Fact 9.5 asserts that the P-value is the smallest significance level at which the null hypothesis can be rejected. Consequently, knowing the P-value allows us to assess significance at any level we desire. For instance, if the P-value of a hypothesis test is 0.03, the null hypothesis can be rejected at any significance level larger than or equal to 0.03, and it cannot be rejected at any significance level smaller than 0.03.

Knowing the P-value also allows us to evaluate the strength of the evidence against the null hypothesis: the smaller the P-value, the stronger will be the evidence against the null hypothesis. Table 9.8 presents guidelines for interpreting the P-value of a hypothesis test.

TABLE 9.8
Guidelines for using the P-value to assess the evidence against the null hypothesis

P-value	Evidence against H_0
$P > 0.10$	Weak or none
$0.05 < P \leq 0.10$	Moderate
$0.01 < P \leq 0.05$	Strong
$P \leq 0.01$	Very strong

Note that we can use the P-value to evaluate the strength of the evidence against the null hypothesis without reference to significance levels. This practice is common among researchers.

Hypothesis Tests Without Significance Levels: Many researchers do not explicitly refer to significance levels. Instead, they simply obtain the P-value and use it (or let the reader use it) to assess the strength of the evidence against the null hypothesis.

Exercises 9.3

Understanding the Concepts and Skills

9.45 State two reasons why including the P-value is prudent when you are reporting the results of a hypothesis test.

9.46 What is the P-value of a hypothesis test? When does it provide evidence against the null hypothesis?

9.47 Explain how the P-value is obtained for a one-mean z-test in case the hypothesis test is
a. left tailed. **b.** right tailed. **c.** two tailed.

9.48 True or false: The P-value is the smallest significance level for which the observed sample data result in rejection of the null hypothesis.

9.49 The P-value for a hypothesis test is 0.06. For each of the following significance levels, decide whether the null hypothesis should be rejected.
a. $\alpha = 0.05$ **b.** $\alpha = 0.10$ **c.** $\alpha = 0.06$

9.50 The P-value for a hypothesis test is 0.083. For each of the following significance levels, decide whether the null hypothesis should be rejected.
a. $\alpha = 0.05$ **b.** $\alpha = 0.10$ **c.** $\alpha = 0.06$

9.51 Which provides stronger evidence against the null hypothesis, a P-value of 0.02 or a P-value of 0.03? Explain your answer.

9.52 Which provides stronger evidence against the null hypothesis, a P-value of 0.06 or a P-value of 0.04? Explain your answer.

9.53 In each part, we have given the P-value for a hypothesis test. For each case, refer to Table 9.8 to determine the strength of the evidence against the null hypothesis.
a. $P = 0.06$ **b.** $P = 0.35$
c. $P = 0.027$ **d.** $P = 0.004$

9.54 In each part, we have given the P-value for a hypothesis test. For each case, refer to Table 9.8 to determine the strength of the evidence against the null hypothesis.
a. $P = 0.184$ **b.** $P = 0.086$
c. $P = 0.001$ **d.** $P = 0.012$

*In Exercises **9.55–9.60**, we have given the value obtained for the test statistic, z, in a one-mean z-test. We have also specified whether the test is two tailed, left tailed, or right tailed. Determine the P-value in each case and decide whether, at the 5% significance level, the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.*

9.55 Right-tailed test:
a. $z = 2.03$ **b.** $z = -0.31$

9.56 Left-tailed test:
a. $z = -1.84$ **b.** $z = 1.25$

9.57 Left-tailed test:
a. $z = -0.74$ **b.** $z = 1.16$

9.58 Two-tailed test:
a. $z = 3.08$ **b.** $z = -2.42$

9.59 Two-tailed test:
a. $z = -1.66$ **b.** $z = 0.52$

9.60 Right-tailed test:
a. $z = 1.24$ **b.** $z = -0.69$

Extending the Concepts and Skills

9.61 Consider a one-mean z-test. Denote z_0 as the observed value of the test statistic z. If the test is right tailed, then the P-value can be expressed as $P(z \geq z_0)$. Determine the corresponding expression for the P-value if the test is
a. left tailed. **b.** two tailed.

9.62 The symbol $\Phi(z)$ is often used to denote the area under the standard normal curve that lies to the left of a specified value of z. Consider a one-mean z-test. Denote z_0 as the observed value of the test statistic z. Express the P-value of the hypothesis test in terms of Φ if the test is
a. left tailed. **b.** right tailed. **c.** two tailed.

9.63 Obtaining the *P*-value. Let x denote the test statistic for a hypothesis test and x_0 its observed value. Then the P-value of the hypothesis test equals
a. $P(x \geq x_0)$ for a right-tailed test,
b. $P(x \leq x_0)$ for a left-tailed test,
c. $2 \cdot \min\{P(x \leq x_0), P(x \geq x_0)\}$ for a two-tailed test,

where the probabilities are computed under the assumption that the null hypothesis is true. Suppose that you are considering a one-mean z-test. Verify that the probability expressions in parts (a)–(c) are equivalent to those obtained in Exercise 9.61.

9.4 Hypothesis Tests for One Population Mean When σ Is Known

As we mentioned earlier, the first hypothesis-testing procedure that we discuss is used to perform a hypothesis test for one population mean when the population standard deviation is known. We call this hypothesis-testing procedure the **one-mean *z*-test** or, when no confusion can arise, simply the ***z*-test.**†

Procedure 9.1 on the next page provides a step-by-step method for performing a one-mean z-test. As you can see, Procedure 9.1 includes options for either the critical-value approach (keep left) or the P-value approach (keep right). The bases for these approaches were discussed in Sections 9.2 and 9.3, respectively.

Properties and guidelines for use of the one-mean z-test are similar to those for the one-mean z-interval procedure. In particular, the one-mean z-test is robust to moderate violations of the normality assumption but, even for large samples, can sometimes be unduly affected by outliers because the sample mean is not resistant to outliers. Key Fact 9.7 lists some general guidelines for use of the one-mean z-test.

KEY FACT 9.7

When to Use the One-Mean *z*-Test‡

- For small samples—say, of size less than 15—the z-test should be used only when the variable under consideration is normally distributed or very close to being so.
- For samples of moderate size—say, between 15 and 30—the z-test can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.
- For large samples—say, of size 30 or more—the z-test can be used essentially without restriction. However, if outliers are present and their removal is not justified, you should perform the hypothesis test once with the outliers and once without them to see what effect the outliers have. If the conclusion is affected, use a different procedure or take another sample, if possible.
- If outliers are present but their removal is justified and results in a data set for which the z-test is appropriate (as previously stated), the procedure can be used.

† The one-mean z-test is also known as the **one-sample *z*-test** and the **one-variable *z*-test.** We prefer "one-mean" because it makes clear the parameter being tested.

‡ We can refine these guidelines further by considering the impact of skewness. Roughly speaking, the more skewed the distribution of the variable under consideration, the larger is the sample size required to use the z-test.

PROCEDURE 9.1 One-Mean z-Test

Purpose To perform a hypothesis test for a population mean, μ

Assumptions

1. Simple random sample
2. Normal population or large sample
3. σ known

Step 1 **The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is**

H_a: $\mu \neq \mu_0$	or	H_a: $\mu < \mu_0$	or	H_a: $\mu > \mu_0$
(Two tailed)		**(Left tailed)**		**(Right tailed)**

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

and denote that value z_0.

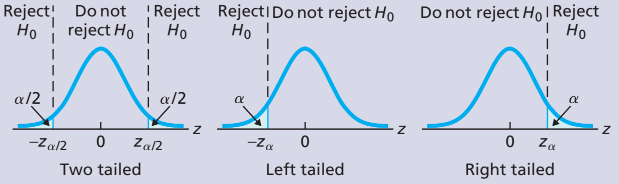
CRITICAL-VALUE APPROACH OR *P*-VALUE APPROACH

CRITICAL-VALUE APPROACH

Step 4 **The critical value(s) are**

$\pm z_{\alpha/2}$	or	$-z_\alpha$	or	z_α
(Two tailed)		**(Left tailed)**		**(Right tailed)**

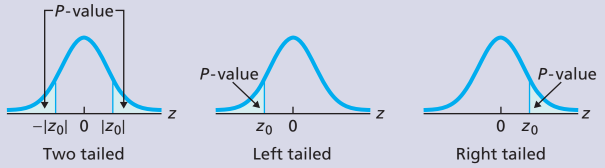
Use Table II to find the critical value(s).



Step 5 **If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.**

P-VALUE APPROACH

Step 4 **Use Table II to obtain the *P*-value.**



Step 5 **If $P \le \alpha$, reject H_0; otherwise, do not reject H_0.**

Step 6 **Interpret the results of the hypothesis test.**

Note: The hypothesis test is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Note: By saying that the hypothesis test is *exact*, we mean that the true significance level equals α; by saying that it is *approximately correct*, we mean that the true significance level only approximately equals α.

Applying the One-Mean z-Test

Examples 9.11–9.13 illustrate use of the z-test, Procedure 9.1.

EXAMPLE 9.11 The One-Mean z-Test

Prices of History Books The R. R. Bowker Company collects information on the retail prices of books and publishes its findings in *The Bowker Annual Library and Book Trade Almanac*. In 2005, the mean retail price of all history books was \$78.01. This year's retail prices for 40 randomly selected history books are shown in Table 9.9.

TABLE 9.9
This year's prices, in dollars, for 40 history books

82.55	72.80	73.89	80.54
80.26	74.43	81.37	82.28
77.55	88.25	73.58	89.23
74.35	77.44	78.91	77.50
77.83	77.49	87.25	98.93
74.25	82.71	78.88	78.25
80.35	77.45	90.29	79.42
67.63	91.48	83.99	80.64
101.92	83.03	95.59	69.26
80.31	98.72	87.81	69.20

At the 1% significance level, do the data provide sufficient evidence to conclude that this year's mean retail price of all history books has increased from the 2005 mean of \$78.01? Assume that the population standard deviation of prices for this year's history books is \$7.61.

Solution We constructed (but did not show) a normal probability plot, a histogram, a stem-and-leaf diagram, and a boxplot for these price data. The boxplot indicated potential outliers, but in view of the other three graphs, we concluded that the data contain no outliers. Because the sample size is 40, which is large, and the population standard deviation is known, we can use Procedure 9.1 to conduct the required hypothesis test.

Step 1 State the null and alternative hypotheses.

Let μ denote this year's mean retail price of all history books. We obtained the null and alternative hypotheses in Example 9.2 as

$$H_0\colon \mu = \$78.01 \text{ (mean price has not increased)}$$
$$H_a\colon \mu > \$78.01 \text{ (mean price has increased).}$$

Note that the hypothesis test is right tailed because a greater-than sign (>) appears in the alternative hypothesis.

Step 2 Decide on the significance level, α.

We are to perform the test at the 1% significance level, or $\alpha = 0.01$.

Step 3 Compute the value of the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

We have $\mu_0 = 78.01$, $\sigma = 7.61$, and $n = 40$. The mean of the sample data in Table 9.9 is $\bar{x} = 81.440$. Thus the value of the test statistic is

$$z = \frac{81.440 - 78.01}{7.61/\sqrt{40}} = 2.85.$$

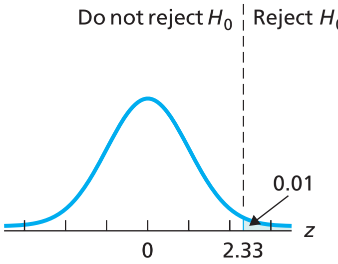
CRITICAL-VALUE APPROACH	OR	*P*-VALUE APPROACH

CRITICAL-VALUE APPROACH

Step 4 The critical value for a right-tailed test is z_α. Use Table II to find the critical value.

Because $\alpha = 0.01$, the critical value is $z_{0.01}$. From Table II (or Table 9.4 on page 371), $z_{0.01} = 2.33$, as shown in Fig. 9.11A.

FIGURE 9.11A



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

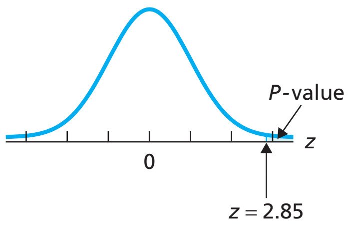
The value of the test statistic found in Step 3 is $z = 2.85$. Figure 9.11A reveals that this value falls in the rejection region, so we reject H_0. The test results are statistically significant at the 1% level.

P-VALUE APPROACH

Step 4 Use Table II to obtain the P-value.

From Step 3, the value of the test statistic is $z = 2.85$. The test is right tailed, so the P-value is the probability of observing a value of z of 2.85 or greater if the null hypothesis is true. That probability equals the shaded area in Fig. 9.11B, which, by Table II, is 0.0022. Hence $P = 0.0022$.

FIGURE 9.11B



Step 5 If $P \le \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P = 0.0022$. Because the P-value is less than the specified significance level of 0.01, we reject H_0. The test results are statistically significant at the 1% level and (see Table 9.8 on page 378) provide very strong evidence against the null hypothesis.

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 1% significance level, the data provide sufficient evidence to conclude that this year's mean retail price of all history books has increased from the 2005 mean of \$78.01.

EXAMPLE 9.12 The One-Mean *z*-Test

Poverty and Dietary Calcium Calcium is the most abundant mineral in the human body and has several important functions. Most body calcium is stored in the bones and teeth, where it functions to support their structure. Recommendations for calcium are provided in *Dietary Reference Intakes*, developed by the Institute of Medicine of the National Academy of Sciences. The recommended adequate intake (RAI) of calcium for adults (ages 19–50 years) is 1000 milligrams (mg) per day.

A simple random sample of 18 adults with incomes below the poverty level gives the daily calcium intakes shown in Table 9.10. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean calcium intake of all adults with incomes below the poverty level is less than the RAI of 1000 mg? Assume that $\sigma = 188$ mg.

TABLE 9.10
Daily calcium intake (mg) for 18 adults with incomes below the poverty level

886	633	943	847	934	841
1193	820	774	834	1050	1058
1192	975	1313	872	1079	809

Solution Because the sample size, $n = 18$, is moderate, we first need to consider questions of normality and outliers. (See the second bulleted item in Key Fact 9.7 on page 379.) Hence we constructed a normal probability plot for the data, shown in Fig. 9.12. The plot reveals no outliers and falls roughly in a straight line. Thus, we can apply Procedure 9.1 to perform the required hypothesis test.

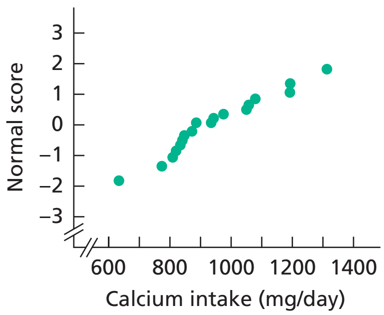


FIGURE 9.12
Normal probability plot of the calcium-intake data in Table 9.10

Step 1 State the null and alternative hypotheses.

Let μ denote the mean calcium intake (per day) of all adults with incomes below the poverty level. The null and alternative hypotheses, which we obtained in Example 9.3, are, respectively,

$$H_0\text{: } \mu = 1000 \text{ mg (mean calcium intake is not less than the RAI)}$$
$$H_a\text{: } \mu < 1000 \text{ mg (mean calcium intake is less than the RAI).}$$

Note that the hypothesis test is left tailed because a less-than sign (<) appears in the alternative hypothesis.

Step 2 Decide on the significance level, α.

We are to perform the test at the 5% significance level, or $\alpha = 0.05$.

Step 3 Compute the value of the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

We have $\mu_0 = 1000$, $\sigma = 188$, and $n = 18$. From the data in Table 9.10, we find that $\bar{x} = 947.4$. Thus the value of the test statistic is

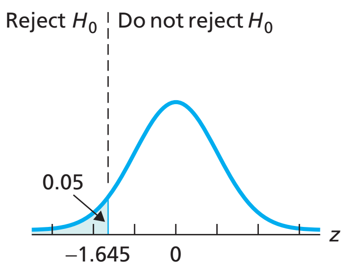
$$z = \frac{947.4 - 1000}{188/\sqrt{18}} = -1.19.$$

CRITICAL-VALUE APPROACH | OR | *P*-VALUE APPROACH

Step 4 The critical value for a left-tailed test is $-z_\alpha$. Use Table II to find the critical value.

Because $\alpha = 0.05$, the critical value is $-z_{0.05}$. From Table II (or Table 9.4 on page 371), $z_{0.05} = 1.645$. Hence the critical value is $-z_{0.05} = -1.645$, as shown in Fig. 9.13A.

FIGURE 9.13A



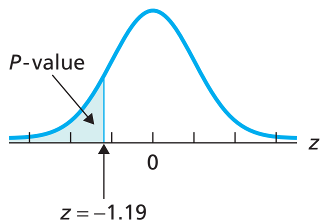
Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

The value of the test statistic found in Step 3 is $z = -1.19$. Figure 9.13A reveals that this value does not fall in the rejection region, so we do not reject H_0. The test results are not statistically significant at the 5% level.

Step 4 Use Table II to obtain the *P*-value.

From Step 3, the value of the test statistic is $z = -1.19$. The test is left tailed, so the P-value is the probability of observing a value of z of -1.19 or less if the null hypothesis is true. That probability equals the shaded area in Fig. 9.13B, which, by Table II, is 0.1170. Hence $P = 0.1170$.

FIGURE 9.13B



Step 5 If $P \le \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P = 0.1170$. Because the P-value exceeds the specified significance level of 0.05, we do not reject H_0. The test results are not statistically significant at the 5% level and (see Table 9.8 on page 378) provide at most weak evidence against the null hypothesis.

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 5% significance level, the data do not provide sufficient evidence to conclude that the mean calcium intake of all adults with incomes below the poverty level is less than the RAI of 1000 mg per day.

SC
Report 9.1

EXAMPLE 9.13 The One-Mean z-Test

Clocking the Cheetah The cheetah (*Acinonyx jubatus*) is the fastest land mammal and is highly specialized to run down prey. The cheetah often exceeds speeds of 60 mph and, according to the online document "Cheetah Conservation in Southern Africa" (*Trade & Environment Database (TED) Case Studies*, Vol. 8, No. 2) by J. Urbaniak, the cheetah is capable of speeds up to 72 mph.

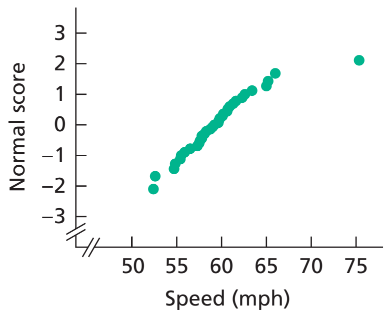
One common estimate of mean top speed for cheetahs is 60 mph. Table 9.11 gives the top speeds, in miles per hour, for a sample of 35 cheetahs.

TABLE 9.11
Top speeds, in miles per hour, for a sample of 35 cheetahs

57.3	57.5	59.0	56.5	61.3
57.6	59.2	65.0	60.1	59.7
62.6	52.6	60.7	62.3	65.2
54.8	55.4	55.5	57.8	58.7
57.8	60.9	75.3	60.6	58.1
55.9	61.6	59.6	59.8	63.4
54.7	60.2	52.4	58.3	66.0

At the 5% significance level, do the data provide sufficient evidence to conclude that the mean top speed of all cheetahs differs from 60 mph? Assume that the population standard deviation of top speeds is 3.2 mph.

FIGURE 9.14
Normal probability plot of the top speeds in Table 9.11



Solution A normal probability plot of the data in Table 9.11, shown in Fig. 9.14, suggests that the top speed of 75.3 mph (third entry in the fifth row) is an outlier. A stem-and-leaf diagram, a boxplot, and a histogram further confirm that 75.3 is an outlier. Thus, as suggested in the third bulleted item in Key Fact 9.7 (page 379), we apply Procedure 9.1 first to the full data set in Table 9.11 and then to that data set with the outlier removed.

Step 1 State the null and alternative hypotheses.

The null and alternative hypotheses are, respectively,

H_0: $\mu = 60$ mph (mean top speed of cheetahs is 60 mph)
H_a: $\mu \neq 60$ mph (mean top speed of cheetahs is not 60 mph),

where μ denotes the mean top speed of all cheetahs. Note that the hypothesis test is two tailed because a does-not-equal sign ($\neq$) appears in the alternative hypothesis.

Step 2 Decide on the significance level, α.

We are to perform the hypothesis test at the 5% significance level, or $\alpha = 0.05$.

Step 3 Compute the value of the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

We have $\mu_0 = 60$, $\sigma = 3.2$, and $n = 35$. From the data in Table 9.11, we find that $\bar{x} = 59.526$. Thus the value of the test statistic is

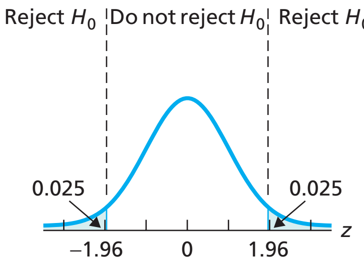
$$z = \frac{59.526 - 60}{3.2/\sqrt{35}} = -0.88.$$

CRITICAL-VALUE APPROACH	OR	*P*-VALUE APPROACH

Step 4 The critical values for a two-tailed test are $\pm z_{\alpha/2}$. Use Table II to find the critical values.

Because $\alpha = 0.05$, we find from Table II (or Table 9.4 or Table IV) the critical values of $\pm z_{0.05/2} = \pm z_{0.025} = \pm 1.96$, as shown in Fig. 9.15A.

FIGURE 9.15A



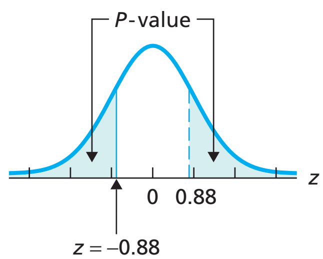
Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

The value of the test statistic found in Step 3 is $z = -0.88$. Figure 9.15A reveals that this value does not fall in the rejection region, so we do not reject H_0. The test results are not statistically significant at the 5% level.

Step 4 Use Table II to obtain the *P*-value.

From Step 3, the value of the test statistic is $z = -0.88$. The test is two tailed, so the P-value is the probability of observing a value of z of 0.88 or greater in magnitude if the null hypothesis is true. That probability equals the shaded area in Fig. 9.15B, which, by Table II, is $2 \cdot 0.1894$ or 0.3788. Hence $P = 0.3788$.

FIGURE 9.15B



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P = 0.3788$. Because the P-value exceeds the specified significance level of 0.05, we do not reject H_0. The test results are not statistically significant at the 5% level and (see Table 9.8 on page 378) provide at most weak evidence against the null hypothesis.

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 5% significance level, the (unabridged) data do not provide sufficient evidence to conclude that the mean top speed of all cheetahs differs from 60 mph.

We have now completed the hypothesis test, using all 35 top speeds in Table 9.11. However, recall that the top speed of 75.3 mph is an outlier. Although in this case, we don't know whether removing this outlier is justified (a common situation), we can still remove it from the sample data and assess the effect on the hypothesis test. With the outlier removed, we determined that the value of the test statistic is $z = -1.71$.

CRITICAL-VALUE APPROACH	OR	*P*-VALUE APPROACH

We see from Fig. 9.15A that the value of the test statistic, $z = -1.71$, for the abridged data does not fall in the rejection region (although it is much closer to the rejection region than the value of the test statistic for the unabridged data, $z = -0.88$). Hence we do not reject H_0. The test results are not statistically significant at the 5% level.

For the abridged data, the P-value is the probability of observing a value of z of 1.71 or greater in magnitude if the null hypothesis is true. Referring to Table II, we find that probability to be $2 \cdot 0.0436$, or 0.0872. Hence $P = 0.0872$.

Because the P-value exceeds the specified significance level of 0.05, we do not reject H_0. The test results are not statistically significant at the 5% level but, as we see from Table 9.8 on page 378, the abridged data do provide moderate evidence against the null hypothesis.

You try it!

Exercise 9.73 on page 388

Interpretation At the 5% significance level, the (abridged) data do not provide sufficient evidence to conclude that the mean top speed of all cheetahs differs from 60 mph. Thus, we see that removing the outlier does not affect the conclusion of this hypothesis test.

Statistical Significance Versus Practical Significance

Recall that the results of a hypothesis test are *statistically significant* if the null hypothesis is rejected at the chosen level of α. Statistical significance means that the data provide sufficient evidence to conclude that the truth is different from the stated null hypothesis. However, it does not necessarily mean that the difference is important in any practical sense.

For example, the manufacturer of a new car, the Orion, claims that a typical car gets 26 miles per gallon. We think that the gas mileage is less. To test our suspicion, we perform the hypothesis test

$$H_0\colon \mu = 26 \text{ mpg (manufacturer's claim)}$$
$$H_a\colon \mu < 26 \text{ mpg (our suspicion)},$$

where μ is the mean gas mileage of all Orions.

We take a random sample of 1000 Orions and find that their mean gas mileage is 25.9 mpg. Assuming $\sigma = 1.4$ mpg, the value of the test statistic for a z-test is $z = -2.26$. This result is statistically significant at the 5% level. Thus, at the 5% significance level, we reject the manufacturer's claim.

What Does It Mean?

Statistical significance does not necessarily imply practical significance!

Because the sample size, 1000, is so large, the sample mean, $\bar{x} = 25.9$ mpg, is probably nearly the same as the population mean. As a result, we rejected the manufacturer's claim because μ is about 25.9 mpg instead of 26 mpg. From a practical point of view, however, the difference between 25.9 mpg and 26 mpg is not important.

The Relation between Hypothesis Tests and Confidence Intervals

Hypothesis tests and confidence intervals are closely related. Consider, for example, a two-tailed hypothesis test for a population mean at the significance level α. In this case, the null hypothesis will be rejected if and only if the value μ_0 given for the mean in the null hypothesis lies outside the $(1-\alpha)$-level confidence interval for μ. You can examine the relation between hypothesis tests and confidence intervals in greater detail in Exercises 9.85–9.87.

THE TECHNOLOGY CENTER

Most statistical technologies have programs that automatically perform a one-mean z-test. In this subsection, we present output and step-by-step instructions for such programs.

EXAMPLE 9.14 Using Technology to Conduct a One-Mean z-Test

Poverty and Dietary Calcium Table 9.10 on page 382 shows the daily calcium intakes for a simple random sample of 18 adults with incomes below the poverty level. Use Minitab, Excel, or the TI-83/84 Plus to decide, at the 5% significance level, whether the data provide sufficient evidence to conclude that the mean calcium intake of all adults with incomes below the poverty level is less than the RAI of 1000 mg per day. Assume that $\sigma = 188$ mg.

Solution Let μ denote the mean calcium intake (per day) of all adults with incomes below the poverty level. We want to perform the hypothesis test

$$H_0\text{: } \mu = 1000 \text{ mg (mean calcium intake is not less than the RAI)}$$
$$H_a\text{: } \mu < 1000 \text{ mg (mean calcium intake is less than the RAI)}$$

at the 5% significance level ($\alpha = 0.05$). Note that the hypothesis test is left tailed.

We applied the one-mean z-test programs to the data, resulting in Output 9.1. Steps for generating that output are presented in Instructions 9.1 at the top of the following page.

OUTPUT 9.1
One-mean z-test on the sample of calcium intakes

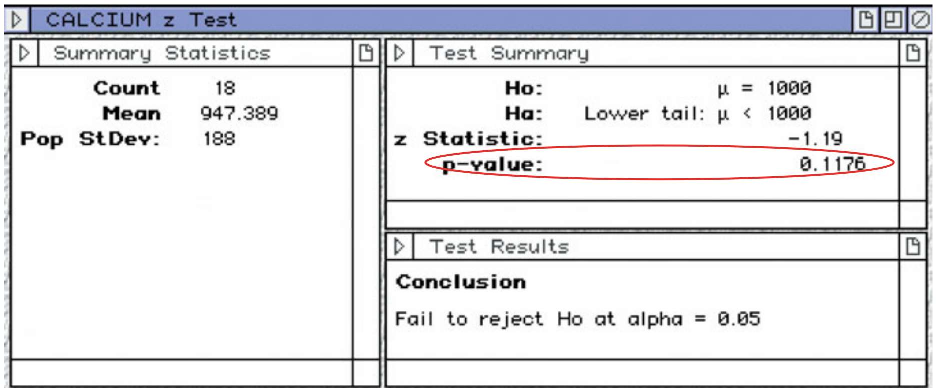
MINITAB

One-Sample Z: CALCIUM

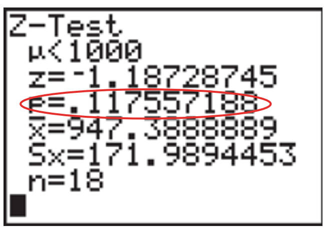
```
Test of mu = 1000 vs < 1000
The assumed standard deviation = 188


                                            95%
                                          Upper
Variable   N   Mean  StDev  SE Mean       Bound      Z       P
CALCIUM   18  947.4  172.0     44.3      1020.3  -1.19   0.118
```

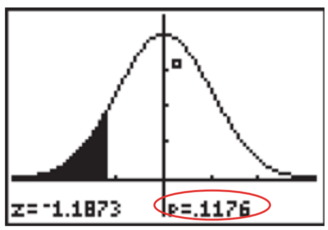
EXCEL



TI-83/84 PLUS



Using **Calculate**



Using **Draw**

As shown in Output 9.1, the P-value for the hypothesis test is 0.118. Because the P-value exceeds the specified significance level of 0.05, we do not reject H_0. At the 5% significance level, the data do not provide sufficient evidence to conclude that the mean calcium intake of all adults with incomes below the poverty level is less than the RAI of 1000 mg per day.

INSTRUCTIONS 9.1 Steps for generating Output 9.1

MINITAB

1. Store the data from Table 9.10 in a column named CALCIUM
2. Choose **Stat ➤ Basic Statistics ➤ 1-Sample Z...**
3. Select the **Samples in columns** option button
4. Click in the **Samples in columns** text box and specify CALCIUM
5. Click in the **Standard deviation** text box and type 188
6. Check the **Perform hypothesis test** check box
7. Click in the **Hypothesized mean** text box and type 1000
8. Click the **Options...** button
9. Click the arrow button at the right of the **Alternative** drop-down list box and select **less than**
10. Click **OK** twice

EXCEL

1. Store the data from Table 9.10 in a range named CALCIUM
2. Choose **DDXL ➤ Hypothesis Tests**
3. Select **1 Var z Test** from the **Function type** drop-down box
4. Specify CALCIUM in the **Quantitative Variable** text box
5. Click **OK**
6. Click the **Set μ0 and sd** button
7. Click in the **Hypothesized μ0** text box and type 1000
8. Click in the **Population std dev** text box and type 188
9. Click **OK**
10. Click the **0.05** button
11. Click the **μ < μ0** button
12. Click the **Compute** button

TI-83/84 PLUS

1. Store the data from Table 9.10 in a list named CALCI
2. Press **STAT**, arrow over to **TESTS**, and press **1**
3. Highlight **Data** and press **ENTER**
4. Press the down-arrow key, type 1000 for μ_0, and press **ENTER**
5. Type 188 for σ and press **ENTER**
6. Press **2nd ➤ LIST**
7. Arrow down to CALCI and press **ENTER** three times
8. Highlight **<** μ_0 and press **ENTER**
9. Press the down-arrow key, highlight **Calculate** or **Draw**, and press **ENTER**

Exercises 9.4

Understanding the Concepts and Skills

9.64 Explain why considering outliers is important when you are conducting a one-mean z-test.

9.65 Each part of this exercise provides a scenario for a hypothesis test for a population mean. Decide whether the z-test is an appropriate method for conducting the hypothesis test. Assume that the population standard deviation is known in each case.

a. Preliminary data analyses reveal that the sample data contain no outliers but that the distribution of the variable under consideration is probably highly skewed. The sample size is 24.

b. Preliminary data analyses reveal that the sample data contain no outliers but that the distribution of the variable under consideration is probably mildly skewed. The sample size is 70.

9.66 Each part of this exercise provides a scenario for a hypothesis test for a population mean. Decide whether the z-test is an appropriate method for conducting the hypothesis test. Assume that the population standard deviation is known in each case.

a. A normal probability plot of the sample data shows no outliers and is quite linear. The sample size is 12.

b. Preliminary data analyses reveal that the sample data contain an outlier. It is determined that the outlier is a legitimate observation and should not be removed. The sample size is 17.

*In each of Exercises **9.67–9.72**, we have provided a sample mean, sample size, and population standard deviation. In each case, use the one-mean z-test to perform the required hypothesis test at the 5% significance level.*

9.67 $\bar{x} = 20$, $n = 32$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu < 22$

9.68 $\bar{x} = 21$, $n = 32$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu < 22$

9.69 $\bar{x} = 24$, $n = 15$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu > 22$

9.70 $\bar{x} = 23$, $n = 15$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu > 22$

9.71 $\bar{x} = 23$, $n = 24$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu \neq 22$

9.72 $\bar{x} = 20$, $n = 24$, $\sigma = 4$, H_0: $\mu = 22$, H_a: $\mu \neq 22$

*Preliminary data analyses indicate that applying the z-test (Procedure 9.1 on page 380) in Exercises **9.73–9.78** is reasonable.*

9.73 Toxic Mushrooms? Cadmium, a heavy metal, is toxic to animals. Mushrooms, however, are able to absorb and accumulate cadmium at high concentrations. The Czech and Slovak governments have set a safety limit for cadmium in dry vegetables at 0.5 part per million (ppm). M. Melgar et al. measured the cadmium levels in a random sample of the edible mushroom *Boletus pinicola* and published the results in the paper "Influence of Some Factors in Toxicity and Accumulation of Cd from Edible Wild Macrofungi in NW Spain" (*Journal of Environmental Science and Health*, Vol. B33(4), pp. 439–455). Here are the data.

0.24	0.59	0.62	0.16	0.77	1.33
0.92	0.19	0.33	0.25	0.59	0.32

At the 5% significance level, do the data provide sufficient evidence to conclude that the mean cadmium level in *Boletus pinicola* mushrooms is greater than the government's recommended limit of 0.5 ppm? Assume that the population standard deviation of cadmium levels in *Boletus pinicola* mushrooms is 0.37 ppm. (*Note:* The sum of the data is 6.31 ppm.)

9.74 Agriculture Books. The R. R. Bowker Company collects information on the retail prices of books and publishes the data in *The Bowker Annual Library and Book Trade Almanac*. In 2005, the mean retail price of agriculture books was \$57.61.

This year's retail prices for 28 randomly selected agriculture books are shown in the following table.

59.54	67.70	57.10	46.11	46.86	62.87	66.40
52.08	37.67	50.47	60.42	38.14	58.21	47.35
50.45	71.03	48.14	66.18	59.36	41.63	53.66
49.95	59.08	58.04	46.65	66.76	50.61	66.68

At the 10% significance level, do the data provide sufficient evidence to conclude that this year's mean retail price of agriculture books has changed from the 2005 mean? Assume that the population standard deviation of prices for this year's agriculture books is \$8.45. (*Note:* The sum of the data is \$1539.14.)

9.75 Iron Deficiency? Iron is essential to most life forms and to normal human physiology. It is an integral part of many proteins and enzymes that maintain good health. Recommendations for iron are provided in *Dietary Reference Intakes*, developed by the Institute of Medicine of the National Academy of Sciences. The recommended dietary allowance (RDA) of iron for adult females under the age of 51 is 18 milligrams (mg) per day. The following iron intakes, in milligrams, were obtained during a 24-hour period for 45 randomly selected adult females under the age of 51.

15.0	18.1	14.4	14.6	10.9	18.1	18.2	18.3	15.0
16.0	12.6	16.6	20.7	19.8	11.6	12.8	15.6	11.0
15.3	9.4	19.5	18.3	14.5	16.6	11.5	16.4	12.5
14.6	11.9	12.5	18.6	13.1	12.1	10.7	17.3	12.4
17.0	6.3	16.8	12.5	16.3	14.7	12.7	16.3	11.5

At the 1% significance level, do the data suggest that adult females under the age of 51 are, on average, getting less than the RDA of 18 mg of iron? Assume that the population standard deviation is 4.2 mg. (*Note:* $\bar{x} = 14.68$ mg.)

9.76 Early-Onset Dementia. Dementia is the loss of the intellectual and social abilities severe enough to interfere with judgment, behavior, and daily functioning. Alzheimer's disease is the most common type of dementia. In the article "Living with Early Onset Dementia: Exploring the Experience and Developing Evidence-Based Guidelines for Practice" (*Alzheimer's Care Quarterly*, Vol. 5, Issue 2, pp. 111–122), P. Harris and J. Keady explored the experience and struggles of people diagnosed with dementia and their families. A simple random sample of 21 people with early-onset dementia gave the following data on age at diagnosis, in years.

60	58	52	58	59	58	51
61	54	59	55	53	44	46
47	42	56	57	49	41	43

At the 1% significance level, do the data provide sufficient evidence to conclude that the mean age at diagnosis of all people with early-onset dementia is less than 55 years old? Assume that the population standard deviation is 6.8 years. (*Note:* $\bar{x} = 52.5$ years.)

9.77 Serving Time. According to the Bureau of Crime Statistics and Research of Australia, as reported on *Lawlink*, the mean length of imprisonment for motor-vehicle-theft offenders in Australia is 16.7 months. One hundred randomly selected motor-vehicle-theft offenders in Sydney, Australia, had a mean length of imprisonment of 17.8 months. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean length of imprisonment for motor-vehicle-theft offenders in Sydney differs from the national mean in Australia? Assume that the population standard deviation of the lengths of imprisonment for motor-vehicle-theft offenders in Sydney is 6.0 months.

9.78 Worker Fatigue. A study by M. Chen et al. titled "Heat Stress Evaluation and Worker Fatigue in a Steel Plant" (*American Industrial Hygiene Association*, Vol. 64, pp. 352–359) assessed fatigue in steel-plant workers due to heat stress. A random sample of 29 casting workers had a mean post-work heart rate of 78.3 beats per minute (bpm). At the 5% significance level, do the data provide sufficient evidence to conclude that the mean post-work heart rate for casting workers exceeds the normal resting heart rate of 72 bpm? Assume that the population standard deviation of post-work heart rates for casting workers is 11.2 bpm.

9.79 Job Gains and Losses. In the article "Business Employment Dynamics: New Data on Gross Job Gains and Losses" (*Monthly Labor Review*, Vol. 127, Issue 4, pp. 29–42), J. Spletzer et al. examined gross job gains and losses as a percentage of the average of previous and current employment figures. A simple random sample of 20 quarters provided the net percentage gains (losses are negative gains) for jobs as presented on the WeissStats CD. Use the technology of your choice to do the following.

a. Decide whether, on average, the net percentage gain for jobs exceeds 0.2. Assume a population standard deviation of 0.42. Apply the one-mean z-test with a 5% significance level.
b. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.
c. Remove the outliers (if any) from the data and then repeat part (a).
d. Comment on the advisability of using the z-test here.

9.80 Hotels and Motels. The daily charges, in dollars, for a sample of 15 hotels and motels operating in South Carolina are provided on the WeissStats CD. The data were found in the report *South Carolina Statistical Abstract*, sponsored by the South Carolina Budget and Control Board.

a. Use the one-mean z-test to decide, at the 5% significance level, whether the data provide sufficient evidence to conclude that the mean daily charge for hotels and motels operating in South Carolina is less than \$75. Assume a population standard deviation of \$22.40.
b. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.
c. Remove the outliers (if any) from the data and then repeat part (a).
d. Comment on the advisability of using the z-test here.

Working with Large Data Sets

9.81 Body Temperature. A study by researchers at the University of Maryland addressed the question of whether the mean body temperature of humans is 98.6°F. The results of the study by P. Mackowiak et al. appeared in the article "A Critical Appraisal of 98.6°F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich" (*Journal of the American Medical Association*, Vol. 268, pp. 1578–1580). Among other data, the researchers obtained the body temperatures of 93 healthy humans, which we provide on the WeissStats CD. Use the technology of your choice to do the following.

a. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.

b. Based on your results from part (a), can you reasonably apply the one-mean z-test to the data? Explain your reasoning.

c. At the 1% significance level, do the data provide sufficient evidence to conclude that the mean body temperature of healthy humans differs from 98.6°F? Assume that $\sigma = 0.63$°F.

9.82 Teacher Salaries. The Educational Resource Service publishes information about wages and salaries in the public schools system in *National Survey of Salaries and Wages in Public Schools*. The mean annual salary of (public) classroom teachers is \$49.0 thousand. A random sample of 90 classroom teachers in Hawaii yielded the annual salaries, in thousands of dollars, presented on the WeissStats CD. Use the technology of your choice to do the following.

a. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.

b. Based on your results from part (a), can you reasonably apply the one-mean z-test to the data? Explain your reasoning.

c. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean annual salary of classroom teachers in Hawaii is greater than the national mean? Assume that the standard deviation of annual salaries for all classroom teachers in Hawaii is \$9.2 thousand.

9.83 Cell Phones. The number of cell phone users has increased dramatically since 1987. According to the *Semi-annual Wireless Survey*, published by the Cellular Telecommunications & Internet Association, the mean local monthly bill for cell phone users in the United States was \$49.94 in 2007. Last year's local monthly bills, in dollars, for a random sample of 75 cell phone users are given on the WeissStats CD. Use the technology of your choice to do the following.

a. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.

b. At the 5% significance level, do the data provide sufficient evidence to conclude that last year's mean local monthly bill for cell phone users decreased from the 2007 mean of \$49.94? Assume that the population standard deviation of last year's local monthly bills for cell phone users is \$25.

c. Remove the two outliers from the data and repeat parts (a) and (b).

d. State your conclusions regarding the hypothesis test.

Extending the Concepts and Skills

9.84 Class Project: Quality Assurance. This exercise can be done individually or, better yet, as a class project. For the pretzel-packaging hypothesis test in Example 9.1 on page 360, the null and alternative hypotheses are, respectively,

H_0: $\mu = 454$ g (machine is working properly)

H_a: $\mu \neq 454$ g (machine is not working properly),

where μ is the mean net weight of all bags of pretzels packaged. The net weights are normally distributed with a standard deviation of 7.8 g.

a. Assuming that the null hypothesis is true, simulate 100 samples of 25 net weights each.

b. Suppose that the hypothesis test is performed at the 5% significance level. Of the 100 samples obtained in part (a), roughly how many would you expect to lead to rejection of the null hypothesis? Explain your answer.

c. Of the 100 samples obtained in part (a), determine the number that lead to rejection of the null hypothesis.

d. Compare your answers from parts (b) and (c), and comment on any observed difference.

9.85 Two-Tailed Hypothesis Tests and CIs. As we mentioned on page 386, the following relationship holds between hypothesis tests and confidence intervals for one-mean z-procedures: For a two-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu \neq \mu_0$ if and only if μ_0 lies outside the $(1 - \alpha)$-level confidence interval for μ. In each case, illustrate the preceding relationship by obtaining the appropriate one-mean z-interval (Procedure 8.1 on page 330) and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.74 **b.** Exercise 9.77

9.86 Left-Tailed Hypothesis Tests and CIs. In Exercise 8.47 on page 337, we introduced one-sided one-mean z-intervals. The following relationship holds between hypothesis tests and confidence intervals for one-mean z-procedures: For a left-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu < \mu_0$ if and only if μ_0 is greater than the $(1 - \alpha)$-level upper confidence bound for μ. In each case, illustrate the preceding relationship by obtaining the appropriate upper confidence bound and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.75 **b.** Exercise 9.76

9.87 Right-Tailed Hypothesis Tests and CIs. In Exercise 8.47 on page 337, we introduced one-sided one-mean z-intervals. The following relationship holds between hypothesis tests and confidence intervals for one-mean z-procedures: For a right-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu > \mu_0$ if and only if μ_0 is less than the $(1 - \alpha)$-level lower confidence bound for μ. In each case, illustrate the preceding relationship by obtaining the appropriate lower confidence bound and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.73 **b.** Exercise 9.78

9.5 Hypothesis Tests for One Population Mean When σ Is Unknown

In Section 9.4, you learned how to perform a hypothesis test for one population mean when the population standard deviation, σ, is known. However, as we have mentioned, the population standard deviation is usually not known.

To develop a hypothesis-testing procedure for a population mean when σ is unknown, we begin by recalling Key Fact 8.5: If a variable x of a population is normally distributed with mean μ, then, for samples of size n, the studentized version of $\bar{x}$,

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

has the t-distribution with $n - 1$ degrees of freedom.

Because of Key Fact 8.5, we can perform a hypothesis test for a population mean when the population standard deviation is unknown by proceeding in essentially the same way as when it is known. The only difference is that we invoke a t-distribution instead of the standard normal distribution. Specifically, for a test with null hypothesis H_0: $\mu = \mu_0$, we employ the variable

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

as our test statistic and use the t-table, Table IV, to obtain the critical value(s) or P-value. We call this hypothesis-testing procedure the **one-mean *t*-test** or, when no confusion can arise, simply the ***t*-test.**[†]

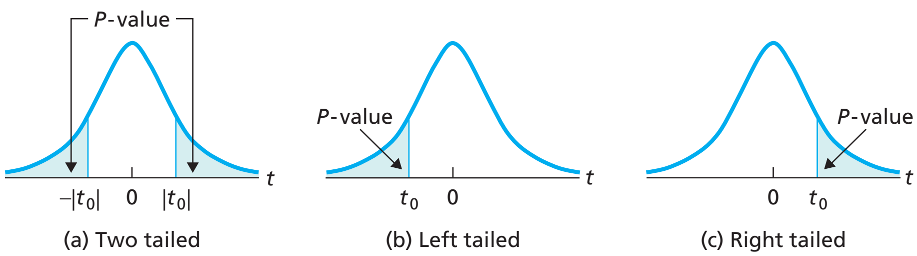
P-Values for a *t*-Test[‡]

Before presenting a step-by-step procedure for conducting a (one-mean) t-test, we need to discuss P-values for such a test. P-values for a t-test are obtained in a manner similar to that for a z-test.

As we know, if the null hypothesis is true, the test statistic for a t-test has the t-distribution with $n - 1$ degrees of freedom, so its probabilities equal areas under the t-curve with df $= n - 1$. Thus, if we let t_0 be the observed value of the test statistic t, we determine the P-value as follows.

- *Two-tailed test:* The P-value equals the probability of observing a value of the test statistic t that is at least as large in magnitude as the value actually observed, which is the area under the t-curve that lies outside the interval from $-|t_0|$ to $|t_0|$, as shown in Fig. 9.16(a).
- *Left-tailed test:* The P-value equals the probability of observing a value of the test statistic t that is as small as or smaller than the value actually observed, which is the area under the t-curve that lies to the left of t_0, as shown in Fig. 9.16(b).
- *Right-tailed test:* The P-value equals the probability of observing a value of the test statistic t that is as large as or larger than the value actually observed, which is the area under the t-curve that lies to the right of t_0, as shown in Fig. 9.16(c).

FIGURE 9.16
P-value for a *t*-test if the test is (a) two tailed, (b) left tailed, or (c) right tailed



Estimating the *P*-Value of a *t*-Test

To obtain the exact P-value of a t-test, we need statistical software or a statistical calculator. However, we can use t-tables, such as Table IV, to estimate the

[†] The one-mean t-test is also known as the **one-sample *t*-test** and the **one-variable *t*-test.** We prefer "one-mean" because it makes clear the parameter being tested.

[‡] Those concentrating on the critical-value approach to hypothesis testing can skip to the subsection on the "The One-Mean t-Test," beginning on page 393.

P-value of a t-test, and an estimate of the P-value is usually sufficient for deciding whether to reject the null hypothesis.

For instance, consider a right-tailed t-test with $n = 15$, $\alpha = 0.05$, and a value of the test statistic of $t = 3.458$. For df $= 15 - 1 = 14$, the t-value 3.458 is larger than any t-value in Table IV, the largest one being $t_{0.005} = 2.977$ (which means that the area under the t-curve that lies to the right of 2.977 equals 0.005). This fact, in turn, implies that the area to the right of 3.458 is less than 0.005; in other words, $P < 0.005$. Because the P-value is less than the designated significance level of 0.05, we reject H_0.

Example 9.15 provides two more illustrations of how Table IV can be used to estimate the P-value of a t-test.

EXAMPLE 9.15 Using Table IV to Estimate the *P*-Value of a *t*-Test

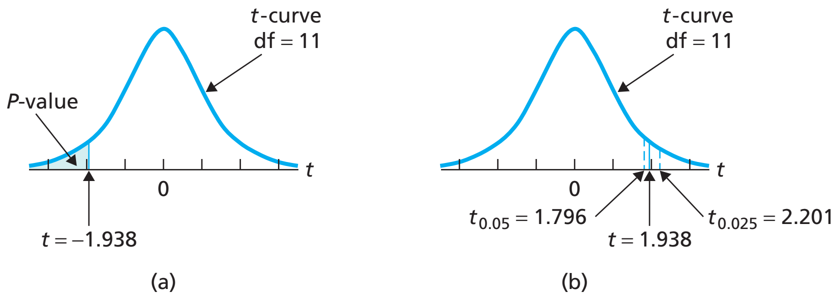
Use Table IV to estimate the P-value of each one-mean t-test.

a. Left-tailed test, $n = 12$, and $t = -1.938$
b. Two-tailed test, $n = 25$, and $t = -0.895$

Solution

a. Because the test is left tailed, the P-value is the area under the t-curve with df $= 12 - 1 = 11$ that lies to the left of -1.938, as shown in Fig. 9.17(a).

FIGURE 9.17
Estimating the P-value of a left-tailed t-test with a sample size of 12 and test statistic $t = -1.938$



A t-curve is symmetric about 0, so the area to the left of -1.938 equals the area to the right of 1.938, which we can estimate by using Table IV. In the df $= 11$ row of Table IV, the two t-values that straddle 1.938 are $t_{0.05} = 1.796$ and $t_{0.025} = 2.201$. Therefore the area under the t-curve that lies to the right of 1.938 is between 0.025 and 0.05, as shown in Fig. 9.17(b).

Consequently, the area under the t-curve that lies to the left of -1.938 is also between 0.025 and 0.05, so $0.025 < P < 0.05$. Hence we can reject H_0 at any significance level of 0.05 or larger, and we cannot reject H_0 at any significance level of 0.025 or smaller. For significance levels between 0.025 and 0.05, Table IV is not sufficiently detailed to help us to decide whether to reject H_0.†

b. Because the test is two tailed, the P-value is the area under the t-curve with df $= 25 - 1 = 24$ that lies either to the left of -0.895 or to the right of 0.895, as shown in Fig. 9.18(a).

Because a t-curve is symmetric about 0, the areas to the left of -0.895 and to the right of 0.895 are equal. In the df $= 24$ row of Table IV, 0.895 is smaller than any other t-value, the smallest being $t_{0.10} = 1.318$. The area under the t-curve that lies to the right of 0.895, therefore, is greater than 0.10, as shown in Fig. 9.18(b).

† This latter case is an example of a P-value estimate that is not good enough. In such cases, use statistical software or a statistical calculator to find the exact P-value.

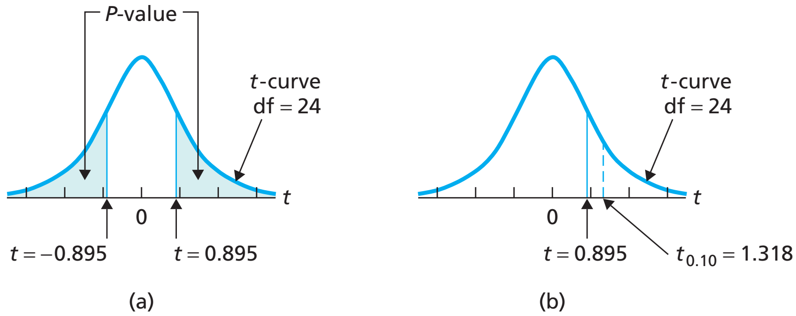


FIGURE 9.18
Estimating the P-value of a two-tailed t-test with a sample size of 25 and test statistic $t = -0.895$

Consequently, the area under the t-curve that lies either to the left of -0.895 or to the right of 0.895 is greater than 0.20, so $P > 0.20$. Hence we cannot reject H_0 at any significance level of 0.20 or smaller. For significance levels larger than 0.20, Table IV is not sufficiently detailed to help us to decide whether to reject H_0.

You try it!

Exercise 9.89 on page 397

The One-Mean *t*-Test

We now present, on the next page, Procedure 9.2, a step-by-step method for performing a one-mean t-test. As you can see, Procedure 9.2 includes both the critical-value approach for a one-mean t-test and the P-value approach for a one-mean t-test.

Properties and guidelines for use of the t-test are the same as those for the z-test, as given in Key Fact 9.7 on page 379. In particular, the t-test is robust to moderate violations of the normality assumption but, even for large samples, can sometimes be unduly affected by outliers because the sample mean and sample standard deviation are not resistant to outliers.

APPLET

Applet 9.1

EXAMPLE 9.16 The One-Mean *t*-Test

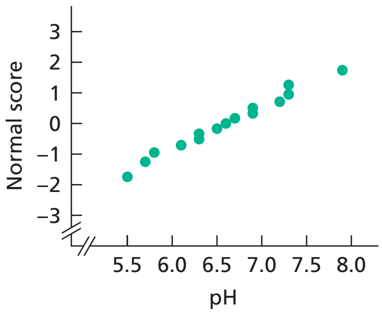
Acid Rain and Lake Acidity Acid rain from the burning of fossil fuels has caused many of the lakes around the world to become acidic. The biology in these lakes often collapses because of the rapid and unfavorable changes in water chemistry. A lake is classified as nonacidic if it has a pH greater than 6.

A. Marchetto and A. Lami measured the pH of high mountain lakes in the Southern Alps and reported their findings in the paper "Reconstruction of pH by Chrysophycean Scales in Some Lakes of the Southern Alps" (*Hydrobiologia*, Vol. 274, pp. 83–90). Table 9.12 shows the pH levels obtained by the researchers for 15 lakes. At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic?

TABLE 9.12
pH levels for 15 lakes

7.2	7.3	6.1	6.9	6.6
7.3	6.3	5.5	6.3	6.5
5.7	6.9	6.7	7.9	5.8

FIGURE 9.19
Normal probability plot of pH levels in Table 9.12



Solution Figure 9.19, a normal probability plot of the data in Table 9.12, reveals no outliers and is quite linear. Consequently, we can apply Procedure 9.2 to conduct the required hypothesis test.

Step 1 State the null and alternative hypotheses.

Let μ denote the mean pH level of all high mountain lakes in the Southern Alps. Then the null and alternative hypotheses are, respectively,

$$H_0\text{: } \mu = 6 \text{ (on average, the lakes are acidic)}$$
$$H_a\text{: } \mu > 6 \text{ (on average, the lakes are nonacidic).}$$

Note that the hypothesis test is right tailed.

PROCEDURE 9.2 One-Mean *t*-Test

Purpose To perform a hypothesis test for a population mean, μ

Assumptions

1. Simple random sample
2. Normal population or large sample
3. σ unknown

Step 1 **The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is**

$$\begin{array}{ccccc} H_a\text{: } \mu \neq \mu_0 & & H_a\text{: } \mu < \mu_0 & & H_a\text{: } \mu > \mu_0 \\ \text{(Two tailed)} & \text{or} & \text{(Left tailed)} & \text{or} & \text{(Right tailed)} \end{array}$$

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

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

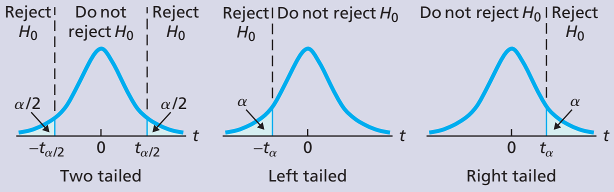
and denote that value t_0.

CRITICAL-VALUE APPROACH	OR	*P*-VALUE APPROACH

Step 4 **The critical value(s) are**

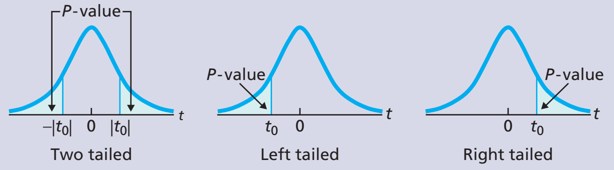
$$\begin{array}{ccccc} \pm t_{\alpha/2} & & -t_\alpha & & t_\alpha \\ \text{(Two tailed)} & \text{or} & \text{(Left tailed)} & \text{or} & \text{(Right tailed)} \end{array}$$

with df $= n - 1$. Use Table IV to find the critical value(s).



Step 5 **If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.**

Step 4 **The t-statistic has df $= n - 1$. Use Table IV to estimate the P-value, or obtain it exactly by using technology.**



Step 5 **If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.**

Step 6 **Interpret the results of the hypothesis test.**

Note: The hypothesis test is exact for normal populations and is approximately correct for large samples from nonnormal populations.

Step 2 **Decide on the significance level, α.**

We are to perform the test at the 5% significance level, so $\alpha = 0.05$.

Step 3 **Compute the value of the test statistic**

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

We have $\mu_0 = 6$ and $n = 15$ and calculate the mean and standard deviation of the sample data in Table 9.12 as 6.6 and 0.672, respectively. Hence the value of the test statistic is

$$t = \frac{6.6 - 6}{0.672/\sqrt{15}} = 3.458.$$

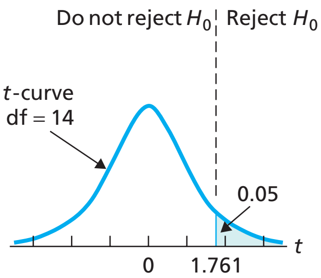
CRITICAL-VALUE APPROACH OR ***P*-VALUE APPROACH**

CRITICAL-VALUE APPROACH

Step 4 The critical value for a right-tailed test is t_α with df $= n - 1$. Use Table IV to find the critical value.

We have $n = 15$ and $\alpha = 0.05$. Table IV shows that for df $= 15 - 1 = 14$, $t_{0.05} = 1.761$. See Fig. 9.20A.

FIGURE 9.20A



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

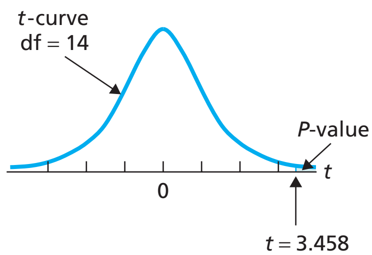
The value of the test statistic, found in Step 3, is $t = 3.458$. Figure 9.20A reveals that it falls in the rejection region. Consequently, we reject H_0. The test results are statistically significant at the 5% level.

***P*-VALUE APPROACH**

Step 4 The t-statistic has df $= n - 1$. Use Table IV to estimate the P-value, or obtain it exactly by using technology.

From Step 3, the value of the test statistic is $t = 3.458$. The test is right tailed, so the P-value is the probability of observing a value of t of 3.458 or greater if the null hypothesis is true. That probability equals the shaded area in Fig. 9.20B.

FIGURE 9.20B



We have $n = 15$, and so df $= 15 - 1 = 14$. From Fig. 9.20B and Table IV, $P < 0.005$. (Using technology, we obtain $P = 0.00192$.)

Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P < 0.005$. Because the P-value is less than the specified significance level of 0.05, we reject H_0. The test results are statistically significant at the 5% level and (see Table 9.8 on page 378) provide very strong evidence against the null hypothesis.

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic.

SC Report 9.2

You try it!

Exercise 9.101 on page 397

THE TECHNOLOGY CENTER

Most statistical technologies have programs that automatically perform a one-mean t-test. In this subsection, we present output and step-by-step instructions for such programs.

EXAMPLE 9.17 Using Technology to Conduct a One-Mean t-Test

Acid Rain and Lake Acidity Table 9.12 on page 393 gives the pH levels of a sample of 15 lakes in the Southern Alps. Use Minitab, Excel, or the TI-83/84 Plus to decide, at the 5% significance level, whether the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic.

Solution Let μ denote the mean pH level of all high mountain lakes in the Southern Alps. We want to perform the hypothesis test

$$H_0\colon \mu = 6 \text{ (on average, the lakes are acidic)}$$
$$H_a\colon \mu > 6 \text{ (on average, the lakes are nonacidic)}$$

at the 5% significance level. Note that the hypothesis test is right tailed.

We applied the one-mean t-test programs to the data, resulting in Output 9.2. Steps for generating that output are presented in Instructions 9.2.

OUTPUT 9.2
One-mean t-test on the sample of pH levels

MINITAB

One-Sample T: PH

```
Test of mu = 6 vs > 6

                                          95%
                                         Lower
Variable   N   Mean  StDev  SE Mean    Bound     T       P
PH        15  6.600  0.672    0.173    6.294  3.46   0.002
```

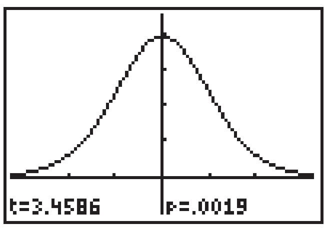
EXCEL

```
PH t Test
Summary Statistics          Test Summary
     Count   15                     Ho:              μ = 6
      Mean   6.6                    Ha:   Upper tail: μ > 6
   Std Dev   0.672                  df:                 14
 Std Error   0.173        t Statistic:                3.46
                              p-value:              0.0019
                          Test Results
                          Conclusion
                          Reject Ho at alpha = 0.05
```

TI-83/84 PLUS

```
T-Test
 μ>6
 t=3.458616098
 p=.0019191607
 x̄=6.6
 Sx=.6718843438
 n=15
```

Using **Calculate**



Using **Draw**

As shown in Output 9.2, the P-value for the hypothesis test is 0.002. The P-value is less than the specified significance level of 0.05, so we reject H_0. At the 5% significance level, the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic.

INSTRUCTIONS 9.2 Steps for generating Output 9.2

MINITAB

1. Store the data from Table 9.12 in a column named PH
2. Choose **Stat ➤ Basic Statistics ➤ 1-Sample t...**
3. Select the **Samples in columns** option button
4. Click in the **Samples in columns** text box and specify PH
5. Check the **Perform hypothesis test** check box
6. Click in the **Hypothesized mean** text box and type 6
7. Click the **Options...** button
8. Click the arrow button at the right of the **Alternative** drop-down list box and select **greater than**
9. Click **OK** twice

EXCEL

1. Store the data from Table 9.12 in a range named PH
2. Choose **DDXL ➤ Hypothesis Tests**
3. Select **1 Var t Test** from the **Function type** drop-down box
4. Specify PH in the **Quantitative Variable** text box
5. Click **OK**
6. Click the **Set μ0** button and type 6
7. Click **OK**
8. Click the **0.05** button
9. Click the **μ > μ0** button
10. Click the **Compute** button

TI-83/84 PLUS

1. Store the data from Table 9.12 in a list named PH
2. Press **STAT**, arrow over to **TESTS**, and press **2**
3. Highlight **Data** and press **ENTER**
4. Press the down-arrow key, type 6 for μ_0, and press **ENTER**
5. Press **2nd ➤ LIST**
6. Arrow down to PH and press **ENTER** three times
7. Highlight **>** μ_0 and press **ENTER**
8. Press the down-arrow key, highlight **Calculate** or **Draw**, and press **ENTER**

Exercises 9.5

Understanding the Concepts and Skills

9.88 What is the difference in assumptions between the one-mean t-test and the one-mean z-test?

*Exercises **9.89–9.94** pertain to P-values for a one-mean t-test. For each exercise, do the following tasks.*
a. Use Table IV in Appendix A to estimate the P-value.
b. Based on your estimate in part (a), state at which significance levels the null hypothesis can be rejected, at which significance levels it cannot be rejected, and at which significance levels it is not possible to decide.

9.89 Right-tailed test, $n = 20$, and $t = 2.235$

9.90 Right-tailed test, $n = 11$, and $t = 1.246$

9.91 Left-tailed test, $n = 10$, and $t = -3.381$

9.92 Left-tailed test, $n = 30$, and $t = -1.572$

9.93 Two-tailed test, $n = 17$, and $t = -2.733$

9.94 Two-tailed test, $n = 8$, and $t = 3.725$

*In each of Exercises **9.95–9.100**, we have provided a sample mean, sample standard deviation, and sample size. In each case, use the one-mean t-test to perform the required hypothesis test at the 5% significance level.*

9.95 $\bar{x} = 20$, $s = 4$, $n = 32$, H_0: $\mu = 22$, H_a: $\mu < 22$

9.96 $\bar{x} = 21$, $s = 4$, $n = 32$, H_0: $\mu = 22$, H_a: $\mu < 22$

9.97 $\bar{x} = 24$, $s = 4$, $n = 15$, H_0: $\mu = 22$, H_a: $\mu > 22$

9.98 $\bar{x} = 23$, $s = 4$, $n = 15$, H_0: $\mu = 22$, H_a: $\mu > 22$

9.99 $\bar{x} = 23$, $s = 4$, $n = 24$, H_0: $\mu = 22$, H_a: $\mu \neq 22$

9.100 $\bar{x} = 20$, $s = 4$, $n = 24$, H_0: $\mu = 22$, H_a: $\mu \neq 22$

*Preliminary data analyses indicate that you can reasonably use a t-test to conduct each of the hypothesis tests required in Exercises **9.101–9.106**.*

9.101 TV Viewing. According to *Communications Industry Forecast & Report*, published by Veronis Suhler Stevenson, the average person watched 4.55 hours of television per day in 2005. A random sample of 20 people gave the following number of hours of television watched per day for last year.

1.0	4.6	5.4	3.7	5.2
1.7	6.1	1.9	7.6	9.1
6.9	5.5	9.0	3.9	2.5
2.4	4.7	4.1	3.7	6.2

At the 10% significance level, do the data provide sufficient evidence to conclude that the amount of television watched per day last year by the average person differed from that in 2005? (*Note:* $\bar{x} = 4.760$ hours and $s = 2.297$ hours.)

9.102 Golf Robots. Serious golfers and golf equipment companies sometimes use golf equipment testing labs to obtain precise information about particular club heads, club shafts, and golf balls. One golfer requested information about the Jazz Fat Cat 5-iron from Golf Laboratories, Inc. The company tested the club by using a robot to hit a Titleist NXT Tour ball six times with a head velocity of 85 miles per hour. The golfer wanted a club that, on average, would hit the ball more than 180 yards at that club speed. The total yards each ball traveled was as follows.

180	187	181	182	185	181

a. At the 5% significance level, do the data provide sufficient evidence to conclude that the club does what the golfer wants? (*Note:* The sample mean and sample standard deviation of the data are 182.7 yards and 2.7 yards, respectively.)

b. Repeat part (a) for a test at the 1% significance level.

9.103 Brewery Effluent and Crops. Because many industrial wastes contain nutrients that enhance crop growth, efforts are being made for environmental purposes to use such wastes on agricultural soils. Two researchers, M. Ajmal and A. Khan, reported their findings on experiments with brewery wastes used for agricultural purposes in the article "Effects of Brewery Effluent on Agricultural Soil and Crop Plants" (*Environmental Pollution (Series A)*, 33, pp. 341–351). The researchers studied the physico-chemical properties of effluent from Mohan Meakin Breweries Ltd. (MMBL), Ghazibad, UP, India, and "... its effects on the physico-chemical characteristics of agricultural soil, seed germination pattern, and the growth of two common crop plants." They assessed the impact of using different concentrations of the effluent: 25%, 50%, 75%, and 100%. The following data, based on the results of the study, provide the percentages of limestone in the soil obtained by using 100% effluent.

2.41	2.31	2.54	2.28	2.72
2.60	2.51	2.51	2.42	2.70

Do the data provide sufficient evidence to conclude, at the 1% level of significance, that the mean available limestone in soil treated with 100% MMBL effluent exceeds 2.30%, the percentage ordinarily found? (*Note:* $\bar{x} = 2.5$ and $s = 0.149$.)

9.104 Apparel and Services. According to the document *Consumer Expenditures*, a publication of the Bureau of Labor Statistics, the average consumer unit spent $1874 on apparel and services in 2006. That same year, 25 consumer units in the Northeast had the following annual expenditures, in dollars, on apparel and services.

1417	1595	2158	1820	1411
2361	2371	2330	1749	1872
2826	2167	2304	1998	2582
1982	1903	2405	1660	2150
2128	1889	2251	2340	1850

At the 5% significance level, do the data provide sufficient evidence to conclude that the 2006 mean annual expenditure on apparel and services for consumer units in the Northeast differed from the national mean of $1874? (*Note:* The sample mean and sample standard deviation of the data are $2060.76 and $350.90, respectively.)

9.105 Ankle Brachial Index. The ankle brachial index (ABI) compares the blood pressure of a patient's arm to the blood pressure of the patient's leg. The ABI can be an indicator of different diseases, including arterial diseases. A healthy (or normal) ABI is 0.9 or greater. In a study by M. McDermott et al. titled "Sex Differences in Peripheral Arterial Disease: Leg Symptoms and Physical Functioning" (*Journal of the American Geriatrics Society*, Vol. 51, No. 2, pp. 222–228), the researchers obtained the ABI of 187 women with peripheral arterial disease. The results were a mean ABI of 0.64 with a standard deviation of 0.15. At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, women with peripheral arterial disease have an unhealthy ABI?

9.106 Active Management of Labor. Active management of labor (AML) is a group of interventions designed to help reduce the length of labor and the rate of cesarean deliveries. Physicians from the Department of Obstetrics and Gynecology at the University of New Mexico Health Sciences Center were interested in determining whether AML would also translate into a reduced cost for delivery. The results of their study can be found in Rogers et al., "Active Management of Labor: A Cost Analysis of a Randomized Controlled Trial" (*Western Journal of Medicine*, Vol. 172, pp. 240–243). According to the article, 200 AML deliveries had a mean cost of $2480 with a standard deviation of $766. At the time of the study, the average cost of having a baby in a U.S. hospital was $2528. At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, AML reduces the cost of having a baby in a U.S. hospital?

*In each of Exercises **9.107–9.110**, decide whether applying the t-test to perform a hypothesis test for the population mean in question appears reasonable. Explain your answers.*

9.107 Cardiovascular Hospitalizations. From the Florida State Center for Health Statistics report, *Women and Cardiovascular Disease Hospitalizations*, we found that, for cardiovascular hospitalizations, the mean age of women is 71.9 years. At one hospital, a random sample of 20 of its female cardiovascular patients had the following ages, in years.

75.9	83.7	87.3	74.5	82.5
78.2	76.1	52.8	56.4	53.8
88.2	78.9	81.7	54.4	52.7
58.9	97.6	65.8	86.4	72.4

9.108 Medieval Cremation Burials. In the article "Material Culture as Memory: Combs and Cremations in Early Medieval Britain" (*Early Medieval Europe*, Vol. 12, Issue 2, pp. 89–128), H. Williams discussed the frequency of cremation burials found in 17 archaeological sites in eastern England. Here are the data.

83	64	46	48	523	35	34	265	2484
46	385	21	86	429	51	258	119	

9.109 Capital Spending. An issue of *Brokerage Report* discussed the capital spending of telecommunications companies in the United States and Canada. The capital spending, in thousands of dollars, for each of 27 telecommunications companies is shown in the following table.

9,310	2,515	3,027	1,300	1,800	70	3,634
656	664	5,947	649	682	1,433	389
17,341	5,299	195	8,543	4,200	7,886	11,189
1,006	1,403	1,982	21	125	2,205	

9.110 Dating Artifacts. In the paper "Reassessment of TL Age Estimates of Burnt Flint from the Paleolithic Site of Tabun Cave, Israel" (*Journal of Human Evolution*, Vol. 45, Issue 5, pp. 401–409), N. Mercier and H. Valladas discussed the re-dating of artifacts and human remains found at Tabun Cave by using new methodological improvements. A random sample of 18 excavated pieces yielded the following new thermoluminescence (TL) ages.

195	243	215	282	361	222
237	266	244	251	282	290
276	248	357	301	224	191

Working with Large Data Sets

9.111 Stressed-Out Bus Drivers. Previous studies have shown that urban bus drivers have an extremely stressful job, and a large proportion of drivers retire prematurely with disabilities due to occupational stress. These stresses come from a combination of physical and social sources such as traffic congestion, incessant time pressure, and unruly passengers. In the paper, "Hassles on the Job: A Study of a Job Intervention With Urban Bus Drivers" (*Journal of Organizational Behavior*, Vol. 20, pp. 199–208), G. Evans et al. examined the effects of an intervention program to improve the conditions of urban bus drivers. Among other variables, the researchers monitored diastolic blood pressure of bus drivers in downtown Stockholm, Sweden. The data, in millimeters of mercury (mm Hg), on the WeissStats CD are based on the blood pressures obtained prior to intervention for the 41 bus drivers in the study. Use the technology of your choice to do the following.

a. Obtain a normal probability plot, boxplot, histogram, and stem-and-leaf diagram of the data.
b. Based on your results from part (a), can you reasonably apply the one-mean t-test to the data? Explain your reasoning.
c. At the 10% significance level, do the data provide sufficient evidence to conclude that the mean diastolic blood pressure of bus drivers in Stockholm exceeds the normal diastolic blood pressure of 80 mm Hg?

9.112 How Far People Drive. In 2005, the average car in the United States was driven 12.4 thousand miles, as reported by the Federal Highway Administration in *Highway Statistics*. On the WeissStats CD, we provide last year's distance driven, in thousands of miles, by each of 500 randomly selected cars. Use the technology of your choice to do the following.

a. Obtain a normal probability plot and histogram of the data.
b. Based on your results from part (a), can you reasonably apply the one-mean t-test to the data? Explain your reasoning.
c. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean distance driven last year differs from that in 2005?

9.113 Fair Market Rent. According to the document *Out of Reach*, published by the National Low Income Housing Coalition, the fair market rent (FMR) for a two-bedroom unit in Maine is \$779. A sample of 100 randomly selected two-bedroom units in Maine yielded the data on monthly rents, in dollars, given on the WeissStats CD. Use the technology of your choice to do the following.

a. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean monthly rent for two-bedroom units in Maine is greater than the FMR of \$779? Apply the one-mean t-test.
b. Remove the outlier from the data and repeat the hypothesis test in part (a).
c. Comment on the effect that removing the outlier has on the hypothesis test.
d. State your conclusion regarding the hypothesis test and explain your answer.

Extending the Concepts and Skills

9.114 Suppose that you want to perform a hypothesis test for a population mean based on a small sample but that preliminary data analyses indicate either the presence of outliers or that the variable under consideration is far from normally distributed.

a. Is either the z-test or t-test appropriate?
b. If not, what type of procedure might be appropriate?

9.115 Two-Tailed Hypothesis Tests and CIs. The following relationship holds between hypothesis tests and confidence intervals for one-mean t-procedures: For a two-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu \neq \mu_0$ if and only if μ_0 lies outside the $(1 - \alpha)$-level confidence interval for μ. In each case, illustrate the preceding relationship by obtaining the appropriate one-mean t-interval (Procedure 8.2 on page 346) and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.101 **b.** Exercise 9.104

9.116 Left-Tailed Hypothesis Tests and CIs. In Exercise 8.113 on page 353, we introduced one-sided one-mean t-intervals. The following relationship holds between hypothesis tests and confidence intervals for one-mean t-procedures: For a left-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu < \mu_0$ if and only if μ_0 is greater than the $(1 - \alpha)$-level upper confidence bound for μ. In each case, illustrate the preceding relationship by obtaining the appropriate upper confidence bound and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.105 **b.** Exercise 9.106

9.117 Right-Tailed Hypothesis Tests and CIs. In Exercise 8.113 on page 353, we introduced one-sided one-mean t-intervals. The following relationship holds between hypothesis tests and confidence intervals for one-mean t-procedures: For a right-tailed hypothesis test at the significance level α, the null hypothesis H_0: $\mu = \mu_0$ will be rejected in favor of the alternative hypothesis H_a: $\mu > \mu_0$ if and only if μ_0 is less than the $(1 - \alpha)$-level lower confidence bound for μ. In each case, illustrate the preceding relationship by obtaining the appropriate lower confidence bound and comparing the result to the conclusion of the hypothesis test in the specified exercise.

a. Exercise 9.102 (both parts)
b. Exercise 9.103

9.6 The Wilcoxon Signed-Rank Test*

Up to this point, we have presented two methods for performing a hypothesis test for a population mean. If the population standard deviation is known, we can use the z-test; if it is unknown, we can use the t-test.

Both procedures require another assumption for their use: The variable under consideration should be approximately normally distributed, or the sample size should be relatively large. For small samples, both procedures should be avoided in the presence of outliers.

In this section, we describe a third method for performing a hypothesis test for a population mean—the **Wilcoxon signed-rank test.**[†] This test, which is sometimes more appropriate than either the z-test or the t-test, is an example of a *nonparametric method.*

What Is a Nonparametric Method?

Recall that descriptive measures for population data, such as μ and σ, are called parameters. Technically, inferential methods concerned with parameters are called **parametric methods;** those that are not are called **nonparametric methods.** However, common statistical practice is to refer to most methods that can be applied without assuming normality as nonparametric. Thus the term *nonparametric method* as used in contemporary statistics is a misnomer.

Nonparametric methods have both advantages and disadvantages. On one hand, they usually entail fewer and simpler computations than parametric methods and are resistant to outliers and other extreme values. On the other hand, they are not as powerful as parametric methods, such as the z-test and t-test, when the requirements for use of parametric methods are met.[‡]

The Logic Behind the Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is based on the assumption that the variable under consideration has a *symmetric distribution*—one that can be divided into two pieces that are mirror images of each other—but does not require that its distribution be normal or have any other specific shape. Thus, for instance, the Wilcoxon signed-rank test applies to a variable that has a normal, triangular, uniform, or symmetric bimodal distribution but not to one that has a right-skewed or left-skewed distribution. The next example explains the reasoning behind this test.

EXAMPLE 9.18 Introducing the Wilcoxon Signed-Rank Test

Weekly Food Costs The U.S. Department of Agriculture publishes information about food costs in *Agricultural Research Service*. According to that document, a typical U.S. family of four spends about \$157 per week on food. Ten randomly selected Kansas families of four have the weekly food costs shown in Table 9.13. Do the data provide sufficient evidence to conclude that the mean weekly food cost for Kansas families of four is less than the national mean of \$157?

TABLE 9.13
Sample of weekly food costs (\$)

143	169	149	135	161
138	152	150	141	159

Solution Let μ denote the mean weekly food cost for all Kansas families of four. We want to perform the hypothesis test

$$H_0\text{: } \mu = \$157 \text{ (mean weekly food cost is not less than \$157)}$$
$$H_a\text{: } \mu < \$157 \text{ (mean weekly food cost is less than \$157).}$$

[†] The Wilcoxon signed-rank text is also known as the **one-sample Wilcoxon signed-rank test** and the **one-variable Wilcoxon signed-rank test.**

[‡] A precise definition of *power* is presented in Section 9.7.

FIGURE 9.21
Stem-and-leaf diagram of sample data in Table 9.13

13	5 8
14	1 3
14	9
15	0 2
15	9
16	1
16	9

As we said, a condition for the use of the Wilcoxon signed-rank test is that the variable under consideration have a symmetric distribution. If the weekly food costs for Kansas families of four have a symmetric distribution, a graphic of the sample data should be roughly symmetric.

Figure 9.21 shows a stem-and-leaf diagram of the sample data in Table 9.13. The diagram is roughly symmetric and so does not reveal any obvious violations of the symmetry condition.† We therefore apply the Wilcoxon signed-rank test to carry out the hypothesis test.

To begin, we rank the data in Table 9.13 according to distance and direction from the null hypothesis mean, $\mu_0 = \$157$. The steps for doing so are presented in Table 9.14.

TABLE 9.14
Steps for ranking the data in Table 9.13 according to distance and direction from the null hypothesis mean

Cost (\$) x	Difference $D = x - 157$	$\lvert D\rvert$	Rank of $\lvert D\rvert$	Signed rank R
143	−14	14	7	−7
138	−19	19	9	−9
169	12	12	6	6
152	−5	5	3	−3
149	−8	8	5	−5
150	−7	7	4	−4
135	−22	22	10	−10
141	−16	16	8	−8
161	4	4	2	2
159	2	2	1	1

Step 1 *Subtract μ_0 from x.*

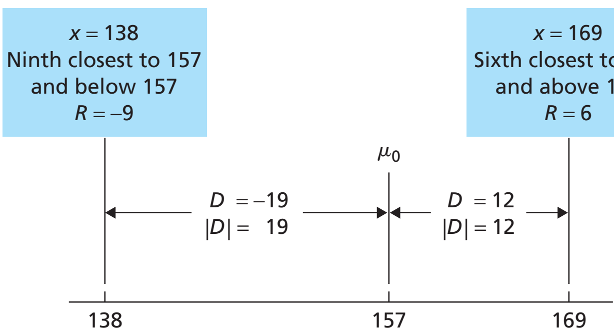
Step 2 *Make each difference positive by taking absolute values.*

Step 3 *Rank the absolute differences in order from smallest (1) to largest (10).*

Step 4 *Give each rank the same sign as the sign in the Difference column.*

The absolute differences, $|D|$, displayed in the third column, identify how far each observation is from 157. The ranks of those absolute differences, displayed in the fourth column, show which observations are closer to 157 and which are farther away. The signed ranks, R, displayed in the last column, indicate in addition whether an observation is greater than 157 (+) or less than 157 (−). Figure 9.22 depicts the information for the second and third rows of Table 9.14.

FIGURE 9.22
Meaning of signed ranks for the observations 138 and 169



† For ease in explaining the Wilcoxon signed-rank test, we have chosen an example in which the sample size is very small. This selection, however, makes it difficult to effectively check the symmetry condition. In general, we must proceed cautiously when dealing with very small samples.

The reasoning behind the Wilcoxon signed-rank test is as follows: If the null hypothesis, $\mu = \$157$, is true, then, because the distribution of weekly food costs is symmetric, we expect the sum of the positive ranks and the sum of the negative ranks to be roughly the same in magnitude. For the sample size of 10, the sum of all the ranks must be $1 + 2 + \cdots + 10 = 55$, and half of 55 is 27.5.

Thus, if the null hypothesis is true, we expect the sum of the positive ranks (and the sum of the negative ranks) to be roughly 27.5. If the sum of the positive ranks is too much smaller than 27.5, we conclude that the null hypothesis is false and, therefore, that the mean weekly food cost is less than \$157. From the last column of Table 9.14, the sum of the positive ranks, which we call W, equals $6 + 2 + 1 = 9$. This value is much smaller than 27.5 (the value we would expect if the mean is \$157).

The question now is, can the difference between the observed and expected values of W be reasonably attributed to sampling error, or does it indicate that the mean weekly food cost for Kansas families of four is actually less than \$157? We answer that question and complete the hypothesis test after we discuss some prerequisite material.

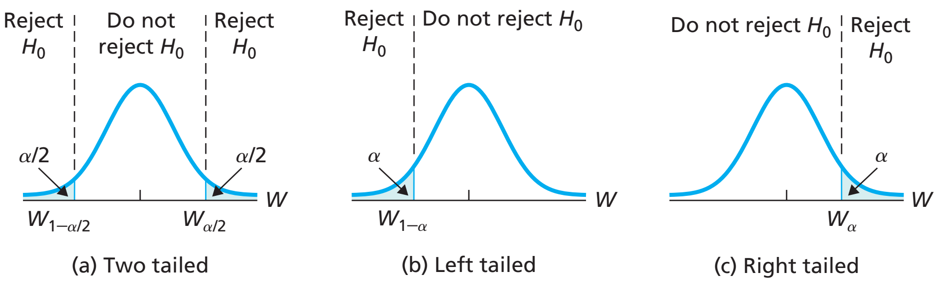
Using the Wilcoxon Signed-Rank Table†

Table V in Appendix A gives values of W_α for a Wilcoxon signed-rank test.‡ The two outside columns of Table V give the sample size, n. As expected, the symbol $\boldsymbol{W_\alpha}$ denotes the W-value with area (percentage, probability) α to its right. Thus the column headed $W_{0.10}$ contains W-values with area 0.10 to their right, the column headed $W_{0.05}$ contains W-values with area 0.05 to their right, and so on.

We can express the critical value(s) for a Wilcoxon signed-rank test at the significance level α as follows:

- For a two-tailed test, the critical values are the W-values with area $\alpha/2$ to its left (or, equivalently, area $1 - \alpha/2$ to its right) and area $\alpha/2$ to its right, which are $W_{1-\alpha/2}$ and $W_{\alpha/2}$, respectively. See Fig. 9.23(a).
- For a left-tailed test, the critical value is the W-value with area α to its left or, equivalently, area $1 - \alpha$ to its right, which is $W_{1-\alpha}$. See Fig. 9.23(b).
- For a right-tailed test, the critical value is the W-value with area α to its right, which is W_α. See Fig. 9.23(c).

FIGURE 9.23
Critical value(s) for a Wilcoxon signed-rank test at the significance level α if the test is (a) two tailed, (b) left tailed, or (c) right tailed



Note the following:

- A critical value from Table V is to be included as part of the rejection region.
- Although the variable W is discrete, we drew the "histograms" in Fig. 9.23 in the shape of a normal curve. This approach is not only convenient, it is also acceptable because W is close to normally distributed except for very small sample sizes. We use this graphical convention throughout this section.

† We can use the Wilcoxon signed-rank table to estimate the P-value of a Wilcoxon signed-rank test. Because doing so can be awkward or tedious, however, using statistical software is preferable. Thus, those concentrating on the P-value approach to hypothesis testing can skip to the subsection "Performing the Wilcoxon Signed-Rank Test."

‡ Actually, the α-levels in Table V are only approximate but are used in practice.

The distribution of the variable W is symmetric about $n(n+1)/4$. This characteristic implies that the W-value with area A to its left (or, equivalently, area $1-A$ to its right) equals $n(n+1)/2$ minus the W-value with area A to its right. In symbols,

$$W_{1-A} = n(n+1)/2 - W_A. \qquad (9.1)$$

Referring to Fig. 9.23, we see that by using Equation (9.1) and Table V, we can determine the critical value for a left-tailed Wilcoxon signed-rank test and the critical values for a two-tailed Wilcoxon signed-rank test. The next example illustrates the use of Table V to determine critical values for a Wilcoxon signed-rank test.

EXAMPLE 9.19 Using the Wilcoxon Signed-Rank Table

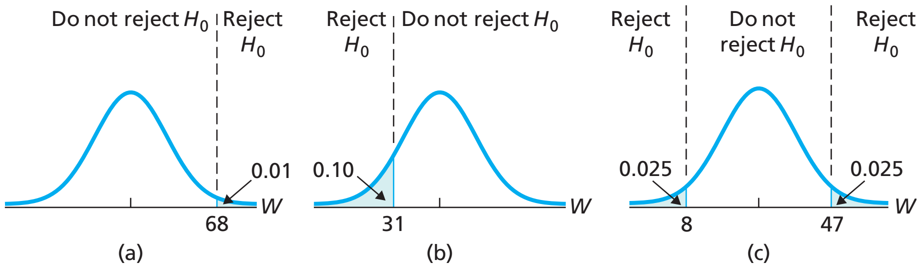
In each case, use Table V to determine the critical value(s) for a Wilcoxon signed-rank test. Sketch graphs to illustrate your results.

a. Sample size = 12; significance level = 0.01; right tailed
b. Sample size = 14; significance level = 0.10; left tailed
c. Sample size = 10; significance level = 0.05; two tailed

Solution In solving these problems, it helps to refer to Fig. 9.23.

a. The critical value for a right-tailed test at the 1% significance level is $W_{0.01}$. To find the critical value, we use Table V. First we go down the outside columns, labeled n, to "12." Then, going across that row to the column labeled $W_{0.01}$, we reach 68, the required critical value. See Fig. 9.24(a).

FIGURE 9.24
Critical value(s) for a Wilcoxon signed-rank test: (a) right tailed, $\alpha = 0.01$, $n = 12$; (b) left tailed, $\alpha = 0.10$, $n = 14$; (c) two tailed, $\alpha = 0.05$, $n = 10$



b. The critical value for a left-tailed test at the 10% significance level is $W_{1-0.10}$. To find the critical value, we use Table V and Equation (9.1). First we go down the outside columns, labeled n, to "14." Then, going across that row to the column labeled $W_{0.10}$, we reach 74; thus $W_{0.10} = 74$. Now we apply Equation (9.1) and the result just obtained to get

$$W_{1-0.10} = 14(14+1)/2 - W_{0.10} = 105 - 74 = 31,$$

which is the required critical value. See Fig. 9.24(b).

c. The critical values for a two-tailed test at the 5% significance level are $W_{1-0.05/2}$ and $W_{0.05/2}$, that is, $W_{1-0.025}$ and $W_{0.025}$. First we use Table V to find $W_{0.025}$. We go down the outside columns, labeled n, to "10." Then, going across that row to the column labeled $W_{0.025}$, we reach 47; thus $W_{0.025} = 47$. Now we apply Equation (9.1) and the result just obtained to get $W_{1-0.025}$:

$$W_{1-0.025} = 10(10+1)/2 - W_{0.025} = 55 - 47 = 8.$$

See Fig. 9.24(c).

You try it!

Exercise 9.125 on page 410

Performing the Wilcoxon Signed-Rank Test

Procedure 9.3 on the next page provides a step-by-step method for performing a Wilcoxon signed-rank test by using either the critical-value approach or the P-value approach. Note that we often use the phrase **symmetric population** to indicate that the variable under consideration has a symmetric distribution.

PROCEDURE 9.3 Wilcoxon Signed-Rank Test

Purpose To perform a hypothesis test for a population mean, μ

Assumptions

1. Simple random sample
2. Symmetric population

Step 1 **The null hypothesis is H_0: $\mu = \mu_0$, and the alternative hypothesis is**

$$H_a: \mu \neq \mu_0 \text{ (Two tailed)} \quad \text{or} \quad H_a: \mu < \mu_0 \text{ (Left tailed)} \quad \text{or} \quad H_a: \mu > \mu_0 \text{ (Right tailed)}$$

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

$$W = \text{sum of the positive ranks}$$

and denote that value W_0. To do so, construct a work table of the following form.

Observation x	Difference $D = x - \mu_0$	$\lvert D\rvert$	Rank of $\lvert D\rvert$	Signed rank R
⋮	⋮	⋮	⋮	⋮

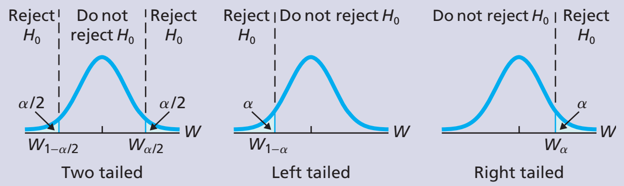
CRITICAL-VALUE APPROACH OR ***P*-VALUE APPROACH**

CRITICAL-VALUE APPROACH

Step 4 **The critical value(s) are**

$$W_{1-\alpha/2} \text{ and } W_{\alpha/2} \text{ (Two tailed)} \quad \text{or} \quad W_{1-\alpha} \text{ (Left tailed)} \quad \text{or} \quad W_{\alpha} \text{ (Right tailed)}$$

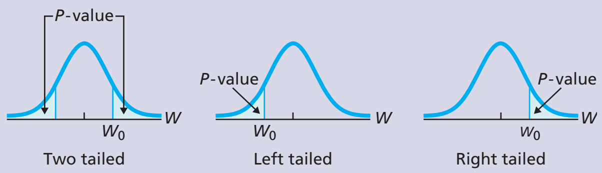
Use Table V to find the critical value(s). For a left-tailed or two-tailed test, you will also need the relation $W_{1-A} = n(n+1)/2 - W_A$.



Step 5 **If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.**

P-VALUE APPROACH

Step 4 **Obtain the *P*-value by using technology.**



Step 5 **If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.**

Step 6 **Interpret the results of the hypothesis test.**

EXAMPLE 9.20 The Wilcoxon Signed-Rank Test

Weekly Food Costs Let's complete the hypothesis test of Example 9.18. A random sample of 10 Kansas families of four yielded the data on weekly food costs shown in Table 9.13 on page 400. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean weekly food cost for Kansas families of four is less than the national mean of \$157?

Solution We apply Procedure 9.3.

Step 1 State the null and alternative hypotheses.

Let μ denote the mean weekly food cost for all Kansas families of four. Then the null and alternative hypotheses are, respectively,

$$H_0\text{: } \mu = \$157 \text{ (mean weekly food cost is not less than \$157)}$$
$$H_a\text{: } \mu < \$157 \text{ (mean weekly food cost is less than \$157).}$$

Note that the hypothesis test is left tailed.

Step 2 Decide on the significance level, α.

The test is to be performed at the 5% significance level, or $\alpha = 0.05$.

Step 3 Compute the value of the test statistic

$$W = \textbf{sum of the positive ranks.}$$

The last column of Table 9.14 on page 401 shows that the sum of the positive ranks equals

$$W = 6 + 2 + 1 = 9.$$

CRITICAL-VALUE APPROACH OR ***P*-VALUE APPROACH**

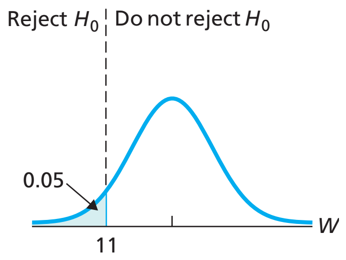
Step 4 The critical value for a left-tailed test is $W_{1-\alpha}$. Use Table V and the relation $W_{1-A} = n(n+1)/2 - W_A$ to find the critical value.

From Table 9.13 on page 400, we see that the sample size is 10. The critical value for a left-tailed test at the 5% significance level is $W_{1-0.05}$. To find the critical value, first we go down the outside columns of Table V, labeled n, to "10." Then, going across that row to the column labeled $W_{0.05}$, we reach 44; thus $W_{0.05} = 44$. Now we apply the aforementioned relation and the result just obtained to get

$$W_{1-0.05} = 10(10+1)/2 - W_{0.05} = 55 - 44 = 11,$$

which is the required critical value. See Fig. 9.25A.

FIGURE 9.25A



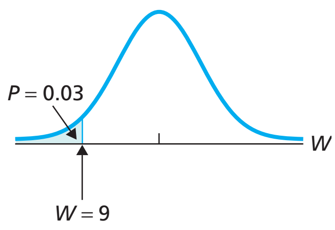
Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

The value of the test statistic is $W = 9$, as found in Step 3, which falls in the rejection region shown in Fig. 9.25A. Thus we reject H_0. The test results are statistically significant at the 5% level.

Step 4 Obtain the *P*-value by using technology.

Using technology, we find that the P-value for the hypothesis test is $P = 0.03$, as shown in Fig. 9.25B.

FIGURE 9.25B



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P = 0.03$. Because the P-value is less than the specified significance level of 0.05, we reject H_0. The test results are statistically significant at the 5% level and (see Table 9.8 on page 378) provide strong evidence against the null hypothesis.

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that the mean weekly food cost for Kansas families of four is less than the national mean of $157.

SC
Report 9.3

As mentioned earlier, one advantage of nonparametric methods is that they are resistant to outliers. We can illustrate that advantage for the Wilcoxon signed-rank test by referring to Example 9.20.

The stem-and-leaf diagram depicted in Fig. 9.21 on page 401 shows that the sample data presented in Table 9.13 contain no outliers. The smallest observation, and also the farthest from the null hypothesis mean of 157, is 135. Replacing 135 by, say, 85, introduces an outlier but has no effect on the value of the test statistic and hence none on the hypothesis test itself. (Why is that so?)

Note: The following points may be relevant when performing a Wilcoxon signed-rank test:

- If an observation equals μ_0 (the value for the mean in the null hypothesis), that observation should be removed and the sample size reduced by 1.
- If two or more absolute differences are tied, each should be assigned the mean of the ranks they would have had if there were no ties.

To illustrate the second bulleted item, suppose that two absolute differences are tied for second place. Then each should be assigned rank $(2 + 3)/2 = 2.5$, and rank 4 should be assigned to the next-largest absolute difference, which really is fourth. Similarly, if three absolute differences are tied for fifth place, each should be assigned rank $(5 + 6 + 7)/3 = 6$, and rank 8 should be assigned to the next-largest absolute difference.

In Example 9.16, we used the one-mean t-test to decide whether, on average, high mountain lakes in the Southern Alps are nonacidic. Now we do so by using the Wilcoxon signed-rank test.

EXAMPLE 9.21 The Wilcoxon Signed-Rank Test

Acid Rain and Lake Acidity A lake is classified as nonacidic if it has a pH greater than 6. A. Marchetto and A. Lami measured the pH of high mountain lakes in the Southern Alps and reported their findings in the paper "Reconstruction of pH by Chrysophycean Scales in Some Lakes of the Southern Alps" (*Hydrobiologia*, Vol. 274, pp 83–90). Table 9.12, which we repeat here as Table 9.15, shows the pH levels obtained by the researchers for 15 lakes.

TABLE 9.15
pH levels for 15 lakes

7.2	7.3	6.1	6.9	6.6
7.3	6.3	5.5	6.3	6.5
5.7	6.9	6.7	7.9	5.8

At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic? Use the Wilcoxon signed-rank test.

Solution Figure 9.26 shows a stem-and-leaf diagram of the sample data in Table 9.15. The diagram is relatively symmetric. Hence, we can reasonably apply Procedure 9.3 to carry out the required hypothesis test.

FIGURE 9.26
Stem-and-leaf diagram of pH levels in Table 9.15

```
5 | 5 7 8
6 | 1 3 3
6 | 5 6 7 9 9
7 | 2 3 3
7 | 9
```

Step 1 State the null and alternative hypotheses.

Let μ denote the mean pH level of all high mountain lakes in the Southern Alps. Then the null and alternative hypotheses are, respectively,

$$H_0\text{: } \mu = 6 \text{ (on average, the lakes are acidic)}$$
$$H_a\text{: } \mu > 6 \text{ (on average, the lakes are nonacidic).}$$

Note that the hypothesis test is right tailed.

Step 2 Decide on the significance level, α.

We are to perform the test at the 5% significance level, so $\alpha = 0.05$.

Step 3 Compute the value of the test statistic

$$W = \text{sum of the positive ranks.}$$

To do so, first construct a worktable to obtain the signed ranks.

We construct the following work table. Note that, in several instances, we applied the aforementioned method to deal with tied absolute differences.

pH x	Difference $D = x - 6$	$\lvert D \rvert$	Rank of $\lvert D \rvert$	Signed rank R
7.2	1.2	1.2	12	12
7.3	1.3	1.3	13.5	13.5
6.1	0.1	0.1	1	1
6.9	0.9	0.9	10.5	10.5
6.6	0.6	0.6	8	8
7.3	1.3	1.3	13.5	13.5
6.3	0.3	0.3	4	4
5.5	−0.5	0.5	6.5	−6.5
6.3	0.3	0.3	4	4
6.5	0.5	0.5	6.5	6.5
5.7	−0.3	0.3	4	−4
6.9	0.9	0.9	10.5	10.5
6.7	0.7	0.7	9	9
7.9	1.9	1.9	15	15
5.8	−0.2	0.2	2	−2

Referring to the last column of the work table, we find that the value of the test statistic is

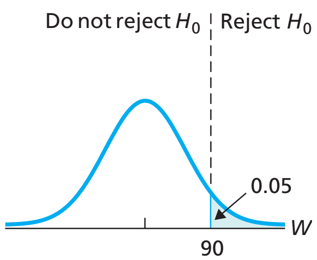
$$W = 12 + 13.5 + 1 + \cdots + 9 + 15 = 107.5.$$

CRITICAL-VALUE APPROACH OR *P*-VALUE APPROACH

Step 4 The critical value for a right-tailed test is W_α. Use Table V to find the critical value.

From Table 9.15, we see that the sample size is 15. The critical value for a right-tailed test at the 5% significance level is $W_{0.05}$. To find the critical value, first we go down the outside columns of Table V, labeled n, to "15." Then, going across that row to the column labeled $W_{0.05}$, we reach 90, the required critical value. See Fig. 9.27A.

FIGURE 9.27A



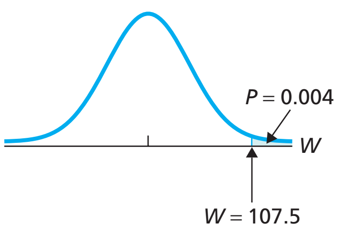
Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

The value of the test statistic is $W = 107.5$, as found in Step 3, which falls in the rejection region shown in Fig. 9.27A. Thus we reject H_0. The test results are statistically significant at the 5% level.

Step 4 Obtain the P-value by using technology.

Using technology, we find that the P-value for the hypothesis test is $P = 0.004$, as shown in Fig. 9.27B.

FIGURE 9.27B



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

From Step 4, $P = 0.004$. Because the P-value is less than the specified significance level of 0.05, we reject H_0. The test results are statistically significant at the 5% level and (see Table 9.8 on page 378) provide very strong evidence against the null hypothesis.

You try it!

Exercise 9.135 on page 411

Step 6 Interpret the results of the hypothesis test.

Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that, on average, high mountain lakes in the Southern Alps are nonacidic.

We note that both the one-mean t-test of Example 9.16 and the Wilcoxon signed-rank test of Example 9.21 reject the null hypothesis that high mountain lakes in the Southern Alps are, on average, acidic in favor of the alternative hypothesis that they are, on average, nonacidic. Furthermore, with both tests, the data provide very strong evidence against that null hypothesis (and, hence, in favor of the alternative hypothesis). Indeed, as we have seen, $P = 0.002$ for the one-mean t-test, and $P = 0.004$ for the Wilcoxon signed-rank test.

Comparing the Wilcoxon Signed-Rank Test and the *t*-Test

As you learned in Section 9.5, a t-test can be used to conduct a hypothesis test for a population mean when the variable under consideration is normally distributed. Because normally distributed variables have symmetric distributions, we can also use the Wilcoxon signed-rank test to perform such a hypothesis test.

For a normally distributed variable, the t-test is more powerful than the Wilcoxon signed-rank test because it is designed expressly for such variables; surprisingly, though, the t-test is not much more powerful than the Wilcoxon signed-rank test. However, if the variable under consideration has a symmetric distribution but is not normally distributed, the Wilcoxon signed-rank test is usually more powerful than the t-test and is often considerably more powerful.

KEY FACT 9.8

Wilcoxon Signed-Rank Test Versus the *t*-Test

Suppose that you want to perform a hypothesis test for a population mean. When deciding between the t-test and the Wilcoxon signed-rank test, follow these guidelines:

- If you are reasonably sure that the variable under consideration is normally distributed, use the t-test.
- If you are not reasonably sure that the variable under consideration is normally distributed but are reasonably sure that it has a symmetric distribution, use the Wilcoxon signed-rank test.

Testing a Population Median with the Wilcoxon Signed-Rank Procedure

Because the mean and median of a symmetric distribution are identical, a Wilcoxon signed-rank test can be used to perform a hypothesis test for a population median, η, as well as for a population mean, μ. To use Procedure 9.3 to carry out a hypothesis test for a population median, simply replace μ by η and μ_0 by η_0.

In some of the exercises at the end of this section, you will be asked to use the Wilcoxon signed-rank test to perform hypothesis tests for a population median.

THE TECHNOLOGY CENTER

Some statistical technologies have programs that automatically perform a Wilcoxon signed-rank test. In this subsection, we present output and step-by-step instructions for such programs. (*Note to TI-83/84 Plus users:* At the time of this writing, the TI-83/84 Plus does not have a built-in program for conducting a Wilcoxon signed-rank test. However, a TI program, WILCOX, to help with the calculations is located in the TI Programs folder on the WeissStats CD. See the *TI-83/84 Plus Manual* for details.)

As we said earlier, a Wilcoxon signed-rank test can be used to perform a hypothesis test for a population median, η, as well as for a population mean, μ. Many statistical technologies present the output of that procedure in terms of the median, but that output can also be interpreted in terms of the mean.

EXAMPLE 9.22 Using Technology to Conduct a Wilcoxon Signed-Rank Test

Weekly Food Costs Table 9.13 on page 400 gives the weekly food costs for 10 Kansas families of four. Use Minitab or Excel to decide, at the 5% significance level, whether the data provide sufficient evidence to conclude that the mean weekly food cost for Kansas families of four is less than the national mean of \$157.

Solution Let μ denote the mean weekly food cost for all Kansas families of four. We want to perform the hypothesis test

H_0: $\mu = \$157$ (mean weekly food cost is not less than \$157)
H_a: $\mu < \$157$ (mean weekly food cost is less than \$157)

at the 5% significance level. Note that the hypothesis test is left tailed.

We applied the Wilcoxon signed-rank test programs to the data, resulting in Output 9.3. Steps for generating that output are presented in Instructions 9.3.

OUTPUT 9.3
Wilcoxon signed-rank test output on the sample of weekly food costs

MINITAB

Wilcoxon Signed Rank Test: COST

```
Test of median = 157.0 versus median < 157.0

               N
              for   Wilcoxon             Estimated
         N   Test   Statistic        P      Median
COST    10     10         9.0    0.033       149.5
```

EXCEL

```
COST One Sample Wilcoxon Signed Rank Test
Test Summary
             Ho:                   Median =  157
             Ha:      Lower tail:  Median <  157
          Count                               10
 Count Adjusted                               10
              W                                9
    Z Statistic                                •
       P-value:                           0.0322
     Conclusion      Reject Ho at alpha = 0.05
```

As shown in Output 9.3, the *P*-value for the hypothesis test is 0.03. Because the *P*-value is less than the specified significance level of 0.05, we reject H_0. At the 5% significance level, the data provide sufficient evidence to conclude that the mean weekly food cost for Kansas families of four is less than the national mean of \$157.

INSTRUCTIONS 10.3
Steps for generating Output 10.3

MINITAB

1 Store the data from Table 9.13 in a column named COST
2 Choose **Stat ➤ Nonparametrics ➤ 1-Sample Wilcoxon...**
3 Specify COST in the **Variables** text box
4 Select the **Test median** option button
5 Click in the **Test median** text box and type 157
6 Click the arrow button at the right of the **Alternative** drop-down list box and select **less than**
7 Click **OK**

EXCEL

1 Store the data from Table 9.13 in a range named COST
2 Choose **DDXL ➤ Nonparametric Tests**
3 Select **1 Var Wilcoxon** from the **Function type** drop-down box
4 Specify COST in the **Quantitative Variable** text box
5 Click **OK**
6 Click the **Set Hypothesized Median** button
7 Click in the **Set Hypothesized Median** text box and type 157
8 Click **OK**
9 Click the **0.05** button
10 Click the **Left Tailed** button
11 Click the **Compute** button

Exercises 9.6

Understanding the Concepts and Skills

9.118 Technically, what is a *nonparametric method?* In current statistical practice, how is that term used?

9.119 Discuss advantages and disadvantages of nonparametric methods relative to parametric methods.

9.120 What distributional assumption must be met in order to use the Wilcoxon signed-rank test?

9.121 We mentioned that if, in a Wilcoxon signed-rank test, an observation equals μ_0 (the value given for the mean in the null hypothesis), that observation should be removed and the sample size reduced by 1. Why does that need to be done?

9.122 Suppose that you want to perform a hypothesis test for a population mean. Assume that the population standard deviation is unknown and that the sample size is relatively small. In each part, we have given the distribution shape of the variable under consideration. Decide whether you would use the t-test, the Wilcoxon signed-rank test, or neither.

a. Uniform **b.** Normal
c. Reverse J shaped

9.123 Suppose that you want to perform a hypothesis test for a population mean. Assume that the population standard deviation is unknown and that the sample size is relatively small. In each part, we have given the distribution shape of the variable under consideration. Decide whether you would use the t-test, the Wilcoxon signed-rank test, or neither.

a. Triangular **b.** Symmetric bimodal
c. Left skewed

9.124 The Wilcoxon signed-rank test can be used to perform a hypothesis test for a population median, η, as well as for a population mean, μ. Why is that so?

*Exercises **9.125–9.128** pertain to critical values for a Wilcoxon signed-rank test. Use Table V in Appendix A to determine the critical value(s) in each case. For a left-tailed or two-tailed test, you will also need the relation $W_{1-A} = n(n+1)/2 - W_A$.*

9.125 Sample size = 8; Significance level = 0.05
a. Right tailed **b.** Left tailed **c.** Two tailed

9.126 Sample size = 10; Significance level = 0.01
a. Right tailed **b.** Left tailed **c.** Two tailed

9.127 Sample size = 19; Significance level = 0.10
a. Right tailed **b.** Left tailed **c.** Two tailed

9.128 Sample size = 15; Significance level = 0.05
a. Right tailed **b.** Left tailed **c.** Two tailed

*In each of Exercises **9.129–9.134**, we have provided a null hypothesis and alternative hypothesis and a sample from the population under consideration. In each case, use the Wilcoxon signed-rank test to perform the required hypothesis test at the 10% significance level.*

9.129 H_0: $\mu = 5$, H_a: $\mu > 5$

12	7	11	9	3	2	8	6

9.130 H_0: $\mu = 10$, H_a: $\mu < 10$

7	6	5	12	15	14	13	4

9.131 H_0: $\mu = 6$, H_a: $\mu \neq 6$

6	4	8	4	1	1	4	7

9.132 H_0: $\mu = 3$, H_a: $\mu \neq 3$

6	6	3	3	2	5	4	7	4

9.133 H_0: $\mu = 12$, H_a: $\mu < 12$

16	11	10	14	13	15	5	8	11

9.134 H_0: $\mu = 8$, H_a: $\mu > 8$

8	10	11	11	5	9	9	12

*In each of Exercises **9.135–9.140**, use the Wilcoxon signed-rank test to perform the required hypothesis test.*

9.135 Global Warming? During the late 1800s, Lake Wingra in Madison, Wisconsin, was frozen over an average of 124.9 days per year. A random sample of eight recent years provided the following data on numbers of days that the lake was frozen over.

103	80	79	135	134	77	80	111

At the 5% significance level, do the data provide sufficient evidence to conclude that the average number of ice days is less now than in the late 1800s?

9.136 Happy-Life Years. In the article, "Apparent Quality-of-Life in Nations: How Long and Happy People Live" (*Social Indicators Research*, Vol. 71, pp. 61–86) R. Veenhoven discussed how the quality of life in nations can be measured by how long and happy people live. In the 1990s, the median number of happy-life years across nations was 46.7. A random sample of eight nations for this year provided the following data on number of happy-life years.

30.3	47.0	56.4	30.5	39.6	47.9	29.7	52.5

At the 5% significance level, do the data provide sufficient evidence to conclude that the median number of happy-life years has changed from that in the 1990s?

9.137 How Old People Are. In 2007, the median age of U.S. residents was 36.6 years, as reported by the U.S. Census Bureau in *Current Population Reports*. A random sample of 10 U.S. residents taken this year yielded the following ages, in years.

43	63	15	58	37
46	50	40	12	27

At the 1% significance level, do the data provide sufficient evidence to conclude that the median age of today's U.S. residents has increased from the 2007 median age of 36.6 years?

9.138 Beverage Expenditures. The Bureau of Labor Statistics publishes information on average annual expenditures by consumers in *Consumer Expenditures*. In 2007, the mean amount spent per consumer unit on nonalcoholic beverages was $333. A random sample of 12 consumer units yielded the following data, in dollars, on last year's expenditures on nonalcoholic beverages.

474	289	297	378	372	394
353	386	372	362	307	371

At the 5% significance level, do the data provide sufficient evidence to conclude that last year's mean amount spent by consumers on nonalcoholic beverages has increased from the 2007 mean of $333?

9.139 Pricing Mustangs. The *Kelley Blue Book* provides information on retail and trade-in values for used cars and trucks. The retail value represents the price a dealer might charge after preparing the vehicle for sale. A 2006 Ford Mustang coupe has a 2009 *Kelley Blue Book* retail value of $13,015. We obtained the following asking prices, in dollars, for a sample of 2006 Ford Mustang coupes for sale in Phoenix, Arizona.

13,645	13,157	13,153	12,965	12,764
12,664	11,665	10,565	12,665	12,765

At the 10% significance level, do the data provide sufficient evidence to conclude that the mean asking price for 2006 Ford Mustang coupes in Phoenix is less than the 2009 *Kelley Blue Book* retail value?

9.140 Birth Weights. The National Center for Health Statistics reports in *Vital Statistics of the United States* that the median birth weight of U.S. babies was 7.4 lb in 2002. A random sample of this year's births provided the following weights, in pounds.

8.6	7.4	5.3	13.8	7.8	5.7	9.2
8.8	8.2	9.2	5.6	6.0	11.6	7.2

Can we conclude that this year's median birth weight differs from that in 2002? Use a significance level of 0.05.

9.141 Brewery Effluent and Crops. Two researchers, M. Ajmal and A. Khan, reported their findings on experiments with brewery wastes used for agricultural purposes in the article "Effects of Brewery Effluent on Agricultural Soil and Crop Plants" (*Environmental Pollution (Series A)*, 33, pp. 341–351). The following data, based on the results of the study, provide the percentages of limestone in the soil obtained by using 100% Mohan Meakin Breweries Ltd. (MMBL) effluent.

2.41	2.31	2.54	2.28	2.72
2.60	2.51	2.51	2.42	2.70

a. Can you conclude that the mean available limestone in soil treated with 100% MMBL effluent exceeds 2.30%, the percentage ordinarily found? Perform a Wilcoxon signed-rank test at the 1% significance level.

b. The hypothesis test considered in part (a) was done in Exercise 9.103 with a t-test. The assumption in that exercise is that the percentage of limestone in the soil obtained by using 100% effluent is normally distributed. If that is the case, why is it permissible to perform a Wilcoxon signed-rank test for the mean available limestone in soil treated with 100% MMBL effluent?

9.142 Ethical Food Choice Motives. In the paper "Measurement of Ethical Food Choice Motives" (*Appetite*, Vol. 34, pp. 55–59), research psychologists M. Lindeman and M. Väänänen of the University of Helsinki published a study on the factors that most influence peoples' choice of food. One of the questions asked of the participants was how important, on a scale of 1 to 4 (1 = not at all important, 4 = very important), is ecological welfare in food choice motive, where ecological welfare includes animal welfare and environmental protection. Following are the responses of a random sample of 18 Helsinkians.

3	2	2	3	3	3	2	2	3
3	3	1	3	4	2	1	3	1

At the 5% significance level, do the data provide sufficient evidence to conclude that, on average, Helsinkians respond with an ecological welfare food choice motive greater than 2?

a. Use the Wilcoxon signed-rank test.
b. Use the t-test.
c. Compare the results of the two tests.

9.143 Checking Advertised Contents. A manufacturer of liquid soap produces a bottle with an advertised content of 310 milliliters (mL). Sixteen bottles are randomly selected and found to have the following contents, in mL.

297	318	306	300	311	303	291	298
322	307	312	300	315	296	309	311

A normal probability plot of the data indicates that you can assume the contents are normally distributed. Let μ denote the mean content of all bottles produced. To decide whether the mean content is less than advertised, perform the hypothesis test

$$H_0: \mu = 310 \text{ mL}$$
$$H_a: \mu < 310 \text{ mL}$$

at the 5% significance level.

a. Use the t-test.
b. Use the Wilcoxon signed-rank test.
c. If the mean content is in fact less than 310 mL, how do you explain the discrepancy between the two tests?

9.144 Education of Jail Inmates. Thirty years ago, the Bureau of Justice Statistics reported in *Profile of Jail Inmates* that the median educational attainment of jail inmates was 10.2 years. Ten current inmates are randomly selected and found to have the following educational attainments, in years.

14	10	5	6	8
10	10	8	9	9

Assume that educational attainments of current jail inmates have a symmetric, nonnormal distribution. At the 10% significance level, do the data provide sufficient evidence to conclude that this year's median educational attainment has changed from what it was 30 years ago?

a. Use the t-test.
b. Use the Wilcoxon signed-rank test.
c. If this year's median educational attainment has in fact changed from what it was 30 years ago, how do you explain the discrepancy between the two tests?

Working with Large Data Sets

9.145 Delaying Adulthood. The convict surgeonfish is a common tropical reef fish that has been found to delay metamorphosis into adult by extending its larval phase. This delay often leads to enhanced survivorship in the species by increasing the chances of finding suitable habitat. In the paper "Delayed Metamorphosis of a Tropical Reef Fish (*Acanthurus triostegus*): A Field Experiment" (*Marine Ecology Progress Series*, Vol. 176, pp. 25–38), M. McCormick published data that he obtained on the larval duration, in days, of 90 convict surgeonfish. The data are given on the WeissStats CD. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean larval duration of convict surgeonfish exceeds 52 days?

a. Employ the Wilcoxon signed-rank test.
b. Employ the t-test.
c. Compare your results from parts (a) and (b).

9.146 Easy Hole at the British Open? The Old Course at St. Andrews in Scotland is home of the British Open, one of the major tournaments in professional golf. The *Hole O'Cross Out*, known by both European and American professional golfers as one of the friendliest holes at St. Andrews, is the fifth hole, a 514-yard, par 5 hole with an open fairway and a large green. As one reporter for pgatour.com put it, "If players think before they drive, they will easily walk away with birdies and pars." The scores on the *Hole O'Cross Out* posted by a sample of 156 golf professionals are presented on the WeissStats CD. Use those data and the technology of your choice to decide whether, on average, professional golfers score better than par on the *Hole O'Cross Out*. Perform the required hypothesis test at the 0.01 level of significance.

a. Employ the Wilcoxon signed-rank test.
b. Employ the t-test.
c. Compare your results from parts (a) and (b).

*In Exercises **9.147–9.149**, we have repeated the contexts of Exercises 9.111–9.113 from Section 9.5. For each exercise, use the technology of your choice to do the following.*

a. Apply the Wilcoxon signed-rank test to perform the required hypothesis test.

b. Compare your result in part (a) to that obtained in the corresponding exercise in Section 9.5, where the t-test was used.

9.147 Stressed-Out Bus Drivers. In the paper "Hassles on the Job: A Study of a Job Intervention With Urban Bus Drivers" (*Journal of Organizational Behavior*, Vol. 20, pp. 199–208), G. Evans et al. examined the effects of an intervention program to improve the conditions of urban bus drivers. Among other variables, the researchers monitored diastolic blood pressure of bus drivers in downtown Stockholm, Sweden. The data, in millimeters of mercury (mm Hg), on the WeissStats CD are based on the blood pressures obtained prior to intervention for the 41 bus drivers in the study. At the 10% significance level, do the data provide sufficient evidence to conclude that the mean diastolic blood pressure of bus drivers in Stockholm exceeds the normal diastolic blood pressure of 80 mm Hg?

9.148 How Far People Drive. In 2005, the average car in the United States was driven 12.4 thousand miles, as reported by the Federal Highway Administration in *Highway Statistics*. On the WeissStats CD, we provide last year's distance driven, in thousands of miles, by each of 500 randomly selected cars. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean distance driven last year differs from that in 2005?

9.149 Fair Market Rent. According to the document *Out of Reach*, published by the National Low Income Housing Coalition, the fair market rent (FMR) for a two-bedroom unit in Maine is \$779. A sample of 100 randomly selected two-bedroom units in Maine yielded the data on monthly rents, in dollars, given on the WeissStats CD. At the 5% significance level, do the data provide sufficient evidence to conclude that the mean monthly rent for two-bedroom units in Maine is greater than the FMR of \$779? Perform the required hypothesis test both with and without the outlier.

Extending the Concepts and Skills

9.150 How Long Do Marriages Last? According to the document "Number, Timing, and Duration of Marriages and Divorces" (*Household Economic Studies*, P70–97) by R. Kreider, the median duration of a first marriage that ended in divorce in 2001 was 8.0 years. Suppose that you take a simple random sample of 50 divorce certificates for first marriages from last year and record the marriage durations. You want to use these data to decide whether the median duration of a first marriage that ended in divorce last year has increased from that in 2001. Which procedure would give the better results, the Wilcoxon signed-rank test or the t-test? Explain your answer.

9.151 The Census Form. The U.S. Census Bureau estimates that the *U.S. Census Form* takes the average household 14 minutes to complete. To check that claim, completion times are recorded for 36 randomly selected households. Which test would give the better results, the Wilcoxon signed-rank test or the t-test? Explain your answer.

9.152 Waiting for the Train. A commuter train arrives punctually at a station every half hour. Each morning, a commuter named John leaves his house and casually strolls to the train station. John thinks that he is unlucky and that he waits longer for the train on average than he should.

a. Assuming that John is not unlucky, how long should he expect to wait for the train, on average?

b. Assuming that John is not unlucky, identify the distribution of the times he waits for the train.

c. The following is a sample of the times, in minutes, that John waited for the train.

24	20	3	19	28	22
26	4	11	5	16	24

Use the Wilcoxon signed-rank test to decide, at the 10% significance level, whether the data provide sufficient evidence to conclude that, on average, John waits more than 15 minutes for the train.

d. Explain why the Wilcoxon signed-rank test is appropriate here.

e. Is the Wilcoxon signed-rank test more appropriate here than the t-test? Explain your answer.

Normal Approximation for *W*. The Wilcoxon signed-rank table, Table V, stops at $n = 20$. For larger samples, a normal approximation can be used. In fact, the normal approximation works well even for sample sizes as small as 10.

Normal Approximation for *W*

Suppose that the variable under consideration has a symmetric distribution. Then, for samples of size n,

- $\mu_W = n(n+1)/4$,
- $\sigma_W = \sqrt{n(n+1)(2n+1)/24}$, and
- W is approximately normally distributed for $n \geq 10$.

Thus, for samples of size 10 or more, the standardized variable

$$z = \frac{W - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

has approximately the standard normal distribution.

9.153 Large-Sample Wilcoxon Signed-Rank Test. Formulate a hypothesis-testing procedure for a Wilcoxon signed-rank test that uses the test statistic z given in the preceding box.

9.154 Birth Weights. Refer to Exercise 9.140.

a. Use the procedure you formulated in Exercise 9.153 to perform the hypothesis test in Exercise 9.140.

b. Compare your result in part (a) to the one you obtained in Exercise 9.140, where the normal approximation was not used.

9.155 The Distribution of *W*. In this exercise, you are to obtain the distribution of the variable W for samples of size 3 so that you can see how the Wilcoxon signed-rank table is constructed.

a. The rows of the following table give all possible signs for the signed ranks in a Wilcoxon signed-rank test with $n = 3$. For instance, the first row covers the possibility that all three observations are greater than μ_0 and thus have positive sign ranks. Fill in the empty column with values of W. (*Hint:* The first entry is 6, and the last is 0.)

Rank			
1	2	3	***W***
+	+	+	
+	+	−	
+	−	+	
+	−	−	
−	+	+	
−	+	−	
−	−	+	
−	−	−	

b. If the null hypothesis H_0: $\mu = \mu_0$ is true, what percentages of samples will match any particular row of the table? (*Hint:* The answer is the same for all rows.)

c. Use the answer from part (b) to obtain the distribution of W for samples of size 3.

d. Draw a relative-frequency histogram of the distribution obtained in part (c).

e. Use your histogram from part (d) to find $W_{0.125}$ for a sample size of 3.

9.156 The Distribution of *W*. Repeat Exercise 9.155 for samples of size 4. (*Hint:* The table will have 16 rows.)

One-Median Sign Test. Recall that the Wilcoxon signed-rank test, which can be used to perform a hypothesis test for a population median, η, requires that the variable under consideration has a symmetric distribution. If that is not the case, the **one-median sign test** (or simply the **sign test**) can be used instead. The one-median sign test is also known as the **one-sample sign test** and the **one-variable sign test.** Technically, like the Wilcoxon signed-rank test, use of the sign test requires that the variable under consideration has a continuous distribution. In practice, however, that restriction is usually ignored.

If the null hypothesis H_0: $\eta = \eta_0$ is true, the probability is 0.5 of an observation exceeding η_0. Therefore, in a simple random sample of size n, the number of observations, s, that exceed η_0 has a binomial distribution with parameters n and 0.5.

To perform a sign test, first assign a "+" sign to each observation in the sample that exceeds η_0 and then obtain the number of "+" signs, which we denote s_0. The P-value for the hypothesis test can be found by applying Exercise 9.63 on page 379 and obtaining the required binomial probability.

9.157 Assuming that the null hypothesis H_0: $\eta = \eta_0$ is true, answer the following questions.

a. Why is the probability of an observation exceeding η_0 equal to 0.5?

b. In a simple random sample of size n, why does the number of observations that exceed η_0 have a binomial distribution with parameters n and 0.5?

9.158 The sign test can be used whether or not the variable under consideration has a symmetric distribution. If the distribution is in fact symmetric, the Wilcoxon signed-rank test is preferable. Why do you think that is so?

9.159 What advantage does the sign test have over the Wilcoxon signed-rank test?

9.160 Explain how to proceed with a sign test if one or more of the observations equals η_0, the value specified in the null hypothesis for the population median.

*In Exercises **9.161–9.166**,*

a. apply the sign test to the specified exercise.

b. compare your result in part (a) to that obtained by using the Wilcoxon signed-rank test earlier in this exercise section.

9.161 **Global Warming?** Exercise 9.135.

9.162 **Happy-Life Years.** Exercise 9.136.

9.163 **How Old People Are.** Exercise 9.137.

9.164 **Beverage Expenditures.** Exercise 9.138.

9.165 **Pricing Mustangs.** Exercise 9.139.

9.166 **Birth Weights.** Exercise 9.140.

9.7 Type II Error Probabilities; Power*

As you learned in Section 9.1, hypothesis tests do not always yield correct conclusions; they have built-in margins of error. An important part of planning a study is to consider both types of errors that can be made and their effects.

Recall that two types of errors are possible with hypothesis tests. One is a Type I error: rejecting a true null hypothesis. The other is a Type II error: not rejecting a false null hypothesis. Also recall that the probability of making a Type I error is called the significance level of the hypothesis test and is denoted α, and that the probability of making a Type II error is denoted β.

In this section, we show how to compute Type II error probabilities. We also investigate the concept of the *power of a hypothesis test.* Although the discussion is limited to the one-mean z-test, the ideas apply to any hypothesis test.

Computing Type II Error Probabilities

The probability of making a Type II error depends on the sample size, the significance level, and the true value of the parameter under consideration.

EXAMPLE 9.23 Computing Type II Error Probabilities

Questioning Gas Mileage Claims The manufacturer of a new model car, the Orion, claims that a typical car gets 26 miles per gallon (mpg). A consumer advocacy group is skeptical of this claim and thinks that the mean gas mileage, μ, of all Orions may be less than 26 mpg. The group plans to perform the hypothesis test

$$H_0\text{: } \mu = 26 \text{ mpg (manufacturer's claim)}$$
$$H_a\text{: } \mu < 26 \text{ mpg (consumer group's conjecture)},$$

at the 5% significance level, using a sample of 30 Orions. Find the probability of making a Type II error if the true mean gas mileage of all Orions is

a. 25.8 mpg. **b.** 25.0 mpg.

Assume that gas mileages of Orions are normally distributed with a standard deviation of 1.4 mpg.

Solution The inference under consideration is a left-tailed hypothesis test for a population mean at the 5% significance level. The test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 26}{1.4/\sqrt{30}}.$$

We first express the decision criterion of whether or not to reject the null hypothesis in terms of the value of the test statistic, z.

- *Critical-value approach:* The critical value is $-z_\alpha = -z_{0.05} = -1.645$. Consequently, the value of the test statistic falls in the rejection region, and hence we reject the null hypothesis, if and only if $z \leq -1.645$.
- *P-value approach:* We reject the null hypothesis if and only if $P \leq \alpha = 0.05$, which happens if and only if the area under the standard normal curve to the left of the value of the test statistic is at most 0.05. Referring to Table II, we see that we reject the null hypothesis if and only if $z \leq -1.645$.

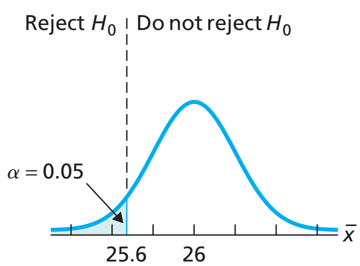
Therefore, a decision criterion for the hypothesis test is: If $z \leq -1.645$, reject H_0; if $z > -1.645$, do not reject H_0.

Computing Type II error probabilities is somewhat simpler if the decision criterion is expressed in terms of $\bar{x}$ instead of z. To do that here, we must find the sample mean that is 1.645 standard deviations below the null hypothesis population mean of 26:

$$\bar{x} = 26 - 1.645 \cdot \frac{1.4}{\sqrt{30}} = 25.6.$$

The decision criterion can thus be expressed in terms of $\bar{x}$ as: If $\bar{x} \leq 25.6$ mpg, reject H_0; if $\bar{x} > 25.6$ mpg, do not reject H_0. See Fig. 9.28.

FIGURE 9.28
Decision criterion for the gas mileage illustration ($\alpha = 0.05$, $n = 30$)

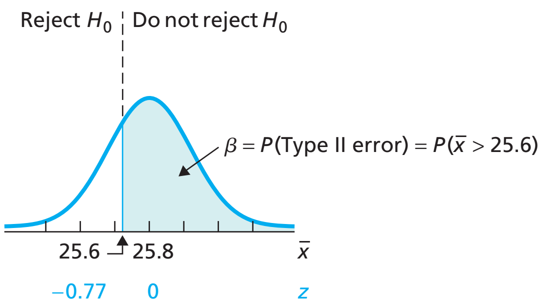


a. If $\mu = 25.8$ mpg, then

- $\mu_{\bar{x}} = \mu = 25.8$,
- $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 1.4/\sqrt{30} = 0.26$, and
- $\bar{x}$ is normally distributed.

Thus, the variable $\bar{x}$ is normally distributed with a mean of 25.8 mpg and a standard deviation of 0.26 mpg. The normal curve for $\bar{x}$ is shown in Fig. 9.29.

FIGURE 9.29
Determining the probability of a Type II error if $\mu = 25.8$ mpg



z-score computation:	Area to the left of z:
$\bar{x} = 25.6 \longrightarrow z = \dfrac{25.6 - 25.8}{0.26} = -0.77$	0.2206

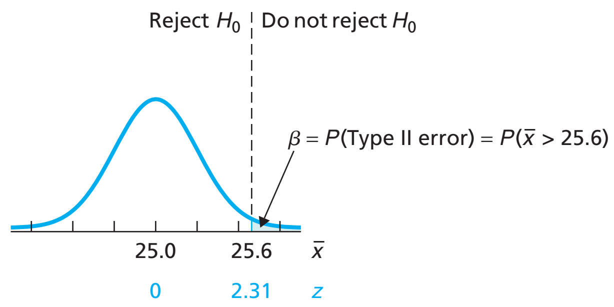
Shaded area = 1 − 0.2206 = 0.7794

A Type II error occurs if we do not reject H_0, that is, if $\bar{x} > 25.6$ mpg. The probability of this happening equals the percentage of all samples whose means exceed 25.6 mpg, which we obtain in Fig. 9.29. Thus, if the true mean gas mileage of all Orions is 25.8 mpg, the probability of making a Type II error is 0.7794; that is, $\beta = 0.7794$.

Interpretation There is roughly a 78% chance that the consumer group will fail to reject the manufacturer's claim that the mean gas mileage of all Orions is 26 mpg when in fact the true mean is 25.8 mpg.

Although this result is a rather high chance of error, we probably would not expect the hypothesis test to detect such a small difference in mean gas mileage (25.8 mpg as opposed to 26 mpg) with a sample size of only 30.

b. We proceed as we did in part (a), but this time we assume that $\mu = 25.0$ mpg. Figure 9.30 shows the required computations.



z-score computation:

$$\bar{x} = 25.6 \longrightarrow z = \frac{25.6 - 25.0}{0.26} = 2.31$$

Area to the left of z: 0.9896

Shaded area = 1 − 0.9896 = 0.0104

FIGURE 9.30
Determining the probability of a Type II error if $\mu = 25.0$ mpg

From Fig. 9.30, if the true mean gas mileage of all Orions is 25.0 mpg, the probability of making a Type II error is 0.0104; that is, $\beta = 0.0104$.

Interpretation There is only about a 1% chance that the consumer group will fail to reject the manufacturer's claim that the mean gas mileage of all Orions is 26 mpg when in fact the true mean is 25.0 mpg.

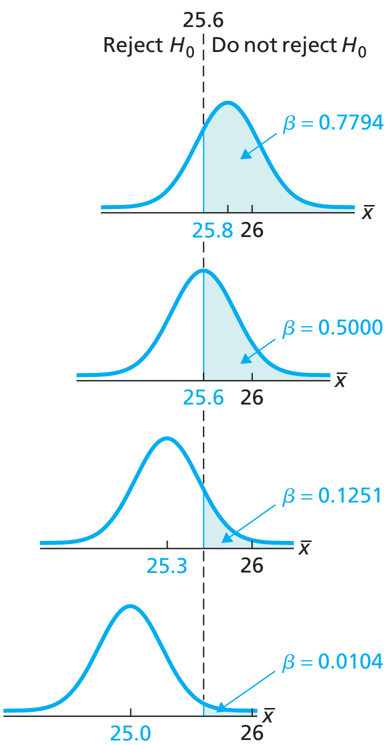


FIGURE 9.31
Type II error probabilities for $\mu = 25.8$, 25.6, 25.3, and 25.0 ($\alpha = 0.05$, $n = 30$)

Combining figures such as Figs. 9.29 and 9.30 gives a better understanding of Type II error probabilities. In Fig. 9.31, we combine those two figures with two others. The Type II error probabilities for the two additional values of μ were obtained by using the same techniques as those in Example 9.23.

Figure 9.31 shows clearly that the farther the true mean is from the null hypothesis mean of 26 mpg, the smaller will be the probability of a Type II error. This result is hardly surprising: We would expect that a false null hypothesis is more likely to be detected when the true mean is far from the null hypothesis mean than when the true mean is close to the null hypothesis mean.

Power and Power Curves

In modern statistical practice, analysts generally use the probability of not making a Type II error, called the **power,** to appraise the performance of a hypothesis test. Once we know the Type II error probability, β, obtaining the power is simple—we just subtract β from 1.

DEFINITION 9.6

What Does It Mean?

The power of a hypothesis test is between 0 and 1 and measures the ability of the hypothesis test to detect a false null hypothesis. If the power is near 0, the hypothesis test is not very good at detecting a false null hypothesis; if the power is near 1, the hypothesis test is extremely good at detecting a false null hypothesis.

Power

The **power** of a hypothesis test is the probability of not making a Type II error, that is, the probability of rejecting a false null hypothesis. We have

$$\text{Power} = 1 - P(\text{Type II error}) = 1 - \beta.$$

In reality, the true value of the parameter in question will be unknown. Consequently, constructing a table of powers for various values of the parameter is helpful in evaluating the effectiveness of the hypothesis test.

For the gas mileage illustration—where the parameter in question is the mean gas mileage, μ, of all Orions—we have already obtained the Type II error probability, β, when the true mean is 25.8 mpg, 25.6 mpg, 25.3 mpg, and 25.0 mpg, as depicted in Fig. 9.31. Similar calculations yield the other β probabilities shown in the second column of Table 9.16. The third column of Table 9.16 shows the power that corresponds to each value of μ, obtained by subtracting β from 1.

TABLE 9.16
Selected Type II error probabilities and powers for the gas mileage illustration ($\alpha = 0.05$, $n = 30$)

True mean μ	P (Type II error) β	Power $1 - \beta$
25.9	0.8749	0.1251
25.8	0.7794	0.2206
25.7	0.6480	0.3520
25.6	0.5000	0.5000
25.5	0.3520	0.6480
25.4	0.2206	0.7794
25.3	0.1251	0.8749
25.2	0.0618	0.9382
25.1	0.0274	0.9726
25.0	0.0104	0.9896
24.9	0.0036	0.9964
24.8	0.0010	0.9990

APPLET

Applet 9.2

You try it!

Exercise 9.175 on page 420

We can use Table 9.16 to evaluate the overall effectiveness of the hypothesis test. We can also obtain from Table 9.16 a visual display of that effectiveness by plotting points of power against μ and then connecting the points with a smooth curve. The resulting curve is called a **power curve** and is shown in Fig. 9.32. In general, the closer a power curve is to 1 (i.e., the horizontal line 1 unit above the horizontal axis), the better the hypothesis test is at detecting a false null hypothesis.

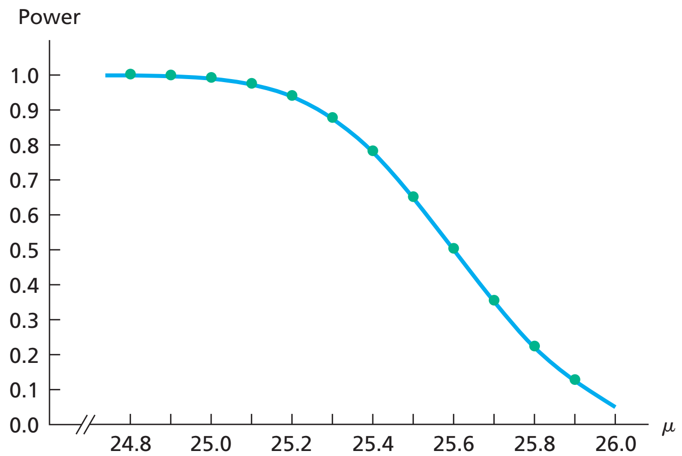


FIGURE 9.32
Power curve for the gas mileage illustration ($\alpha = 0.05$, $n = 30$)

Sample Size and Power

Ideally, both Type I and Type II errors should have small probabilities. In terms of significance level and power, then, we want to specify a small significance level (close to 0) and yet have large power (close to 1).

Key Fact 9.1 (page 363) implies that the smaller we specify the significance level, the smaller will be the power. However, by using a large sample, we can have both a small significance level and large power, as shown in the next example.

EXAMPLE 9.24 The Effect of Sample Size on Power

Questioning Gas Mileage Claims Consider again the hypothesis test for the gas mileage illustration of Example 9.23,

$$H_0\colon \mu = 26 \text{ mpg (manufacturer's claim)}$$
$$H_a\colon \mu < 26 \text{ mpg (consumer group's conjecture)},$$

where μ is the mean gas mileage of all Orions. In Table 9.16, we presented selected powers when $\alpha = 0.05$ and $n = 30$. Now suppose that we keep the significance level at 0.05 but increase the sample size from 30 to 100.

a. Construct a table of powers similar to Table 9.16.
b. Use the table from part (a) to draw the power curve for $n = 100$, and compare it to the power curve drawn earlier for $n = 30$.
c. Interpret the results from parts (a) and (b).

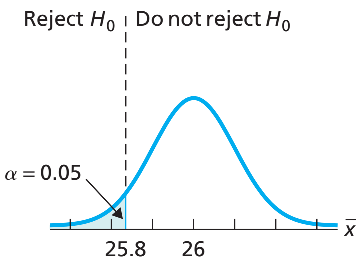
Solution The inference under consideration is a left-tailed hypothesis test for a population mean at the 5% significance level. The test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 26}{1.4/\sqrt{100}}.$$

From Example 9.23, a decision criterion for the hypothesis test is: If $z \le -1.645$, reject H_0; if $z > -1.645$, do not reject H_0.

As we noted earlier, computing Type II error probabilities is somewhat simpler if the decision criterion is expressed in terms of $\bar{x}$ instead of z. To do that here, we must find the sample mean that is 1.645 standard deviations below the null hypothesis population mean of 26:

FIGURE 9.33
Decision criterion for the gas mileage illustration ($\alpha = 0.05$, $n = 100$)



$$\bar{x} = 26 - 1.645 \cdot \frac{1.4}{\sqrt{100}} = 25.8.$$

The decision criterion can thus be expressed in terms of $\bar{x}$ as: If $\bar{x} \le 25.8$ mpg, reject H_0; if $\bar{x} > 25.8$ mpg, do not reject H_0. See Fig. 9.33.

a. Now that we have expressed the decision criterion in terms of $\bar{x}$, we can obtain Type II error probabilities by using the same techniques as in Example 9.23. We computed the Type II error probabilities that correspond to several values of μ, as shown in Table 9.17. The third column of Table 9.17 displays the powers.
b. Using Table 9.17, we can draw the power curve for the gas mileage illustration when $n = 100$, as shown in Fig. 9.34. For comparison purposes, we have also reproduced from Fig. 9.32 the power curve for $n = 30$.
c. **Interpretation** Comparing Tables 9.16 and 9.17 shows that each power is greater when $n = 100$ than when $n = 30$. Figure 9.34 displays that fact visually.

You try it!

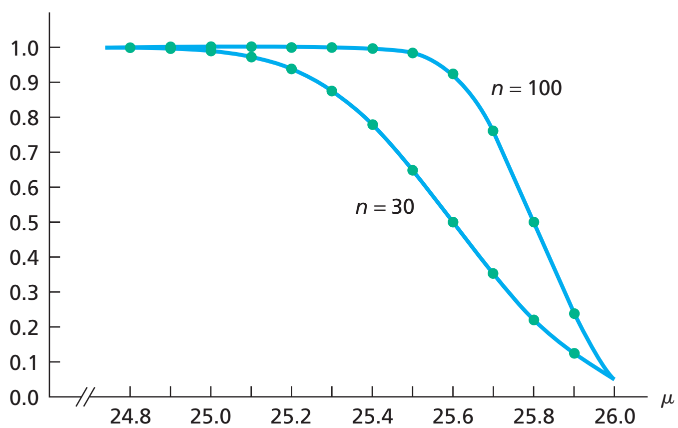
Exercise 9.181 on page 420

TABLE 9.17
Selected Type II error probabilities and powers for the gas mileage illustration ($\alpha = 0.05$, $n = 100$)

True mean μ	P (Type II error) β	Power $1 - \beta$
25.9	0.7611	0.2389
25.8	0.5000	0.5000
25.7	0.2389	0.7611
25.6	0.0764	0.9236
25.5	0.0162	0.9838
25.4	0.0021	0.9979
25.3	0.0002	0.9998
25.2	0.0000†	1.0000‡
25.1	0.0000	1.0000
25.0	0.0000	1.0000
24.9	0.0000	1.0000
24.8	0.0000	1.0000

† For $\mu \leq 25.2$, the β probabilities are 0 to four decimal places.
‡ For $\mu \leq 25.2$, the powers are 1 to four decimal places.

FIGURE 9.34
Power curves for the gas mileage illustration when $n = 30$ and $n = 100$ ($\alpha = 0.05$)



In the preceding example, we found that increasing the sample size without changing the significance level increased the power. This relationship is true in general.

KEY FACT 9.9

Sample Size and Power

For a fixed significance level, increasing the sample size increases the power.

What Does It Mean?

By using a sufficiently large sample size, we can obtain a hypothesis test with as much power as we want.

In practice, larger sample sizes tend to increase the cost of a study. Consequently, we must balance, among other things, the cost of a large sample against the cost of possible errors.

As we have indicated, power is a useful way to evaluate the overall effectiveness of a hypothesis-testing procedure. However, power can also be used to compare different procedures. For example, a researcher might decide between two hypothesis-testing procedures on the basis of which test is more powerful for the situation under consideration.

THE TECHNOLOGY CENTER

As we have shown, obtaining Type II error probabilities or powers is computationally intensive. Moreover, determining those quantities by hand can result in substantial roundoff error. Therefore, in practice, Type II error probabilities and powers are almost always calculated by computer.

Exercises 9.7

Understanding the Concepts and Skills

9.167 Why don't hypothesis tests always yield correct decisions?

9.168 Define each term.
a. Type I error **b.** Type II error **c.** Significance level

9.169 Explain the meaning of each of the following in the context of hypothesis testing.
a. α **b.** β **c.** $1 - \beta$

9.170 What does the power of a hypothesis test tell you? How is it related to the probability of making a Type II error?

9.171 Why is it useful to obtain the power curve for a hypothesis test?

9.172 What happens to the power of a hypothesis test if the sample size is increased without changing the significance level? Explain your answer.

9.173 What happens to the power of a hypothesis test if the significance level is decreased without changing the sample size? Explain your answer.

9.174 Suppose that you must choose between two procedures for performing a hypothesis test—say, Procedure A and Procedure B. Further suppose that, for the same sample size and significance level, Procedure A has less power than Procedure B. Which procedure would you choose? Explain your answer.

*In Exercises **9.175–9.180**, we have given a hypothesis testing situation and* (i) *the population standard deviation, σ,* (ii) *a significance level,* (iii) *a sample size, and* (iv) *some values of μ. For each exercise,*

a. express the decision criterion for the hypothesis test in terms of $\bar{x}$.
b. determine the probability of a Type I error.
c. construct a table similar to Table 9.16 on page 417 that provides the probability of a Type II error and the power for each of the given values of μ.
d. use the table obtained in part (c) to draw the power curve.

9.175 Toxic Mushrooms? The null and alternative hypotheses obtained in Exercise 9.5 on page 364 are, respectively,

$$H_0\colon \mu = 0.5 \text{ ppm}$$
$$H_a\colon \mu > 0.5 \text{ ppm},$$

where μ is the mean cadmium level in *Boletus pinicola* mushrooms.
i. $\sigma = 0.37$ ii. $\alpha = 0.05$ iii. $n = 12$
iv. $\mu = 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85$

9.176 Agriculture Books. The null and alternative hypotheses obtained in Exercise 9.6 on page 365 are, respectively,

$$H_0\colon \mu = \$57.61$$
$$H_a\colon \mu \neq \$57.61,$$

where μ is this year's mean retail price of agriculture books.
i. $\sigma = 8.45$ ii. $\alpha = 0.10$ iii. $n = 28$
iv. $\mu = 53, 54, 55, 56, 57, 58, 59, 60, 61, 62$

9.177 Iron Deficiency? The null and alternative hypotheses obtained in Exercise 9.7 on page 365 are, respectively,

$$H_0\colon \mu = 18 \text{ mg}$$
$$H_a\colon \mu < 18 \text{ mg},$$

where μ is the mean iron intake (per day) of all adult females under the age of 51 years.
i. $\sigma = 4.2$ ii. $\alpha = 0.01$ iii. $n = 45$
iv. $\mu = 15.50, 15.75, 16.00, 16.25, 16.50, 16.75, 17.00, 17.25, 17.50, 17.75$

9.178 Early-Onset Dementia. The null and alternative hypotheses obtained in Exercise 9.8 on page 365 are, respectively,

$$H_0\colon \mu = 55 \text{ years old}$$
$$H_a\colon \mu < 55 \text{ years old},$$

where μ is the mean age at diagnosis of all people with early-onset dementia.
i. $\sigma = 6.8$ ii. $\alpha = 0.01$ iii. $n = 21$
iv. $\mu = 47, 48, 49, 50, 51, 52, 53, 54$

9.179 Serving Time. The null and alternative hypotheses obtained in Exercise 9.9 on page 365 are, respectively,

$$H_0\colon \mu = 16.7 \text{ months}$$
$$H_a\colon \mu \neq 16.7 \text{ months},$$

where μ is the mean length of imprisonment for motor-vehicle-theft offenders in Sydney, Australia.
i. $\sigma = 6.0$ ii. $\alpha = 0.05$ iii. $n = 100$
iv. $\mu = 14.0, 14.5, 15.0, 15.5, 16.0, 16.5, 17.0, 17.5, 18.0, 18.5, 19.0$

9.180 Worker Fatigue. The null and alternative hypotheses obtained in Exercise 9.10 on page 365 are, respectively,

$$H_0\colon \mu = 72 \text{ bpm}$$
$$H_a\colon \mu > 72 \text{ bpm},$$

where μ is the mean post-work heart rate of all casting workers.
i. $\sigma = 11.2$ ii. $\alpha = 0.05$ iii. $n = 29$
iv. $\mu = 73, 74, 75, 76, 77, 78, 79, 80$

9.181 Toxic Mushrooms? Repeat parts (a)–(d) of Exercise 9.175 for a sample size of 20. Compare your power curves for the two sample sizes, and explain the principle being illustrated.

9.182 Agriculture Books. Repeat parts (a)–(d) of Exercise 9.176 for a sample size of 50. Compare your power curves for the two sample sizes, and explain the principle being illustrated.

9.183 Serving Time. Repeat parts (a)–(d) of Exercise 9.179 for a sample size of 40. Compare your power curves for the two sample sizes, and explain the principle being illustrated.

9.184 Early-Onset Dementia. Repeat parts (a)–(d) of Exercise 9.178 for a sample size of 15. Compare your power curves for the two sample sizes, and explain the principle being illustrated.

Extending the Concepts and Skills

9.185 Consider a right-tailed hypothesis test for a population mean with null hypothesis H_0: $\mu = \mu_0$.
a. Draw the ideal power curve.
b. Explain what your curve in part (a) portrays.

9.186 Consider a left-tailed hypothesis test for a population mean with null hypothesis H_0: $\mu = \mu_0$.
a. Draw the ideal power curve.
b. Explain what your curve in part (a) portrays.

9.187 Consider a two-tailed hypothesis test for a population mean with null hypothesis H_0: $\mu = \mu_0$.
a. Draw the ideal power curve.
b. Explain what your curve in part (a) portrays.

9.188 Class Project: Questioning Gas Mileage. This exercise can be done individually or, better yet, as a class project. Refer to the gas mileage hypothesis test of Example 9.23 on page 414. Recall that the null and alternative hypotheses are

$$H_0: \mu = 26 \text{ mpg (manufacturer's claim)}$$
$$H_a: \mu < 26 \text{ mpg (consumer group's conjecture),}$$

where μ is the mean gas mileage of all Orions. Also recall that the mileages are normally distributed with a standard deviation of 1.4 mpg. Figure 9.28 on page 415 portrays the decision criterion for a test at the 5% significance level with a sample size of 30. Suppose that, in reality, the mean gas mileage of all Orions is 25.4 mpg.
a. Determine the probability of making a Type II error.
b. Simulate 100 samples of 30 gas mileages each.
c. Determine the mean of each sample in part (b).
d. For the 100 samples obtained in part (b), about how many would you expect to lead to nonrejection of the null hypothesis? Explain your answer.
e. For the 100 samples obtained in part (b), determine the number that lead to nonrejection of the null hypothesis.
f. Compare your answers from parts (d) and (e), and comment on any observed difference.

9.8 Which Procedure Should Be Used?*

In this chapter, you learned three procedures for performing a hypothesis test for one population mean: the z-test, the t-test, and the Wilcoxon signed-rank test. The z-test and t-test are designed to be used when the variable under consideration has a normal distribution. In such cases, the z-test applies when the population standard deviation is known, and the t-test applies when the population standard deviation is unknown.

Recall that both the z-test and the t-test are approximately correct when the sample size is large, regardless of the distribution of the variable under consideration. Moreover, these two tests should be used cautiously when outliers are present. Refer to Key Fact 9.7 on page 379 for guidelines covering use of the z-test and t-test.

Recall further that the Wilcoxon signed-rank test is designed to be used when the variable under consideration has a symmetric distribution. Unlike the z-test and t-test, the Wilcoxon signed-rank test is resistant to outliers.

We summarize the three procedures in Table 9.18. Each row of the table gives the type of test, the conditions required for using the test, the test statistic, and the procedure to use. Note that we used the abbreviations "normal population" for "the variable under consideration is normally distributed," "W-test" for "Wilcoxon signed-rank test," and "symmetric population" for "the variable under consideration has a symmetric distribution."

TABLE 9.18
Summary of hypothesis-testing procedures for one population mean, μ. The null hypothesis for all tests is H_0: $\mu = \mu_0$

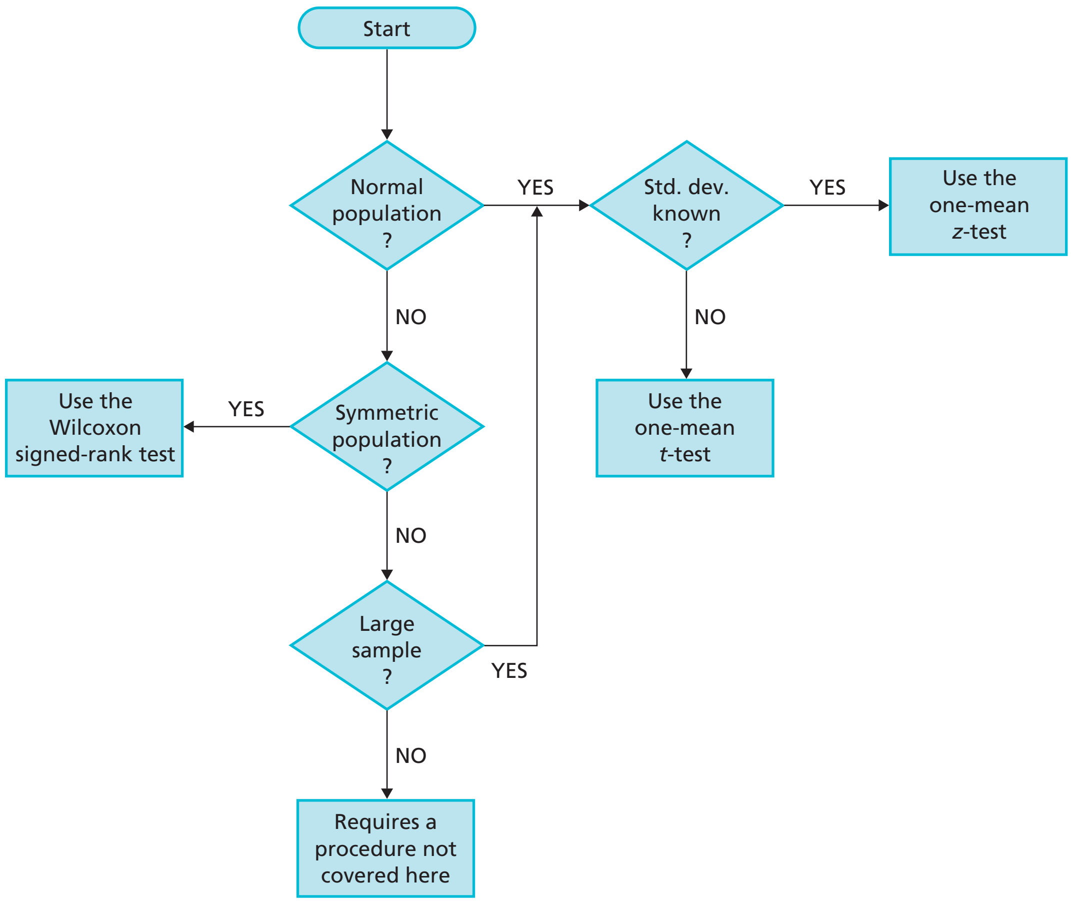
Type	Assumptions	Test statistic	Procedure to use
z-test	1. Simple random sample 2. Normal population or large sample 3. σ known	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	9.1 (page 380)
t-test	1. Simple random sample 2. Normal population or large sample 3. σ unknown	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ (df $= n - 1$)	9.2 (page 394)
W-test	1. Simple random sample 2. Symmetric population	W = sum of positive ranks	9.3 (page 404)

*The parametric and nonparametric methods discussed in this chapter are prerequisite to this section.

In selecting the correct procedure, keep in mind that the best choice is the procedure expressly designed for the type of distribution under consideration, if such a procedure exists, and that the *z*-test and *t*-test are only approximately correct for large samples from nonnormal populations. For instance, suppose that the variable under consideration is normally distributed and that the population standard deviation is known. Then both the *z*-test and Wilcoxon signed-rank test apply. The *z*-test applies because the variable under consideration is normally distributed and σ is known; the W-test applies because a normal distribution is symmetric. The correct procedure, however, is the *z*-test because it is designed specifically for variables that have a normal distribution.

The flowchart shown in Fig. 9.35 summarizes the preceding discussion.

FIGURE 9.35 Flowchart for choosing the correct hypothesis testing procedure for a population mean



In practice, you need to look at the sample data to ascertain the type of distribution before selecting the appropriate procedure. We recommend using a normal probability plot and either a stem-and-leaf diagram (for small or moderate-size samples) or a histogram (for moderate-size or large samples).

EXAMPLE 9.25 Choosing the Correct Hypothesis-Testing Procedure

Chicken Consumption The U.S. Department of Agriculture publishes data on chicken consumption in *Food Consumption, Prices, and Expenditures*. In 2006, the average person consumed 61.3 lb of chicken. A simple random sample of 17 people had the chicken consumption for last year shown in Table 9.19.

Suppose that we want to use the sample data in Table 9.19 to decide whether last year's mean chicken consumption has changed from the 2006 mean of 61.3 lb.

Then we want to perform the hypothesis test

$$H_0\text{: } \mu = 61.3 \text{ lb (mean chicken consumption has not changed)}$$
$$H_a\text{: } \mu \neq 61.3 \text{ lb (mean chicken consumption has changed),}$$

where μ is last year's mean chicken consumption. Which procedure should be used to perform the hypothesis test?

TABLE 9.19
Sample of last year's chicken consumption (lb)

57	69	63	49	63	61
72	65	91	59	0	82
60	75	55	80	73	

Solution We begin by drawing a normal probability plot and a stem-and-leaf diagram of the sample data in Table 9.19, as shown in Fig. 9.36.

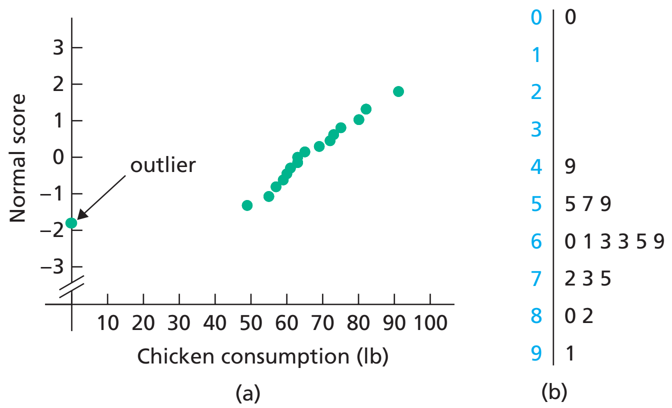


FIGURE 9.36
(a) Normal probability plot and (b) stem-and-leaf diagram of the chicken-consumption data in Table 9.19

Next, we consult the flowchart in Fig. 9.35 and the graphs in Fig. 9.36. The first question is whether the variable under consideration is normally distributed. The normal probability plot in Fig. 9.36(a) shows an outlier, so the answer to the first question is probably "No."

This result leads to the next question: Does the variable under consideration have a symmetric distribution? The stem-and-leaf diagram in Fig. 9.36(b) suggests that we can reasonably assume that the answer to that question is "Yes."

The "Yes" answer to the preceding question leads us to the box in Fig. 9.35 that states Use the Wilcoxon signed-rank test.

Interpretation An appropriate procedure for carrying out the hypothesis test is the Wilcoxon signed-rank test.

Exercises 9.8

Understanding the Concepts and Skills

9.189 In this chapter, we presented three procedures for conducting a hypothesis test for one population mean.
a. Identify the three procedures by name.
b. List the assumptions for using each procedure.
c. Identify the test statistic for each procedure.

9.190 Suppose that you want to perform a hypothesis test for a population mean. Assume that the variable under consideration is normally distributed and that the population standard deviation is unknown.
a. Can the t-test be used to perform the hypothesis test? Explain your answer.
b. Can the Wilcoxon signed-rank test be used to perform the hypothesis test? Explain your answer.
c. Which procedure is preferable, the t-test or the Wilcoxon signed-rank test? Explain your answer.

9.191 Suppose that you want to perform a hypothesis test for a population mean. Assume that the variable under consideration has a symmetric nonnormal distribution and that the population standard deviation is unknown. Further assume that the sample size is large and that no outliers are present in the sample data.
a. Can the t-test be used to perform the hypothesis test? Explain your answer.
b. Can the Wilcoxon signed-rank test be used to perform the hypothesis test? Explain your answer.
c. Which procedure is preferable, the t-test or the Wilcoxon signed-rank test? Explain your answer.

9.192 Suppose that you want to perform a hypothesis test for a population mean. Assume that the variable under consideration has a highly skewed distribution and that the population standard deviation is known. Further assume that the sample size is large and that no outliers are present in the sample data.

a. Can the z-test be used to perform the hypothesis test? Explain your answer.

b. Can the Wilcoxon signed-rank test be used to perform the hypothesis test? Explain your answer.

*In Exercises **9.193–9.200**, we have provided a normal probability plot and either a stem-and-leaf diagram or a frequency histogram for a set of sample data. The intent is to employ the sample data to perform a hypothesis test for the mean of the population from which the data were obtained. In each case, consult the graphs provided and the flowchart in Fig. 9.35 to decide which procedure should be used.*

9.193 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.37; σ is known.

9.194 The normal probability plot and histogram of the data are shown in Fig. 9.38; σ is known.

9.195 The normal probability plot and histogram of the data are shown in Fig. 9.39; σ is unknown.

9.196 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.40; σ is unknown.

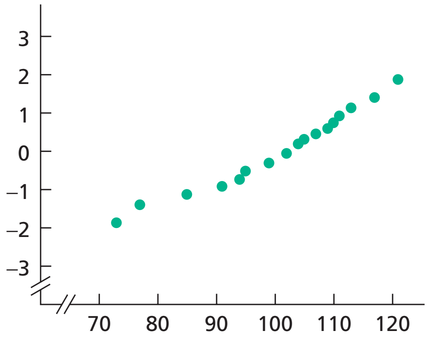
9.197 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.41; σ is unknown.

9.198 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.42; σ is unknown. (*Note:* The decimal parts of the observations were removed before the stem-and-leaf diagram was constructed.)

9.199 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.43; σ is known.

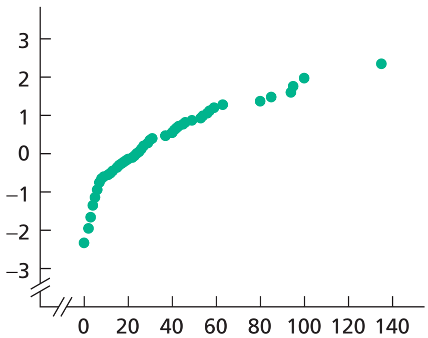
9.200 The normal probability plot and stem-and-leaf diagram of the data are shown in Fig. 9.44; σ is known.

FIGURE 9.37
Normal probability plot and stem-and-leaf diagram for Exercise 9.193



7	3
7	7
8	
8	5
9	1 4
9	5 5 9
10	2 2 2 4
10	5 7 9
11	0 1 3
11	7
12	1

FIGURE 9.38
Normal probability plot and histogram for Exercise 9.194



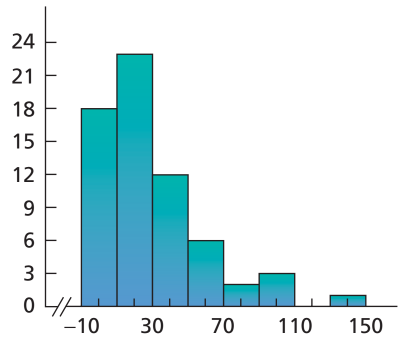
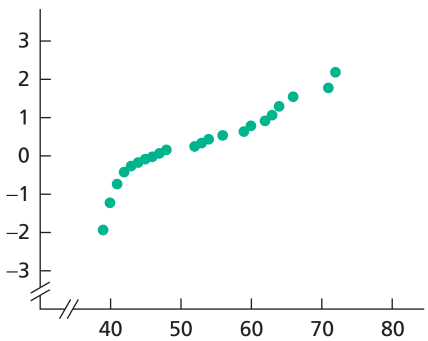


FIGURE 9.39
Normal probability plot and histogram for Exercise 9.195



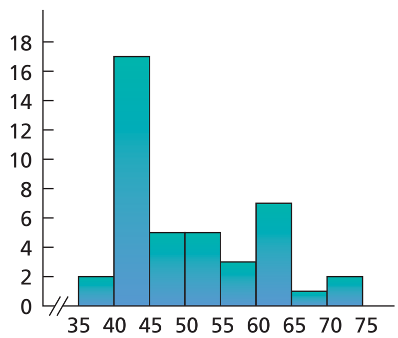
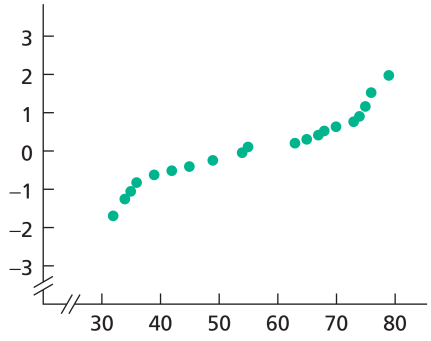
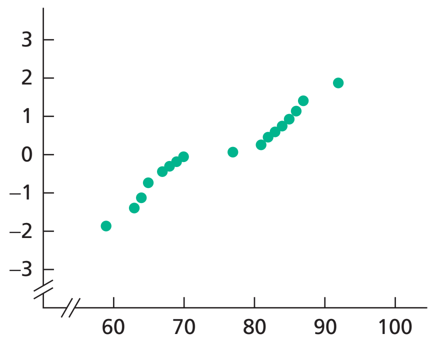


FIGURE 9.40
Normal probability plot and stem-and-leaf diagram for Exercise 9.196



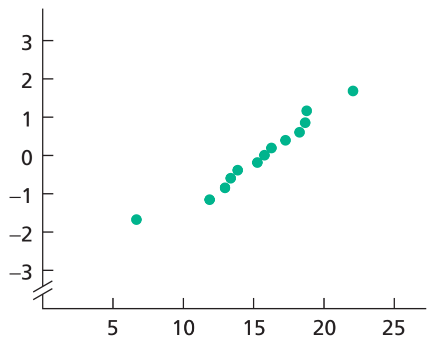
3	2 2 4
3	5 6 6 9
4	2
4	5 9 9
5	4 4
5	5
6	3
6	5 7 8
7	0 3 4
7	5 5 6 9

FIGURE 9.41
Normal probability plot and stem-and-leaf diagram for Exercise 9.197



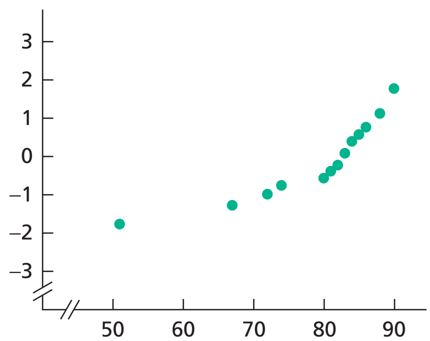
5	9
6	3 4
6	5 5 5 7 8 9
7	0
7	7
8	1 1 2 3 4
8	5 6 7
9	2

FIGURE 9.42
Normal probability plot and stem-and-leaf diagram for Exercise 9.198



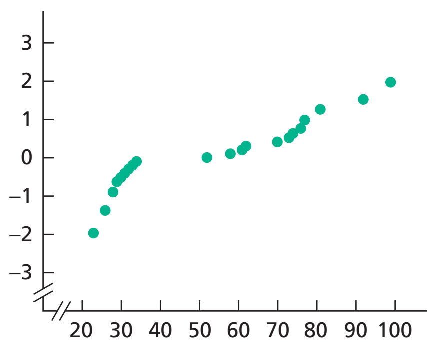
0	6
0	
1	1
1	3 3 3
1	5 5
1	6 7
1	8 8 8
2	
2	2

FIGURE 9.43
Normal probability plot and stem-and-leaf diagram for Exercise 9.199



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

FIGURE 9.44
Normal probability plot and stem-and-leaf diagram for Exercise 9.200



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

CHAPTER IN REVIEW

You Should Be Able to

1. use and understand the formulas in this chapter.
2. define and apply the terms that are associated with hypothesis testing.
3. choose the null and alternative hypotheses for a hypothesis test.
4. explain the basic logic behind hypothesis testing.
5. define and apply the concepts of Type I and Type II errors.
6. understand the relation between Type I and Type II error probabilities.
7. state and interpret the possible conclusions for a hypothesis test.
8. understand and apply the critical-value approach to hypothesis testing and/or the P-value approach to hypothesis testing.
9. perform a hypothesis test for one population mean when the population standard deviation is known.
10. perform a hypothesis test for one population mean when the population standard deviation is unknown.

*11. perform a hypothesis test for one population mean when the variable under consideration has a symmetric distribution.

*12. compute Type II error probabilities for a one-mean z-test.

*13. calculate the power of a hypothesis test.

*14. draw a power curve.

*15. understand the relationship between sample size, significance level, and power.

*16. decide which procedure should be used to perform a hypothesis test for one population mean.

Key Terms

REVIEW PROBLEMS

Understanding the Concepts and Skills

1. Explain the meaning of each term.
a. null hypothesis **b.** alternative hypothesis
c. test statistic **d.** significance level

2. The following statement appeared on a box of Tide laundry detergent: "Individual packages of Tide may weigh slightly more or less than the marked weight due to normal variations incurred with high speed packaging machines, but each day's production of Tide will average slightly above the marked weight."
a. Explain in statistical terms what the statement means.
b. Describe in words a hypothesis test for checking the statement.
c. Suppose that the marked weight is 76 ounces. State in words the null and alternative hypotheses for the hypothesis test. Then express those hypotheses in statistical terminology.

3. Regarding a hypothesis test:
a. What is the procedure, generally, for deciding whether the null hypothesis should be rejected?
b. How can the procedure identified in part (a) be made objective and precise?

4. There are three possible alternative hypotheses in a hypothesis test for a population mean. Identify them, and explain when each is used.

5. Two types of incorrect decisions can be made in a hypothesis test: a Type I error and a Type II error.
a. Explain the meaning of each type of error.
b. Identify the letter used to represent the probability of each type of error.

c. If the null hypothesis is in fact true, only one type of error is possible. Which type is that? Explain your answer.
d. If you fail to reject the null hypothesis, only one type of error is possible. Which type is that? Explain your answer.

6. For a fixed sample size, what happens to the probability of a Type II error if the significance level is decreased from 0.05 to 0.01?

*Problems **7–12** pertain to the critical-value approach to hypothesis testing.*

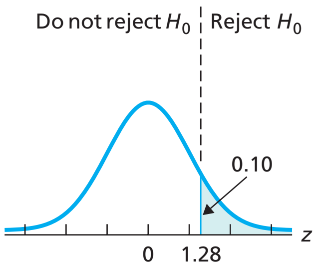
7. Explain the meaning of each term.
a. rejection region
b. nonrejection region
c. critical value(s)

8. True or false: A critical value is considered part of the rejection region.

9. Suppose that you want to conduct a left-tailed hypothesis test at the 5% significance level. How must the critical value be chosen?

10. Determine the critical value(s) for a one-mean z-test at the 1% significance level if the test is
a. right tailed.
b. left tailed.
c. two tailed.

11. The following graph portrays the decision criterion for a one-mean z-test, using the critical-value approach to hypothesis testing. The curve in the graph is the normal curve for the test statistic under the assumption that the null hypothesis is true.



Determine the
a. rejection region. **b.** nonrejection region.
c. critical value(s). **d.** significance level.
e. Draw a graph that depicts the answers that you obtained in parts (a)–(d).
f. Classify the hypothesis test as two tailed, left tailed, or right tailed.

12. State the general steps of the critical-value approach to hypothesis testing.

*Problems **13–20** pertain to the P-value approach to hypothesis testing.*

13. Define the P-value of a hypothesis test.

14. True or false: A P-value of 0.02 provides more evidence against the null hypothesis than a P-value of 0.03. Explain your answer.

15. State the decision criterion for a hypothesis test, using the P-value.

16. Explain why the P-value of a hypothesis test is also referred to as the observed significance level.

17. How is the P-value of a hypothesis test actually determined?

18. In each part, we have given the value obtained for the test statistic, z, in a one-mean z-test. We have also specified whether the test is two tailed, left tailed, or right tailed. Determine the P-value in each case and decide whether, at the 5% significance level, the data provide sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis.
a. $z = -1.25$; left-tailed test
b. $z = 2.36$; right-tailed test
c. $z = 1.83$; two-tailed test

19. State the general steps of the P-value approach to hypothesis testing.

20. Assess the evidence against the null hypothesis if the P-value of the hypothesis test is 0.062.

21. What is meant when we say that a hypothesis test is
a. exact? **b.** approximately correct?

22. Discuss the difference between statistical significance and practical significance.

23. In each part, we have identified a hypothesis-testing procedure for one population mean. State the assumptions required and the test statistic used in each case.
a. one-mean t-test **b.** one-mean z-test
***c.** Wilcoxon signed-rank test

***24.** Identify two advantages of nonparametric methods over parametric methods. When is a parametric procedure preferred? Explain your answer.

***25.** Regarding the power of a hypothesis test:
a. What does it represent?
b. What happens to the power of a hypothesis test if the significance level is kept at 0.01 while the sample size is increased from 50 to 100?

26. Cheese Consumption. The U.S. Department of Agriculture reports in *Food Consumption, Prices, and Expenditures* that the average American consumed 30.0 lb of cheese in 2001. Cheese consumption has increased steadily since 1960, when the average American ate only 8.3 lb of cheese annually. Suppose that you want to decide whether last year's mean cheese consumption is greater than the 2001 mean.
a. Identify the null hypothesis.
b. Identify the alternative hypothesis.
c. Classify the hypothesis test as two tailed, left tailed, or right tailed.

27. Cheese Consumption. The null and alternative hypotheses for the hypothesis test in Problem 26 are, respectively,

$$H_0\text{: } \mu = 30.0 \text{ lb (mean has not increased)}$$
$$H_a\text{: } \mu > 30.0 \text{ lb (mean has increased)},$$

where μ is last year's mean cheese consumption for all Americans. Explain what each of the following would mean.
a. Type I error **b.** Type II error **c.** Correct decision

Now suppose that the results of carrying out the hypothesis test lead to rejection of the null hypothesis. Classify that decision by error type or as a correct decision if in fact last year's mean cheese consumption

d. has not increased from the 2001 mean of 30.0 lb.
e. has increased from the 2001 mean of 30.0 lb.

28. Cheese Consumption. Refer to Problem 26. The following table provides last year's cheese consumption, in pounds, for 35 randomly selected Americans.

45	28	32	37	41	39	33
32	31	35	27	46	25	41
35	31	44	23	38	27	32
43	32	25	36	26	30	35
36	36	35	21	43	35	28

a. At the 10% significance level, do the data provide sufficient evidence to conclude that last year's mean cheese consumption for all Americans has increased over the 2001 mean? Assume that $\sigma = 6.9$ lb. Use a z-test. (*Note:* The sum of the data is 1183 lb.)
b. Given the conclusion in part (a), if an error has been made, what type must it be? Explain your answer.

29. Purse Snatching. The Federal Bureau of Investigation (FBI) compiles information on robbery and property crimes by type and selected characteristic and publishes its findings in *Population-at-Risk Rates and Selected Crime Indicators*. According to that document, the mean value lost to purse snatching was \$417 in 2004. For last year, 12 randomly selected purse-snatching offenses yielded the following values lost, to the nearest dollar.

364	488	314	428	324	252
521	436	499	430	320	472

Use a t-test to decide, at the 5% significance level, whether last year's mean value lost to purse snatching has decreased from the 2004 mean. The mean and standard deviation of the data are \$404.0 and \$86.8, respectively.

***30. Purse Snatching.** Refer to Problem 29.
a. Perform the required hypothesis test, using the Wilcoxon signed-rank test.
b. In performing the hypothesis test in part (a), what assumption did you make about the distribution of last year's values lost to purse snatching?
c. In Problem 29, we used the t-test to perform the hypothesis test. The assumption in that problem is that last year's values lost to purse snatching are normally distributed. If that assumption is true, why is it permissible to perform a Wilcoxon signed-rank test for the mean value lost?

***31. Purse Snatching.** Refer to Problems 29 and 30. If in fact last year's values lost to purse snatching are normally distributed, which is the preferred procedure for performing the hypothesis test—the t-test or the Wilcoxon signed-rank test? Explain your answer.

32. Betting the Spreads. College basketball, and particularly the NCAA basketball tournament, is a popular venue for gambling, from novices in office betting pools to high rollers. To encourage uniform betting across teams, Las Vegas oddsmakers assign a point spread to each game. The *point spread* is the oddsmakers' prediction for the number of points by which the favored team will win. If you bet on the favorite, you win the bet provided the favorite wins by more than the point spread; otherwise, you lose the bet. Is the point spread a good measure of the relative ability of the two teams? H. Stern and B. Mock addressed this question in the paper "College Basketball Upsets: Will a 16-Seed Ever Beat a 1-Seed?" (*Chance*, Vol. 11(1), pp. 27–31). They obtained the difference between the actual margin of victory and the point spread, called the *point-spread error,* for 2109 college basketball games. The mean point-spread error was found to be −0.2 point with a standard deviation of 10.9 points. For a particular game, a point-spread error of 0 indicates that the point spread was a perfect estimate of the two teams' relative abilities.
a. If, on average, the oddsmakers are estimating correctly, what is the (population) mean point-spread error?
b. Use the data to decide, at the 5% significance level, whether the (population) mean point-spread error differs from 0.
c. Interpret your answer in part (b).

***33. Cheese Consumption.** Refer to Problem 26. Suppose that you decide to use a z-test with a significance level of 0.10 and a sample size of 35. Assume that $\sigma = 6.9$ lb.
a. Determine the probability of a Type I error.
b. If last year's mean cheese consumption was 33.5 lb, identify the distribution of the variable $\bar{x}$, that is, the sampling distribution of the mean for samples of size 35.
c. Use part (b) to determine the probability, β, of a Type II error if in fact last year's mean cheese consumption was 33.5 lb.
d. Repeat parts (b) and (c) if in fact last year's mean cheese consumption was 30.5 lb, 31.0 lb, 31.5 lb, 32.0 lb, 32.5 lb, 33.0 lb, and 34.0 lb.
e. Use your answers from parts (c) and (d) to construct a table of selected Type II error probabilities and powers similar to Table 9.16 on page 417.
f. Use your answer from part (e) to construct the power curve.

Using a sample size of 60 instead of 35, repeat
g. part(b). **h.** part (c). **i.** part (d).
j. part (e). **k.** part (f).
l. Compare your power curves for the two sample sizes and explain the principle being illustrated.

*Problems **34** and **35** each include a normal probability plot and either a frequency histogram or a stem-and-leaf diagram for a set of sample data. The intent is to use the sample data to perform a hypothesis test for the mean of the population from which the data were obtained. In each case, consult the graphs provided to decide whether to use the z-test, the t-test, or neither. Explain your answer.*

34. The normal probability plot and histogram of the data are depicted in Fig. 9.45; σ is known.

35. The normal probability plot and stem-and-leaf diagram of the data are depicted in Fig. 9.46; σ is unknown.

***36.** Refer to Problems 34 and 35.
a. In each case, consult the appropriate graphs to decide whether using the Wilcoxon signed-rank test is reasonable for performing a hypothesis test for the mean of the population from which the data were obtained. Give reasons for your answers.

FIGURE 9.45
Normal probability plot and histogram for Problem 34

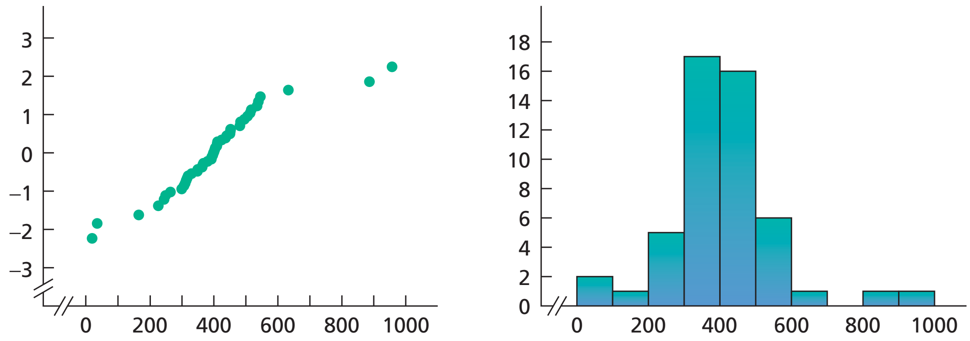
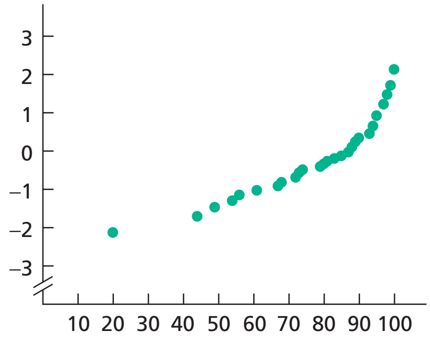


FIGURE 9.46
Normal probability plot and stem-and-leaf diagram for Problem 35



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

b. For each case where using either the z-test or the t-test is reasonable and where using the Wilcoxon signed-rank test is also appropriate, decide which test is preferable. Give reasons for your answers.

***37. Nursing-Home Costs.** The cost of staying in a nursing home in the United States is rising dramatically, as reported in the August 5, 2003, issue of *The Wall Street Journal*. In May 2002, the average cost of a private room in a nursing home was \$168 per day. For August 2003, a random sample of 11 nursing homes yielded the following daily costs, in dollars, for a private room in a nursing home.

73	199	192	181	182	250
159	182	208	129	282	

a. Apply the t-test to decide at the 10% significance level whether the average cost for a private room in a nursing home in August 2003 exceeded that in May 2002.
b. Repeat part (a) by using the Wilcoxon signed-rank test.
c. Obtain a normal probability plot, a boxplot, a stem-and-leaf diagram, and a histogram of the sample data.
d. Discuss the discrepancy in results between the t-test and the Wilcoxon signed-rank test.

Working with Large Data Sets

38. Beef Consumption. According to *Food Consumption, Prices, and Expenditures*, published by the U.S. Department of Agriculture, the mean consumption of beef per person in 2002 was 64.5 lb (boneless, trimmed weight). A sample of 40 people taken this year yielded the data, in pounds, on last year's beef consumption given on the WeissStats CD. Use the technology of your choice to do the following.

a. Obtain a normal probability plot, a boxplot, a histogram, and a stem-and-leaf diagram of the data on beef consumptions.
b. Decide, at the 5% significance level, whether last year's mean beef consumption is less than the 2002 mean of 64.5 lb. Apply the one-mean t-test.
c. The sample data contain four potential outliers: 0, 0, 8, and 20. Remove those four observations, repeat the hypothesis test in part (b), and compare your result with that obtained in part (b).
d. Assuming that the four potential outliers are not recording errors, comment on the advisability of removing them from the sample data before performing the hypothesis test.
e. What action would you take regarding this hypothesis test?

***39. Beef Consumption.** Use the technology of your choice to do the following.
a. Repeat parts (b) and (c) of Problem 38 by using the Wilcoxon signed-rank test.
b. Compare your results from part (a) with those in Problem 38.
c. Discuss the reasonableness of using the Wilcoxon signed-rank test here.

40. Body Mass Index. Body mass index (BMI) is a measure of body fat based on height and weight. According to *Dietary Guidelines for Americans*, published by the U.S. Department of Agriculture and the U.S. Department of Health and Human Services, for adults, a BMI of greater than 25 indicates an above healthy weight (i.e., overweight or obese). The BMIs of 75 randomly selected U.S. adults provided the data on the WeissStats CD. Use the technology of your choice to do the following.
a. Obtain a normal probability plot, a boxplot, and a histogram of the data.
b. Based on your graphs from part (a), is it reasonable to apply the one-mean z-test to the data? Explain your answer.
c. At the 5% significance level, do the data provide sufficient evidence to conclude that the average U.S. adult has an

above healthy weight? Apply the one-mean z-test, assuming a standard deviation of 5.0 for the BMIs of all U.S. adults.

41. Beer Drinking. According to the *Beer Institute Annual Report*, the mean annual consumption of beer per person in the United States is 30.4 gallons (roughly 324 twelve-ounce bottles). A random sample of 300 Missouri residents yielded the annual beer consumptions provided on the WeissStats CD. Use the technology of your choice to do the following.

a. Obtain a histogram of the data.

b. Does your histogram in part (a) indicate any outliers?

c. At the 1% significance level, do the data provide sufficient evidence to conclude that the mean annual consumption of beer per person in Missouri differs from the national mean? (*Note:* See the third bulleted item in Key Fact 9.7 on page 379.)

FOCUSING ON DATA ANALYSIS

UWEC UNDERGRADUATES

Recall from Chapter 1 (see pages 30–31) that the Focus database and Focus sample contain information on the undergraduate students at the University of Wisconsin—Eau Claire (UWEC). Now would be a good time for you to review the discussion about these data sets.

According to *ACT High School Profile Report*, published by ACT, Inc., the national means for ACT composite, English, and math scores are 21.1, 20.6, and 21.0, respectively. You will use these national means in the following problems.

a. Apply the one-mean t-test to the ACT composite score data in the Focus sample (FocusSample) to decide, at the 5% significance level, whether the mean ACT composite score of UWEC undergraduates exceeds the national mean of 21.1 points. Interpret your result.

b. In practice, the population mean of the variable under consideration is unknown. However, in this case, we actually do have the population data, namely, in the Focus database (Focus). If your statistical software package will accommodate the entire Focus database, open that worksheet and then obtain the mean ACT composite score of all UWEC undergraduate students. (*Answer:* 23.6)

c. Was the decision concerning the hypothesis test in part (a) correct? Would it necessarily have to be? Explain your answers.

d. Repeat parts (a)–(c) for ACT English scores. (*Note:* The mean ACT English score of all UWEC undergraduate students is 23.0.)

e. Repeat parts (a)–(c) for ACT math scores. (*Note:* The mean ACT math score of all UWEC undergraduate students is 23.5.)

CASE STUDY DISCUSSION

GENDER AND SENSE OF DIRECTION

At the beginning of this chapter, we discussed research by J. Sholl et al. on the relationship between gender and sense of direction. Recall that, in their study, the spatial orientation skills of 30 male and 30 female students were challenged in a wooded park near the Boston College campus in Newton, Massachusetts. The participants were asked to rate their own sense of direction as either good or poor.

In the park, students were instructed to point to predesignated landmarks and also to the direction of south. For the female students who had rated their sense of direction to be good, the table on page 359 provides the pointing errors (in degrees) when they attempted to point south.

a. If, on average, women who consider themselves to have a good sense of direction do no better than they would by just randomly guessing at the direction of south, what would their mean pointing error be?

b. At the 1% significance level, do the data provide sufficient evidence to conclude that women who consider themselves to have a good sense of direction really do better, on average, than they would by just randomly guessing at the direction of south? Use a one-mean t-test.

c. Obtain a normal probability plot, boxplot, and stem-and-leaf diagram of the data. Based on these plots, is use of the t-test reasonable? Explain your answer.

d. Use the technology of your choice to perform the data analyses in parts (b) and (c).

***e.** Solve part (b) by using the Wilcoxon signed-rank test.

***f.** Based on the plots you obtained in part (c), is use of the Wilcoxon signed-rank test reasonable? Explain your answer.

***g.** Use the technology of your choice to perform the required Wilcoxon signed-rank test of part (e).

BIOGRAPHY

JERZY NEYMAN: A PRINCIPAL FOUNDER OF MODERN STATISTICAL THEORY

Jerzy Neyman was born on April 16, 1894, in Bendery, Russia. His father, Czeslaw, was a member of the Polish nobility, a lawyer, a judge, and an amateur archaeologist. Because Russian authorities prohibited the family from living in Poland, Jerzy Neyman grew up in various cities in Russia. He entered the university in Kharkov in 1912. At Kharkov he was at first interested in physics, but, because of his clumsiness in the laboratory, he decided to pursue mathematics.

After World War I, when Russia was at war with Poland over borders, Neyman was jailed as an enemy alien. In 1921, as a result of a prisoner exchange, he went to Poland for the first time. In 1924, he received his doctorate from the University of Warsaw. Between 1924 and 1934, Neyman worked with Karl Pearson (see Biography in Chapter 13) and his son Egon Pearson and held a position at the University of Kraków. In 1934, Neyman took a position in Karl Pearson's statistical laboratory at University College in London. He stayed in England, where he worked with Egon Pearson until 1938, at which time he accepted an offer to join the faculty at the University of California at Berkeley.

When the United States entered World War II, Neyman set aside development of a statistics program and did war work. After the war ended, Neyman organized a symposium to celebrate its end and "the return to theoretical research." That symposium, held in August 1945, and succeeding ones, held every 5 years until 1970, were instrumental in establishing Berkeley as a preeminent statistical center.

Neyman was a principal founder of the theory of modern statistics. His work on hypothesis testing, confidence intervals, and survey sampling transformed both the theory and the practice of statistics. His achievements were acknowledged by the granting of many honors and awards, including election to the U.S. National Academy of Sciences and receiving the Guy Medal in Gold of the Royal Statistical Society and the U.S. National Medal of Science.

Neyman remained active until his death of heart failure on August 5, 1981, at the age of 87, in Oakland, California.