Hypothesis Tests Regarding a Parameter

CHAPTER 10

Outline

DECISIONS

Many of the products we buy have labels that indicate the net weight of the contents. For example, a candy bar wrapper might state that the net weight of the candy bar is 4 ounces. Should we believe the manufacturer? See the Decisions project on page 486.

Putting It All Together

In Chapter 9, we mentioned that there are two areas of inferential statistics: (1) estimation and (2) hypothesis testing. We have already discussed procedures for estimating the population mean and the population proportion.

We now focus our attention on hypothesis testing. Hypothesis testing is used to test statements regarding a characteristic of one or more populations. In this chapter, we will test hypotheses regarding a single population parameter. The hypotheses that we test regard the population mean and the population proportion.

10.1 The Language of Hypothesis Testing

Preparing for This Section Before getting started, review the following:

- Simple random sampling (Section 1.2, pp. 16–19)
- Sampling distribution of $\bar{x}$ (Section 8.1, pp. 374–388)
- Parameter versus statistic (Section 3.1, p. 107)
- Table 9 (Section 6.2, p. 302)

Objectives

1. **Determine the null and alternative hypotheses**
2. **Understand Type I and Type II errors**
3. **State conclusions to hypothesis tests**

Let's begin with an example that introduces the idea behind hypothesis testing.

EXAMPLE 1 Illustrating Hypothesis Testing

Problem: According to the National Center for Chronic Disease Prevention and Health Promotion, 73.8% of females between the ages of 18 and 29 exercise. Kathleen believes that more women between the ages of 18 and 29 are now exercising, so she obtains a simple random sample of 1000 women and finds that 750 of them are exercising. Is this evidence that the percentage of women between the ages of 18 and 29 who are exercising has increased? What if Kathleen's sample resulted in 920 women exercising?

Approach: Here is the situation Kathleen faces. If 73.8% of 18- to 29-year-old females exercise, she would expect 738 of the 1000 respondents in the sample to exercise. The questions that Kathleen wants to answer are, "How likely is it to obtain a sample of 750 out of 1000 women exercising from a population when the percentage of women who exercise is 73.8%? How likely is a sample that has 920 women exercising?"

Solution: The result of 750 women who exercise is close to what we would expect, so Kathleen is not inclined to believe that the percentage of women exercising has increased. However, the likelihood of obtaining a sample of 920 women who exercise is extremely low if the actual percentage of women who exercise is 73.8%. For the case of obtaining a sample of 920 women who exercise, Kathleen can conclude one of two things: Either the percentage of women who exercise is 73.8% and her sample just happens to include a lot of women who exercise, or the percentage of women who exercise has increased. Provided the sampling was performed in a correct fashion, Kathleen is more inclined to believe that the percentage of women who exercise has increased.

1 Determine the Null and Alternative Hypotheses

Example 1 presents the basic premise behind hypothesis testing: A statement is made regarding the nature of the population, information is collected, and this information is used to test the statement. The steps in conducting a hypothesis test are presented next.

Steps in Hypothesis Testing

1. A statement is made regarding the nature of the population.
2. Evidence (sample data) is collected to test the statement.
3. The data are analyzed to assess the plausibility of the statement.

In this section, we introduce the language of hypothesis testing. Sections 10.2 to 10.4 discuss the formal process of testing a hypothesis.

Definition

A **hypothesis** is a statement regarding a characteristic of one or more populations.

In this chapter, we look at hypotheses regarding a single population parameter. Consider the following example.

(A) According to a Gallup poll conducted in 1995, 74% of Americans felt that men were more aggressive than women. A researcher wonders if the percentage of Americans that feel men are more aggressive than women is different today (a statement regarding a population proportion).

(B) The packaging on a lightbulb states that the bulb will last 500 hours under normal use. A consumer advocate would like to know if the mean lifetime of a bulb is less than 500 hours (a statement regarding the population mean).

(C) The standard deviation of the rate of return for a certain class of mutual funds is 0.08. A mutual fund manager believes the standard deviation of the rate of return for his fund is less than 0.08 (a statement regarding the population standard deviation).

CAUTION

If population data are available, there is no need for inferential statistics.

We test these types of statements using sample data because it is usually impossible or impractical to gain access to the entire population. The procedure (or process) that we use to test these statements is called *hypothesis testing*.

Definition

Hypothesis testing is a procedure, based on sample evidence and probability, used to test statements regarding a characteristic of one or more populations.

Because a statement can either be true or false, hypothesis testing is based on two types of hypotheses.

Definitions

The **null hypothesis**, denoted H_0 (read "H-naught"), is a statement to be tested. The null hypothesis is assumed true until evidence indicates otherwise. In this chapter, it will be a statement regarding the value of a population parameter.

The **alternative hypothesis**, denoted H_1 (read "H-one"), is a statement to be tested. We are trying to find evidence for the alternative hypothesis. In this chapter, it will be a statement regarding the value of a population parameter.

In Other Words

The null hypothesis is a statement of status quo or no difference and always contains a statement of equality. The null hypothesis is assumed to be true until we have evidence to the contrary. We seek evidence for the statement in the alternative hypothesis.

For the lightbulb manufacturer in Situation B, the consumer advocate wishes to know whether the mean lifetime of the bulb is less than 500 hours. Because we are trying to obtain evidence for this statement, it is expressed as the alternative hypothesis using the notation H_1: $\mu < 500$. The statement made by the manufacturer is that the bulb lasts 500 hours on average. We give the manufacturer the benefit of the doubt and assume this statement to be true, so this becomes the statement to be tested. We express the statement to be tested using the notation H_0: $\mu = 500$.

In this chapter, there are three ways to set up the null and alternative hypotheses.

1. Equal hypothesis versus not equal hypothesis **(two-tailed test)**

   H_0: parameter $=$ some value

   H_1: parameter $\neq$ some value

2. Equal versus less than **(left-tailed test)**

   H_0: parameter $=$ some value

   H_1: parameter $<$ some value

3. Equal versus greater than **(right-tailed test)**

   H_0: parameter $=$ some value

   H_1: parameter $>$ some value

Left- and right-tailed tests are referred to as **one-tailed tests**. Notice that in the left-tailed test the direction of the inequality sign in the alternative hypothesis points to the left ($<$), while in the right-tailed test the direction of the inequality sign in the alternative hypothesis points to the right ($>$). In all three tests the null hypothesis contains a statement of equality. The statement of equality comes from existing information.

Refer to the three claims made on page 455. In Situation A, the null hypothesis is H_0: $p = 0.74$. This is a statement of *status quo* or no difference. The Latin phrase *status quo* means "the existing state or condition." It means that American opinions have not changed from 1995. In Situation B, the null hypothesis is H_0: $\mu = 500$. This is a statement of no difference between the population mean and the lifetime stated on the label. In Situation C, the null hypothesis is H_0: $\sigma = 0.08$. This is a statement of no difference between the population standard deviation rate of return of the manager's mutual fund and all mutual funds.

The statement we are trying to gather evidence for, which is dictated by the researcher before any data are collected, determines the structure of the alternative hypothesis (two-tailed, left-tailed, or right-tailed). For example, the label on a can of soda states that the can contains 12 ounces of liquid. A consumer advocate would be concerned only if the mean contents are less than 12 ounces, so the alternative hypothesis is H_1: $\mu < 12$. However, a quality-control engineer for the soda manufacturer would be concerned if there is too little or too much soda in the can, so the alternative hypothesis would be H_1: $\mu \neq 12$. In both cases, however, the null hypothesis is a statement of no difference between the manufacturer's assertion on the label and the actual mean contents of the can. So the null hypothesis is H_0: $\mu = 12$.

EXAMPLE 2 Forming Hypotheses

Problem: Determine the null and alternative hypotheses. State whether the test is two-tailed, left-tailed, or right-tailed.

(a) The Medco pharmaceutical company has just developed a new antibiotic for children. Among the competing antibiotics, 2% of children who take the drug experience headaches as a side effect. A researcher for the Food and Drug Administration wishes to know if the percentage of children taking the new antibiotic who experience headaches as a side effect is more than 2%.

(b) The *Blue Book* value of a used 3-year-old Chevy Corvette is \$37,500. Grant wonders if the mean price of a used 3-year-old Chevy Corvette in the Miami metropolitan area is different from \$37,500.

(c) The standard deviation of the contents in a 64-ounce bottle of detergent using an old filling machine was known to be 0.23 ounce. The company purchased a new filling machine and wants to know if the standard deviation for the new filling machine is less than 0.23 ounce.

Approach: In each case, we must determine the parameter to be tested. We then determine the statement of no change or no difference (status quo), and finally we identify the statement we are attempting to gather evidence for.

> **In Other Words**
> Structuring the null and alternative hypothesis:
> 1. Identify the parameter to be tested.
> 2. Determine the status quo value of the parameter.
> 3. Determine the statement that reflects what we are trying to gather evidence for.

Solution

(a) The hypothesis deals with a population proportion, p. If the new drug is no different from current drugs on the market, the proportion of individuals taking the new drug who experience a headache will be 0.02, so the null hypothesis is H_0: $p = 0.02$. We are trying to show that the proportion of individuals who experience a headache is "more than" 0.02. Therefore, the alternative hypothesis is H_1: $p > 0.02$. This is a right-tailed test because the alternative hypothesis contains a $>$ symbol.

(b) The hypothesis deals with a population mean, μ. If the mean price of a 3-year-old Corvette in Grant's neighborhood is no different from the *Blue Book* price, the population mean in Grant's neighborhood will be \$37,500, so the null hypothesis is H_0: $\mu = 37{,}500$. Grant wishes to determine if the mean price is different from \$37,500, so that the alternative hypothesis is H_1: $\mu \neq 37{,}500$. This is a two-tailed test because the alternative hypothesis contains a $\neq$ symbol.

(c) The hypothesis deals with a population standard deviation, σ. If the new machine is no different from the old machine, the standard deviation of the amount in the bottles filled by the new machine will be 0.23 ounce, so the null hypothesis is H_0: $\sigma = 0.23$. The phrase *less than* is represented symbolically by $<$, so the alternative hypothesis is H_1: $\sigma < 0.23$. This is a left-tailed test because the alternative hypothesis contains a $<$ symbol.

In Other Words

Look for key phrases when forming the alternative hypothesis. For example, more than means $>$; different from means $\neq$; less than means $<$; and so on. See Table 9 on page 302 for a list of key phrases and the symbols they translate into.

Now Work Problem 17(a).

2 Understand Type I and Type II Errors

As stated earlier, we use sample data to determine whether to reject or not reject the null hypothesis. Because the decision to reject or not reject the null hypothesis is based on incomplete (sample) information, there is always the possibility of making an incorrect decision. In fact, there are four possible outcomes from hypothesis testing.

In Other Words

When you are testing a hypothesis, there is always the possibility that your conclusion will be wrong. To make matters worse, you won't know whether you are wrong or not! Don't fret, however; we have tools to help manage these incorrect conclusions.

Four Outcomes from Hypothesis Testing

1. We reject the null hypothesis when the alternative hypothesis is true. This decision would be correct.
2. We do not reject the null hypothesis when the null hypothesis is true. This decision would be correct.
3. We reject the null hypothesis when the null hypothesis is true. This decision would be incorrect. This type of error is called a **Type I error**.
4. We do not reject the null hypothesis when the alternative hypothesis is true. This decision would be incorrect. This type of error is called a **Type II error**.

Figure 1 illustrates the two types of errors that can be made in hypothesis testing.

Figure 1

		Reality	
		H_0 Is True	H_1 Is True
Conclusion	Do Not Reject H_0	Correct Conclusion	Type II Error
	Reject H_0	Type I Error	Correct Conclusion

We illustrate the idea of Type I and Type II errors by looking at hypothesis testing from the point of view of a criminal trial. In any trial, the defendant is assumed to be innocent. (We give the defendant the benefit of the doubt.) The district attorney must present evidence proving that the defendant is guilty.

Because we are seeking evidence for guilt, it becomes the alternative hypothesis. Innocence is assumed, so this is the null hypothesis. The hypotheses for a trial are written

$$H_0\text{: the defendant is innocent}$$
$$H_1\text{: the defendant is guilty}$$

The trial is the process whereby information (sample data) is obtained. The jury then deliberates about the evidence (the data analysis). Finally, the jury either convicts the defendant (rejects the null hypothesis) or declares the defendant not guilty (fails to reject the null hypothesis).

Note that the defendant is never declared innocent. That is, we never conclude that the null hypothesis is true. Using this analogy, the two correct decisions are to conclude that an innocent person is not guilty or conclude that a guilty person is guilty. The two incorrect decisions are to convict an innocent person (a Type I error) or to let a guilty person go free (a Type II error). It is helpful to think in this way when trying to remember the difference between a Type I and a Type II error.

In Other Words

A Type I error is like putting an innocent person in jail. A Type II error is like letting a guilty person go free.

EXAMPLE 3 Type I and Type II Errors

Problem: The Medco pharmaceutical company has just developed a new antibiotic. Among the competing antibiotics, 2% of children who take the drug experience headaches as a side effect. A researcher for the Food and Drug Administration wishes to know if the percentage of children taking the new antibiotic who experience a headache as a side effect is more than 2%. The researcher conducts a hypothesis test with H_0: $p = 0.02$ and H_1: $p > 0.02$. Provide statements explaining what it would mean to make (a) a Type I error and (b) a Type II error.

Approach: A Type I error occurs if the null hypothesis is rejected when, in fact, the null hypothesis is true. A Type II error occurs if the null hypothesis is not rejected when, in fact, the alternative hypothesis is true.

Solution

(a) A Type I error is made if the sample evidence leads the researcher to believe that $p > 0.02$ (that is, we reject the null hypothesis) when, in fact, the proportion of children who experience a headache is not greater than 0.02.

(b) A Type II error is made if the researcher does not reject the null hypothesis that the proportion of children experiencing a headache is equal to 0.02 when, in fact, the proportion of children who experience a headache is more than 0.02. For example, the sample evidence led the researcher to believe $p = 0.02$ when in fact the true proportion is some value larger than 0.2.

Now Work Problems 17(b) and 17(c).

Understand the Probability of Making a Type I or Type II Error

Recall that we never know whether a confidence interval contains the unknown parameter. We only know the likelihood that a confidence interval captures the parameter. Similarly, we never know whether or not the outcome of a hypothesis test results in an error. However, just as we place a level of confidence in the construction of a confidence interval, we can determine the probability of making errors. The following notation is commonplace:

$$\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$$

$$\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true})$$

The symbol β is the Greek letter beta (pronounced "BAY tah"). The probability of making a Type I error, α, is chosen by the researcher *before* the sample data are collected. This probability is referred to as the *level of significance*.

Definition

The **level of significance**, α, is the probability of making a Type I error.

The choice of the level of significance depends on the consequences of making a Type I error. If the consequences are severe, the level of significance should be small (say, $\alpha = 0.01$). However, if the consequences of making a Type I error are not severe, a higher level of significance can be chosen (say $\alpha = 0.05$ or $\alpha = 0.10$).

Why is the level of significance not always set at $\alpha = 0.01$? By reducing the probability of making a Type I error, you increase the probability of making a Type II error, β. Using our court analogy, a jury is instructed that the prosecution must provide proof of guilt "beyond all reasonable doubt." This implies that we are choosing to make α small so the probability that we will send an innocent person to jail is very small. The consequence of the small α, however, is a large β, which means many guilty defendants will go free. For now, we are content to recognize the inverse relation between α and β (as one goes up the other goes down).

In Other Words

As the probability of a Type I error increases, the probability of a Type II error decreases, and vice versa.

3 State Conclusions to Hypothesis Tests

Once the decision to reject or not reject the null hypothesis is made, the researcher must state his or her conclusion. It is important to recognize that we never *accept* the null hypothesis. Again, the court system analogy helps to illustrate the idea. The null hypothesis is H_0: innocent. When the evidence presented to the jury is not enough to convict beyond all reasonable doubt, the jury comes back with a verdict of "not guilty."

Notice that the verdict does not state that the null hypothesis of innocence is true; it states that there is not enough evidence to conclude guilt. This is a huge difference. Being told that you are not guilty is very different from being told that you are innocent!

So sample evidence can never prove the null hypothesis to be true. When we do not reject the null hypothesis, we are saying that the evidence indicates that the null hypothesis *could* be true.

CAUTION

We never *accept* the null hypothesis, because, without having access to the entire population, we don't know the exact value of the parameter stated in the null. Rather, we say that we do not reject the null hypothesis. This is just like the court system. We never declare a defendant innocent, but rather say the defendant is not guilty.

EXAMPLE 4 Stating the Conclusion

Problem: The Medco pharmaceutical company has just developed a new antibiotic. Among the competing antibiotics, 2% of children who take the drug experience a headache as a side effect. A researcher for the Food and Drug Administration believes that the proportion of children taking the new antibiotic who experience a headache as a side effect is more than 0.02. From Example 2(a), we know the null hypothesis is H_0: $p = 0.02$ and the alternative hypothesis is H_1: $p > 0.02$.

(a) Suppose the sample evidence indicates that the null hypothesis is rejected. State the conclusion.

(b) Suppose the sample evidence indicates that the null hypothesis is not rejected. State the conclusion.

Approach: When the null hypothesis is rejected, we say that there is sufficient evidence to support the statement in the alternative hypothesis. When the null hypothesis is not rejected, we say that there is not sufficient evidence to support the statement in the alternative hypothesis. We never say that the null hypothesis is true!

Solution

(a) The statement in the alternative hypothesis is that the proportion of children taking the new antibiotic who experience a headache as a side effect is more than 0.02. Because the null hypothesis $(p = 0.02)$ is rejected, we conclude there is sufficient evidence to conclude that the proportion of children who experience a headache as a side effect is more than 0.02.

(b) Because the null hypothesis is not rejected, we conclude that there is not sufficient evidence to say that the proportion of children who experience a headache as a side effect is more than 0.02.

Now Work Problem 25.

10.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. Explain what it means to make a Type I error. Explain what it means to make a Type II error.
2. Suppose the consequences of making a Type I error are severe. Would you choose the level of significance, α, to equal 0.01, 0.05, or 0.10? Why?
3. What happens to the probability of making a Type II error, β, as the level of significance, α, decreases? Why is this result intuitive?
4. If a hypothesis is tested at the $\alpha = 0.05$ level of significance, what is the probability of making a Type I error?
5. The following is a quotation from Sir Ronald A. Fisher, a famous statistician.

   > "For the logical fallacy of believing that a hypothesis has been proved true, merely because it is not contradicted by the available facts, has no more right to insinuate itself in statistics than in other kinds of scientific reasoning .... It would, therefore, add greatly to the clarity with which the tests of significance are regarded if it were generally understood that tests of significance, when used accurately, are capable of rejecting or invalidating hypotheses, in so far as they are contradicted by the data: but that they are never capable of establishing them as certainly true ...."

   In your own words, explain what this quotation means.
6. In your own words, explain the difference between "beyond all reasonable doubt" and "beyond all doubt."
7. *True or False:* Sample evidence can prove that a null hypothesis is true.
8. *True or False:* Type I and Type II errors are independent events.

Skill Building

In Problems 9–14, a null and alternative hypothesis is given. Determine whether the hypothesis test is left-tailed, right-tailed, or two-tailed. What parameter is being tested?

9. $H_0: \mu = 5$
$H_1: \mu > 5$

10. $H_0: p = 0.2$
$H_1: p < 0.2$

11. $H_0: \sigma = 4.2$
$H_1: \sigma \neq 4.2$

12. $H_0: p = 0.76$
$H_1: p > 0.76$

13. $H_0: \mu = 120$
$H_1: \mu < 120$

14. $H_0: \sigma = 7.8$
$H_1: \sigma \neq 7.8$

In Problems 15–22, (a) determine the null and alternative hypotheses, (b) explain what it would mean to make a Type I error, and (c) explain what it would mean to make a Type II error.

15. Teenage Mothers According to the U.S. Census Bureau, 11.8% of registered births in the United States in 2000 were to teenage mothers. A sociologist believes that this percentage has decreased since then.

16. Charitable Contributions According to *Giving and Volunteering in the United States, 2001 Edition*, the mean charitable contribution per household in the United States in 2000 was $1623. A researcher believes that the level of giving has changed since then.

17. Single-Family Home Price NW According to the Federal Housing Finance Board, the mean price of a single-family home in October 2003 was $243,756. A real estate broker believes that because of changes in interest rates, as well as other economic factors, the mean price has increased since then.

18. Fair Packaging and Labeling Federal law requires that a jar of peanut butter that is labeled as containing 32 ounces must contain at least 32 ounces. A consumer advocate feels that a certain peanut butter manufacturer is shorting customers by underfilling the jars so that the mean content is less than the 32 ounces stated on the label.

19. Valve Pressure The standard deviation in the pressure required to open a certain sprung-type valve is known to be $\sigma = 0.7$ psi. Due to changes in the manufacturing process, the quality-control manager feels that the pressure variability has been reduced.

20. Overweight According to the Centers for Disease Control and Prevention, 16% of children aged 6 to 11 years are overweight. A school nurse thinks that the percentage of 6- to 11-year-olds who are overweight is higher in her school district.

21. Cell Phone Service According to the *Statistical Abstract of the United States,* the mean monthly cell phone bill was \$49.91 in 2003. A researcher suspects that the mean monthly cell phone bill is different today.

22. SAT Math Scores In 2004, the standard deviation SAT math score for all students taking the exam was 114. A teacher believes that, due to changes to the SAT Reasoning Test in 2005, the standard deviation of SAT math scores will increase.

In Problems 23–34, state the conclusion based on the results of the test.

23. For the hypotheses in Problem 15, suppose the null hypothesis is rejected.

24. For the hypotheses in Problem 16, suppose the null hypothesis is not rejected.

25. NW For the hypotheses in Problem 17, suppose the null hypothesis is not rejected.

26. For the hypotheses in Problem 18, suppose the null hypothesis is rejected.

27. For the hypotheses in Problem 19, suppose the null hypothesis is not rejected.

28. For the hypotheses in Problem 20, suppose the null hypothesis is not rejected.

29. For the hypotheses in Problem 21, suppose the null hypothesis is rejected.

30. For the hypotheses in Problem 22, suppose the null hypothesis is not rejected.

31. For the hypotheses in Problem 15, suppose the null hypothesis is not rejected.

32. For the hypotheses in Problem 16, suppose the null hypothesis is rejected.

33. For the hypotheses in Problem 17, suppose the null hypothesis is rejected.

34. For the hypotheses in Problem 18, suppose the null hypothesis is not rejected.

Applying the Concepts

35. Fruits and Vegetables According to the *Statistical Abstract of the United States*, the mean consumption of fruits in 2003 was 98.4 pounds. A dietician believes more people are becoming health conscious and that fruit consumption has risen since then.

(a) Determine the null and alternative hypotheses.
(b) Suppose sample data indicate that the null hypothesis should be rejected. State the conclusion of the researcher.
(c) Suppose, in fact, the mean consumption of fruits is 98.4 pounds. Was a Type I or Type II error committed? If we tested this hypothesis at the $\alpha = 0.05$ level of significance, what is the probability of committing a Type I error?

36. Test Preparation The mean score on the SAT Math Reasoning exam is 518. A test preparation company claims that the mean scores of students who take its course are higher than the mean of 518.

(a) Determine the null and alternative hypotheses.
(b) Suppose sample data indicate that the null hypothesis should not be rejected. State the conclusion of the company.
(c) Suppose, in fact, the mean score of students taking the preparatory course is 522. Was a Type I or Type II error committed? If we tested this hypothesis at the $\alpha = 0.01$ level, what is the probability of committing a Type I error?
(d) If we wanted to decrease the probability of making a Type II error, would we need to increase or decrease the level of significance?

37. Marijuana Use According to the Centers for Disease Control and Prevention, in 2001, 10.2% of high school students had tried marijuana for the first time before the age of 13. The Drug Abuse and Resistance Education (DARE) program underwent several major changes to keep up with technology and issues facing students in the 21st century. After the changes, a school resource officer (SRO) thinks that the proportion of high school students who have tried marijuana for the first time before the age of 13 has decreased from the 2001 level.

(a) Determine the null and alternative hypotheses.
(b) Suppose sample data indicate that the null hypothesis should not be rejected. State the conclusion of the SRO.
(c) Suppose, in fact, the proportion of high school students who have tried marijuana for the first time before the age of 13 was 9.5%. Was a Type I or Type II error committed?

38. Internet Use According to the *Statistical Abstract of the United States*, in 2000, 10.7% of Americans over 65 years of age used the Internet. A researcher believes the proportion of Americans over 65 years of age who use the Internet is higher than 10.7% today.

(a) Determine the null and alternative hypotheses.

(b) Suppose sample data indicate that the null hypothesis should be rejected. State the conclusion of the researcher.

(c) Suppose, in fact, the percentage of Americans over 65 years of age who use the Internet is still 10.7%. Was a Type I or Type II error committed?

39. Consumer Reports The following is an excerpt from a *Consumer Reports* article from February 2001.

> The Platinum Gasaver makes some impressive claims. The device, \$188 for two, is guaranteed to increase gas mileage by 22% says the manufacturer, National Fuelsaver. Also, the company quotes "the government" as concluding, "Independent testing shows greater fuel savings with *Gasaver* than the 22 percent claimed by the developer." Readers have told us they want to know more about it.
>
> The Environmental Protection Agency (EPA), after its lab tests of the Platinum Gasaver, concluded in 1991, "Users of the device would not be expected to realize either an emission or fuel economy benefit." The Federal Trade Commission says, "No government agency endorses gas-saving products for cars."

Determine the null and alternative hypotheses that the EPA used in to draw the conclusion stated in the second paragraph.

40. Prolong Engine Treatment The manufacturer of Prolong Engine Treatment claims that if you add one 12-ounce bottle of its \$20 product, your engine will be protected from excessive wear. An infomercial claims that a woman drove 4 hours without oil, thanks to Prolong. *Consumer Reports* magazine tested engines in which they added Prolong to the motor oil, ran the engines, drained the oil, and then determined the time until the engines seized.

(a) Determine the null and alternative hypotheses *Consumer Reports* will test.

(b) Both engines took exactly 13 minutes to seize. What conclusion might *Consumer Reports* draw based on this evidence?

41. Refer to the claim made in Problem 18. Researchers must choose the level of significance based on the consequences of making a Type I error. In your opinion, is a Type I error or Type II error more serious? Why? On the basis of your answer, decide on a level of significance, α. Be sure to support your opinion.

10.2 Hypothesis Tests for a Population Mean Assuming the Population Standard Deviation Is Known

Preparing for This Section Before getting started, review the following:

- Using probabilities to identify unusual events (Section 5.1, p. 225)
- z_α Notation (Section 7.2, p. 340)
- Sampling distribution of $\bar{x}$ (Section 8.1, pp. 374–388)
- Computing normal probabilities (Section 7.3, pp. 345–349)

Objectives

1. **Understand the logic of hypothesis testing**
2. **Test hypotheses about a population mean with σ known using the classical approach**
3. **Test hypotheses about a population mean with σ known using *P*-values**
4. **Test hypotheses about a population mean with σ known using confidence intervals**
5. **Understand the difference between statistical significance and practical significance**

1 Understand the Logic of Hypothesis Testing

Now that we know the language of hypothesis testing, we are ready to present the methods for conducting a hypothesis test.

In this section, we present three approaches to testing hypotheses about a population mean, μ. As we did for confidence intervals, we begin by assuming that we know the value of the population standard deviation, σ. The assumption is made because it allows us to use the normal model to test hypotheses regarding the population mean. Because the normal model is fairly easy to use, we can concentrate on the techniques of hypothesis testing without getting bogged down with other details. The assumption that σ is known will be dropped in the next section.

Historical Note

Jerzy Neyman was born on April 16, 1894, in Bendery, Russia. In 1921, he moved to Poland. He received his Ph.D. from the University of Warsaw in 1924. He read some of Karl Pearson's works and became interested in statistics; however, Neyman was not impressed with Pearson's mathematical abilities. In 1927, he met Pearson's son, Egon Pearson, who was working on a formal approach to hypothesis testing. It was Neyman who provided the mathematical rigor to their work. Together, they developed the phrases *null hypothesis* and *alternative hypothesis*. In 1938, Neyman joined the faculty at the University of California at Berkeley. He died on August 5, 1981.

To test hypotheses regarding the population mean assuming the population standard deviation is known, two requirements must be satisfied.

- A simple random sample is obtained.
- The population from which the sample is drawn is normally distributed or the sample size is large ($n \geq 30$).

If these requirements are met, then the distribution of $\bar{x}$ is normal with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

The first method that we use in testing hypotheses regarding a population mean is referred to as the classical (traditional) approach, the second method is the *P*-value approach, and the third method uses confidence intervals. Your instructor may choose to cover one, two, or all three approaches to hypothesis testing. Do not be alarmed if one or two of the approaches are not covered by your instructor.

Let's lay out a scenario that will be used to help understand both the classical approach to hypothesis testing and the *P*-value approach. Suppose a consumer advocate is concerned that a manufacturer of potato chips is underfilling its bags. The bag states the contents weigh 12.5 ounces. In hypothesis testing we assume that the manufacturer is "not guilty," which means we assume that the population mean contents of the bags of chips is $\mu = 12.5$ ounces. We are looking for evidence that shows the manufacturer is underfilling the bags. We have the following hypotheses:

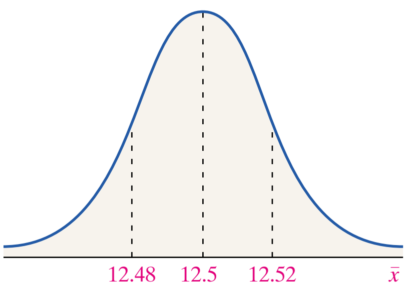
$$H_0: \mu = 12.5 \quad \text{versus} \quad H_1: \mu < 12.5$$

Suppose the consumer advocate gathers evidence by obtaining a simple random sample of $n = 36$ bags of chips, weighing the contents, and obtaining a sample mean of 12.45 ounces. Does this sample suggest that the manufacturer is underfilling its bags? What is convincing or *statistically significant* evidence?

Definition

When observed results are unlikely under the assumption that the null hypothesis is true, we say the result is **statistically significant**. When results are found to be statistically significant, we reject the null hypothesis.

Figure 2



Before we can test the hypothesis, we need to know the distribution of the sample mean, since a different sample of 36 bags of chips will likely result in a different sample mean. Since the sample size is large, the Central Limit Theorem says that the shape of the distribution of the sample mean is approximately normal. Regardless of the size of the sample, the mean of the distribution of the sample mean is $\mu_{\bar{x}} = \mu = 12.5$ ounces because we assume the statement in the null hypothesis to be true, until we have evidence to the contrary. Suppose the population standard deviation is known to be 0.12 ounce; then the standard deviation of the distribution of the sample mean is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.12}{\sqrt{36}} = 0.02$ ounce. Figure 2 shows the sampling distribution of the sample mean for our potato chip example.

Now that we have a model that describes the distribution of the sample mean, we can look at two approaches to testing if the potato chip company is underfilling the bags.

The Logic of the Classical Approach

One criterion we may use for sufficient evidence is to reject the null hypothesis if the sample mean is too many standard deviations below the hypothesized (or status quo) population mean of 12.5 ounces. For example, our criterion might be to reject the null hypothesis if the sample mean is more than 2 standard deviations below the assumed mean of 12.5 ounces.

CAUTION
We always test hypotheses assuming that the null hypothesis is true.

Recall that $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ represents the number of standard deviations that $\bar{x}$ is from the population mean, μ. Our simple random sample of 36 bags results in sample mean weight of $\bar{x} = 12.45$ ounces, so under the assumption that the null hypothesis is true, we have

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{12.45 - 12.5}{0.12/\sqrt{36}} = -2.5$$

The sample mean is 2.5 standard deviations below the hypothesized mean. Because the sample mean is more than 2 standard deviations (that is, "too far") below the hypothesized population mean, we will reject the null hypothesis and conclude that there is sufficient evidence to support the belief that the bag has less than 12.5 ounces of potato chips. This conclusion will lead the consumer advocate to wage a "truth in advertising campaign" against the potato chip manufacturer.

Why does it make sense to reject the null hypothesis if the sample mean is more than two standard deviations away from the hypothesized mean? The area under the standard normal curve to the left of $Z = -2$ is 0.0228, as shown in Figure 3.

Figure 4 shows that, if the null hypothesis is true, 97.72% of all sample means will be 12.46 ounces or more and only 2.28% of the sample means will be less than 12.46 ounces, as indicated by the region shaded blue in Figure 4. Remember, the 12.46 comes from the fact that 12.46 is 2 standard deviations below the hypothesized mean of 12.5 ounces. If a sample mean lies in the green region we are inclined to believe that it came from a population whose mean is less than 12.5, rather than believe that the population mean equals 12.5 and our sample just happened to result in an unusual outcome (a bunch of underfilled bags).

Figure 3

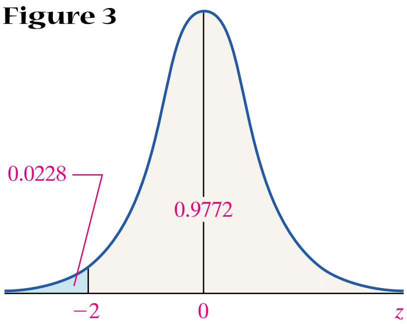
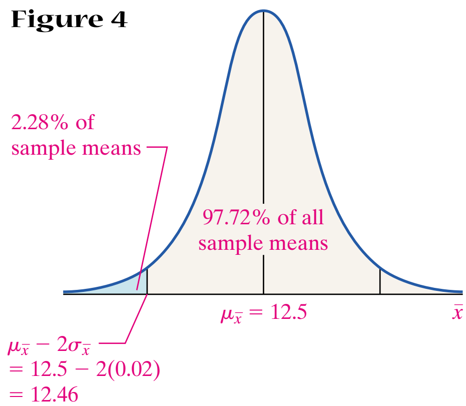


Figure 4



Notice that our criterion for rejecting the null hypothesis will lead to making a Type I error (rejecting a true null hypothesis) 2.28% of the time. That is, the probability of making a Type I error is 2.28%.

The previous discussion leads to the following premise of hypothesis testing using the classical approach:

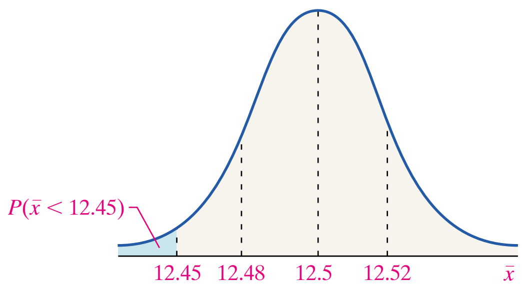
> If the sample mean is too many standard deviations from the mean stated in the null hypothesis, we reject the null hypothesis.

The Logic of the *P*-Value Approach

A second criterion we may use for sufficient evidence to support the belief that the manufacturer is underfilling the bags is to compute how likely it is to obtain a sample mean of 12.45 ounces or less from a population whose mean is assumed to be 12.5 ounces. If a sample mean of 12.45 or less is unlikely (or unusual), we have evidence against the null hypothesis. If the sample mean of 12.45 is not unlikely (not unusual), we do not have sufficient evidence against the null hypothesis.

We can compute the probability of obtaining a sample mean of 12.45 or less from a population whose mean is 12.5 using the normal model. Figure 5 shows the area that represents $P(\bar{x} < 12.45)$.

Figure 5



Because

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{12.45 - 12.5}{0.02} = -2.5$$

we compute

$$P(\bar{x} \leq 12.45) = P(Z \leq -2.5) = 0.0062$$

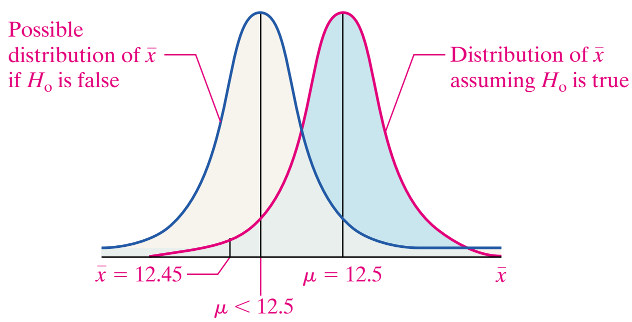
The probability of obtaining a sample mean of 12.45 ounces or less from a population whose mean is 12.5 ounces is 0.0062. This means that less than 1 sample in 100 will give a mean as low or lower than the one we obtained *if* the population mean really is 12.5 ounces. Because these results are so unusual, we take this as evidence against the statement in the null hypothesis.

This discussion leads to the following premise of testing a hypothesis using the *P*-value approach:

> Assuming H_0 is true, if the probability of getting a sample mean as extreme or more extreme than the one obtained is small, we reject the null hypothesis.

Figure 6 further illustrates the situation for both the classical and *P*-value approach. The distribution in red shows the distribution of the sample mean assuming the statement in the null hypothesis is true. The sample mean of 12.45 is too far from the assumed population mean of 12.5. Therefore, we reject the null hypothesis that $\mu = 12.5$ and conclude that the sample came from a population with some population mean less than 12.5 ounces, as indicated by the distribution in blue. We don't know what the population mean weight of the bags is, but we have evidence that it is less than 12.5 ounces.

Figure 6



2 Test Hypotheses about a Population Mean with σ Known Using the Classical Approach

We now formalize the procedure for testing hypotheses regarding the population mean when the population standard deviation, σ, is known using the classical approach.

Testing Hypotheses Regarding the Population Mean with σ Known Using the Classical Approach

To test hypotheses regarding the population mean with σ known, we can use the steps that follow provided the following two requirements are satisfied.

1. The sample is obtained using simple random sampling.
2. The sample has no outliers and the population from which the sample is drawn is normally distributed or the sample size, n, is large ($n \geq 30$).

***Step 1*:** Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two-Tailed	Left-Tailed	Right-Tailed
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$	H_1: $\mu < \mu_0$	H_1: $\mu > \mu_0$

Note: μ_0 is the assumed or status quo value of the population mean.

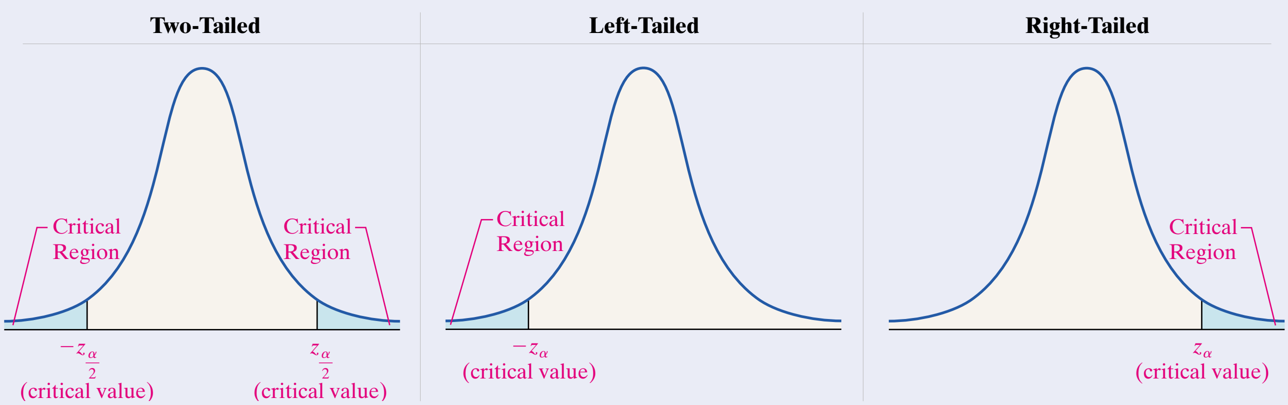
***Step 2*:** Select a level of significance α based on the seriousness of making a Type I error.

***Step 3*:** Provided the population from which the sample is drawn is normal or the sample size is large ($n \geq 30$) and the population standard deviation σ, is known, the distribution of the sample mean, $\bar{x}$, is normal with mean μ_0 and standard deviation $\frac{\sigma}{\sqrt{n}}$. Therefore,

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

represents the number of standard deviations the sample mean is from the assumed mean, μ_0. This value is called the **test statistic**.

***Step 4*:** The level of significance is used to determine the *critical value*. The **critical value** represents the maximum number of standard deviations the sample mean can be from μ_0 before the null hypothesis is rejected. For example, the critical value in the left-tailed test is $-z_\alpha$. The shaded region(s) represents the *critical* (*or rejection*) *region(s)*. The **critical region** or **rejection region** is the set of all values such that the null hypothesis is rejected.



***Step 5*:** Compare the critical value with the test statistic:

Two-Tailed	Left-Tailed	Right-Tailed
If $z_0 < -z_{\frac{\alpha}{2}}$ or $z_0 > z_{\frac{\alpha}{2}}$, reject the null hypothesis.	If $z_0 < -z_\alpha$, reject the null hypothesis.	If $z_0 > z_\alpha$, reject the null hypothesis.

The comparison of the test statistic and critical value is called the **decision rule**.

***Step 6*:** State the conclusion.

The procedure is **robust**, which means that minor departures from normality will not adversely affect the results of the test. However, for small samples, if the data have outliers, the procedure should not be used.

For small samples, we will verify that the data come from a population that is normal by constructing normal probability plots (to assess normality) and boxplots (to determine whether there are outliers). If the normal probability plot indicates that the data do not come from a population that is normally distributed or the boxplot reveals outliers, nonparametric tests, which are not discussed in this text, should be performed.

EXAMPLE 1 The Classical Approach of Hypothesis Testing: Right-Tailed, Large Sample

Problem: According to the U.S. Federal Highway Administration, the mean number of miles driven annually is 12,200. Patricia believes that residents of the state of Montana drive more than the national average. She obtains a simple random sample of 35 drivers from a list of registered drivers in the state of Montana. The mean number of miles driven for the 35 drivers is 12,895.9. Assuming $\sigma = 3800$ miles, does the sample provide sufficient evidence that residents of the state of Montana drive more than the national average at the $\alpha = 0.1$ level of significance?

Approach: Because the sample size is large, we can proceed to Steps 1 through 6.

Solution

Step 1: Patricia wants to know if people in Montana are driving more than 12,200 miles annually. This can be written $\mu > 12{,}200$. This is a right-tailed test and we have

$$H_0: \mu = 12{,}200 \qquad \text{versus} \qquad H_1: \mu > 12{,}200$$

Step 2: The level of significance is $\alpha = 0.1$.

Step 3: Patricia found the sample mean, $\bar{x}$, to be 12,895.9 miles. The test statistic is

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{12{,}895.9 - 12{,}200}{3800/\sqrt{35}} = 1.08$$

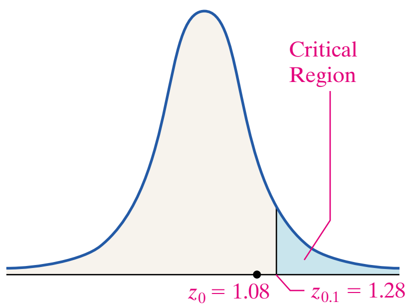
The sample mean of 12,895.9 miles is 1.08 standard deviations above the mean of 12,200.

Step 4: Because Patricia is performing a right-tailed test, we determine the critical value at the $\alpha = 0.1$ level of significance to be $z_{0.1} = 1.28$. The critical region is displayed in Figure 7.

CAUTION

In Example 1, we see that the sample mean, $\bar{x} = 12{,}895.9$, is not far enough from the status quo value of the population mean, 12,200 miles. Therefore, we do not have enough evidence to reject the null hypothesis. However, this does not mean we are accepting the null hypothesis that the mean number of miles driven in Montana is 12,200. We are saying that we don't have enough evidence to say it is greater than 12,200 miles. Be sure you understand the difference between these two comments.

Figure 7



Step 5: Because the test statistic $z_0 = 1.08$ is less than the critical value $z_{0.1} = 1.28$, we do not reject the null hypothesis. That is, the value of the test statistic does not fall within the critical region, so we do not reject H_0. We label the test statistic in Figure 7.

***Step 6*:** There is not sufficient evidence at the $\alpha = 0.1$ level of significance to support the belief that residents of the state of Montana drive more than the national average of 12,200 miles.

Now Work Problem 19 Using the Classical Approach.

Now let's look at a two-tailed hypothesis test.

EXAMPLE 2 The Classical Approach of Hypothesis Testing: Two-Tailed, Small Sample

Table 1

94.25	38.94	79.15	56.78
70.07	115.59	77.56	37.01
55.00	76.05	27.29	52.48

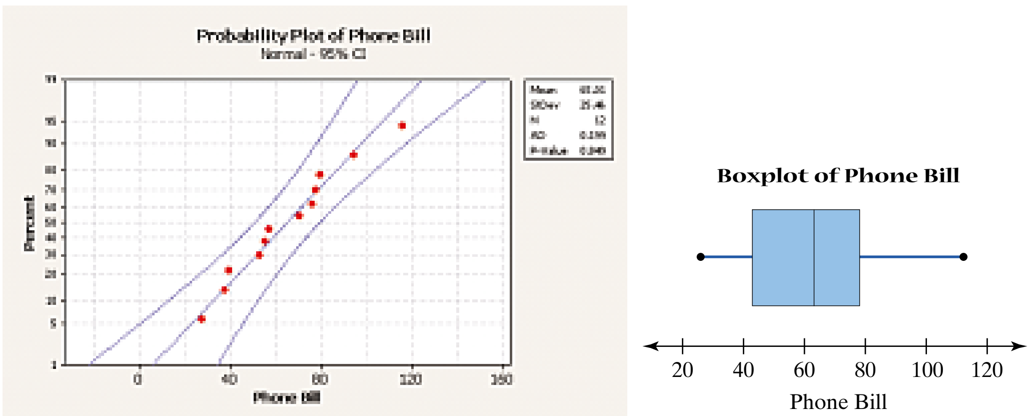
***Problem*:** According to CTIA–The Wireless Association, the mean monthly cell phone bill in 2004 was \$50.64. A market researcher believes that the mean monthly cell phone bill is different today, but is not sure whether bills have declined because of technological advances or increased due to additional use. The researcher phones a simple random sample of 12 cell phone subscribers and obtains the monthly bills shown in Table 1.

Assuming $\sigma = \$18.49$, use these data to determine whether the mean monthly cell phone bills is different from \$50.64 at the $\alpha = 0.05$ level of significance.

***Approach*:** Because the sample size, n, is less than 30, we must verify that the data come from a population that is approximately normal with no outliers. We will construct a normal probability plot and boxplot to verify these requirements. We then proceed to follow Steps 1 through 6.

***Solution*:** Figure 8 displays the normal probability plot and boxplot.

Figure 8



The normal probability plot indicates that the data could come from a population that is normal. The boxplot does not show any outliers.

***Step 1*:** The market researcher wants to know if the mean cell phone bill is different from \$50.64, which can be written $\mu \neq 50.64$. This is a two-tailed test and we have

$$H_0: \mu = 50.64 \quad \text{versus} \quad H_1: \mu \neq 50.64$$

***Step 2*:** The level of significance is $\alpha = 0.05$.

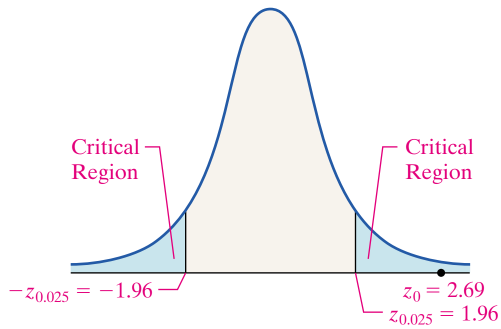
***Step 3*:** From the data in Table 1, the sample mean is computed to be \$65.014. The test statistic is

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{65.014 - 50.64}{18.49/\sqrt{12}} = 2.69$$

The sample mean of \$65.014 is 2.69 standard deviations above the assumed population mean of \$50.64.

***Step 4*:** Because this is a two-tailed test, we determine the critical values at the $\alpha = 0.05$ level of significance to be $-z_{0.05/2} = -1.96$ and $z_{0.05/2} = 1.96$. The critical regions are displayed in Figure 9.

Figure 9



***Step 5*:** Because the test statistic, $z_0 = 2.69$, is greater than the critical value $z_{0.025} = 1.96$, we reject the null hypothesis. That is, the value of the test statistic falls within the critical region, so we reject H_0. We label this point in Figure 9.

***Step 6*:** There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the mean monthly cell phone bill is different from the mean amount in 2004, \$50.64.

Now Work Problem 25 Using the Classical Approach.

3 Test Hypotheses about a Population Mean with σ Known Using *P*-Values

Now let's look at testing hypotheses with *P*-values.

Definition

A ***P*-value** is the probability of observing a sample statistic as extreme or more extreme than the one observed under the assumption that the null hypothesis is true. Put another way, the *P*-value is the likelihood or probability that a sample will result in a sample mean such as the one obtained if the null hypothesis is true.

In Other Words

The smaller the *P*-value, the greater the evidence against the null hypothesis.

A small *P*-value implies that the sample mean is unlikely if the null hypothesis is true and would be considered as evidence against the null hypothesis.

The following procedures can be used to compute *P*-values when testing a hypothesis about a population mean with σ known.

Testing Hypotheses Regarding the Population Mean Using *P*-Values

To test hypotheses regarding the population mean with σ known, we can use the steps that follow to compute the *P*-value, provided that the following two requirements are satisfied.

1. The sample is obtained using simple random sampling.
2. The sample has no outliers and the population from which the sample is drawn is normally distributed or the sample size, n, is large $(n \geq 30)$.

***Step 1*:** Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two-Tailed	Left-Tailed	Right-Tailed
$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$	$H_1: \mu < \mu_0$	$H_1: \mu > \mu_0$

Note: μ_0 is the assumed value of the population mean.

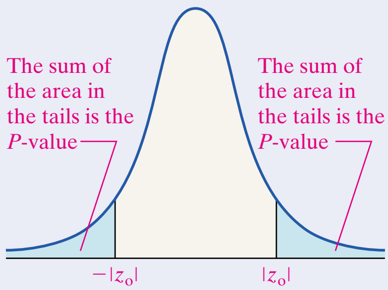
***Step 2*:** Decide on a level of significance, α, depending on the seriousness of making a Type I error.

***Step 3*:** Compute the test statistic, $z_0 = \dfrac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$.

***Step 4*:** Determine the *P*-value.

Two-Tailed

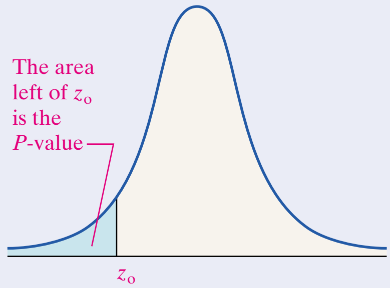
$P\text{-value} = P(Z < -|z_0| \text{ or } Z > |z_0|)$
$= 2P(Z > |z_0|)$



Interpretation: The *P*-value is the probability of obtaining a sample mean that is more than $|z_0|$ standard deviations from the mean stated in the null hypothesis, μ_0.

Left-Tailed

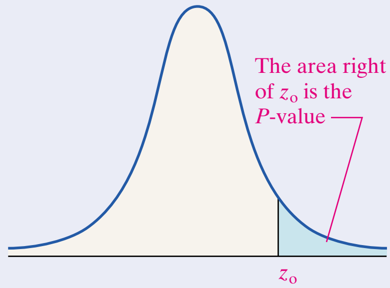
$P\text{-value} = P(Z < z_0)$



Interpretation: The *P*-value is the probability of obtaining a sample mean of $\overline{x}$ or smaller under the assumption that H_0 is true. In other words, it is the probability of obtaining a sample mean that is more than z_0 standard deviations to the left of μ_0.

Right-Tailed

$P\text{-value} = P(Z > z_0)$



Interpretation: The *P*-value is the probability of obtaining a sample mean of $\overline{x}$ or larger under the assumption that H_0 is true. In other words, it is the probability of obtaining a sample mean that is more than z_0 standard deviations to the right of μ_0.

***Step 5*:** Reject the null hypothesis if the *P*-value is less than the level of significance, α. The comparison of the *P*-value and the level of significance is called the **decision rule**.

***Step 6*:** State the conclusion.

EXAMPLE 3 The *P*-Value Approach of Hypothesis Testing: Right-Tailed, Large Sample

***Problem*:** According to the U.S. Federal Highway Administration, the mean number of miles driven annually is 12,200. Patricia believes that residents of the state of Montana drive more than the national average. She obtains a simple random sample of 35 drivers from a list of registered drivers in the state of Montana. The mean number of miles driven for the 35 drivers is 12,895.9. Assuming $\sigma = 3800$ miles, does the sample provide sufficient evidence that residents of the state of Montana drive more than the national average at the $\alpha = 0.1$ level of significance? Use the *P*-value approach.

***Approach*:** Because the sample size is large, we can proceed to Steps 1 through 6.

Solution

***Step 1*:** Patricia wants to know if people in Montana are driving more than 12,200 miles annually. This can be written $\mu > 12{,}200$. This is a right-tailed test and we have

$$H_0: \mu = 12{,}200 \quad \text{versus} \quad H_1: \mu > 12{,}200$$

***Step 2*:** The level of significance is $\alpha = 0.1$.

***Step 3*:** Patricia found the sample mean, $\overline{x}$, to be 12,895.9 miles. The test statistic is

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{12{,}895.9 - 12{,}200}{3800/\sqrt{35}} = 1.08$$

The sample mean of 12,895.9 miles is 1.08 standard deviations above the mean of 12,200.

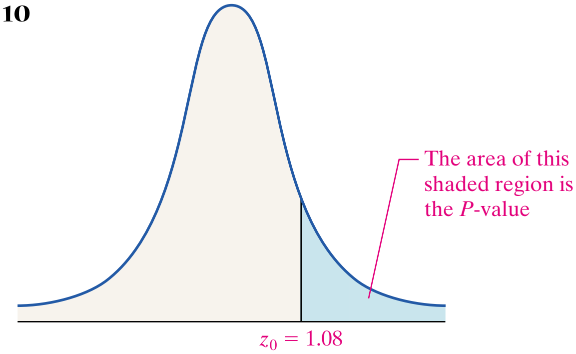
In Other Words
For this test (a right-tailed test), the P-value is the area under the standard normal curve to the right of $z_0 = 1.08$.

***Step 4*:** Because Patricia is performing a right-tailed test,

$$P\text{-value} = P(Z > z_0) = P(Z > 1.08)$$

We have to determine the area under the standard normal curve to the right of $z_0 = 1.08$, as shown in Figure 10.

Figure 10



Using Table IV, we have

$$P\text{-value} = P(Z > 1.08) = 1 - P(Z \le 1.08) = 1 - 0.8599 = 0.1401$$

The probability of obtaining a sample mean of 12,895.9 miles or higher from a population whose mean is 12,200 miles is 0.1401. A P-value of 0.1401 means we would expect to obtain the results we obtained about 14 times out of 100 if the statement in the null hypothesis were true. The results we obtained are not unusual assuming $\mu = 12{,}200$.

In Other Words
If P-value $< \alpha$, then reject the null hypothesis. Put another way—If the P-value is low, the null must go!

***Step 5*:** The decision rule that we use is this: If the P-value is less than the level of significance, α, we reject the null hypothesis. Because $0.1401 > 0.1$, we do not reject the null hypothesis.

***Step 6*:** There is not sufficient evidence at the $\alpha = 0.1$ level of significance to support Patricia's belief that residents of the state of Montana drive more than the national average of 12,200 miles.

Now Work Problem 19 Using the *P*-Value Approach.

EXAMPLE 4 The *P*-Value Approach of Hypothesis Testing: Two-Tailed, Small Sample

Table 2

94.25	38.94	79.15	56.78
70.07	115.59	77.56	37.01
55.00	76.05	27.29	52.48

***Problem*:** According to CTIA—The Wireless Association, the mean monthly cell phone bill in 2004 was \$50.64. A market researcher believes that the mean monthly cell phone bill is different today, but is not sure whether bills have declined because of technological advances or increased due to additional use. The researcher phones a simple random sample of 12 cell phone subscribers and obtains the monthly bills shown in Table 2.

Assuming $\sigma = \$18.49$, use these data to determine whether the mean monthly cell phone bill is different from \$50.64 at the $\alpha = 0.05$ level of significance. Use the P-value approach.

***Approach*:** Because the sample size, n, is less than 30, we must verify that the data come from a population that is approximately normal with no outliers. We will construct a normal probability plot and boxplot to verify these requirements. We then proceed to follow Steps 1 through 6.

***Solution*:** Figure 8 on page 468 displays the normal probability plot and boxplot. The normal probability plot indicates that the data could come from a population that is normal. The boxplot does not show any outliers.

***Step 1*:** The market researcher wants to know if the mean cell phone bill is different from \$50.64, which can be written $\mu \neq 50.64$. This is a two-tailed test and we have

$$H_0: \mu = 50.64 \quad \text{versus} \quad H_1: \mu \neq 50.64$$

***Step 2*:** The level of significance is $\alpha = 0.05$.

***Step 3*:** From the data in Table 2, the sample mean is computed to be \$65.014 and $n = 12$. We assume $\sigma = \$18.49$. The test statistic is

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{65.014 - 50.64}{18.49/\sqrt{12}} = 2.69$$

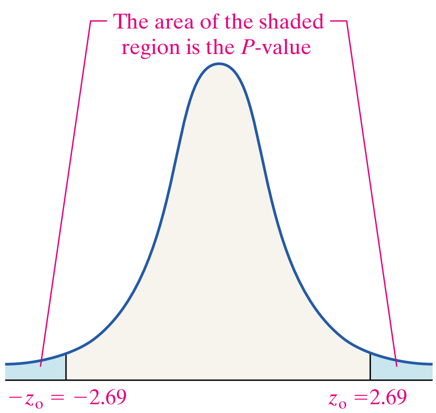
The sample mean of \$65.014 is 2.69 standard deviations above the assumed population mean of \$50.64.

***Step 4*:** Because we are performing a two-tailed test,

$$P\text{-value} = P(Z < -2.69 \text{ or } Z > 2.69)$$

We need to determine the area under the standard normal curve to the right of $Z = 2.69$ and to the left of $Z = -2.69$, as shown in Figure 11.

Figure 11



In Other Words

To find the P-value for a two-tailed test, first determine whether the test statistic, z_0, is positive or negative. If z_0 is negative, determine the area under the standard normal curve to the left of z_0 and then multiply this area by 2. If z_0 is positive, find the area under the standard normal curve to the right of z_0 and double this value.

Using Table IV, we have

$$\begin{aligned} P\text{-value} = P(Z < -2.69 \text{ or } Z > 2.69) &= P(Z < -2.69) + P(Z > 2.69) \\ &= 2P(Z > 2.69) \\ &= 2[1 - P(Z \leq 2.69)] \\ &= 2(1 - 0.9964) \\ &= 2(0.0036) \\ &= 0.0072 \end{aligned}$$

The probability of obtaining a sample mean that is more than 2.69 standard deviations from the status quo population mean of \$50.64 is 0.0072. This means less than 1 sample in 100 will result in a sample mean such as the one we obtained if the statement in the null hypothesis is true.

***Step 5*:** Because the P-value is less than the level of significance $(0.0072 < 0.05)$, we reject the null hypothesis.

***Step 6*:** There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the mean monthly cell phone bill is different from \$50.64, the mean amount in 2004.

Now Work Problem 25 Using the *P*-Value Approach.

EXAMPLE 5 Testing Hypotheses about a Population Mean Using Technology

Problem: According to CTIA—The Wireless Association, the mean monthly cell phone bill in 2004 was \$50.64. A market researcher believes that the mean monthly cell phone bill is different today, but is not sure whether bills have declined because of technological advances, or increased due to additional use. The researcher phones a simple random sample of 12 cell phone subscribers and obtains the data in Table 2.

Assuming $\sigma = \$18.49$, use these data to determine whether the mean monthly cell phone bill is different from \$50.64 at the $\alpha = 0.05$ level of significance.

Approach: We will use MINITAB to test the hypothesis. The steps for testing hypotheses about a population mean with σ known using the TI-83/84 Plus graphing calculators, MINITAB, and Excel are given in the Technology Step by Step on page 479.

Result: Figure 12 shows the results obtained from MINITAB.

Figure 12

One-Sample Z: Cell Phone Bill

```
Test of mu = 50.64 vs not = 50.64
The assumed standard deviation = 18.49

Variable       N    Mean    StDev  SE Mean        95% CI            Z      P
Cell Phone
Bill          12  65.0142  25.4587   5.3376  (54.5527, 75.4757) 2.69  0.007
```

Using Technology

The P-value obtained from technology may be slightly different from the P-value obtained by hand because of rounding.

The P-value is highlighted. From MINITAB, we have $P\text{-value} = 0.007$.

Interpretation: Because the P-value is less than the level of significance $(0.007 < 0.05)$, we reject the null hypothesis. There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the mean monthly cell phone bill is different from \$50.64, the mean amount in 2004.

One advantage of using P-values over the classical approach in hypothesis testing is that P-values provide information regarding the strength of the evidence. In Example 2, we rejected the null hypothesis, but did not learn anything about the strength of the evidence against the null hypothesis. Example 4 tested the same claim using P-values. The P-value was 0.0072. This result not only led us to reject the null hypothesis, but also indicates the strength of the evidence against the null hypothesis: Less than 1 sample in 100 would give us the sample mean that we got if the null hypothesis, H_0: $\mu = 50.64$, were true.

Another advantage of P-values is that they are interpreted the same way, regardless of the type of hypothesis test being performed. If $P\text{-value} < \alpha$, reject the null hypothesis.

4 Test Hypotheses about a Population Mean with σ Known Using Confidence Intervals

Recall that the level of confidence in a confidence interval is a probability that represents the percentage of intervals that will contain μ if repeated samples are obtained. The level of confidence is denoted $(1 - \alpha) \cdot 100\%$. We can use confidence intervals to test H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$ using the following criterion.

When testing H_0: $\mu = \mu_0$ versus H_1: $\mu \neq \mu_0$, if a $(1 - \alpha) \cdot 100\%$ confidence interval contains μ_0, we do not reject the null hypothesis. However, if the confidence interval does not contain μ_0, we have sufficient evidence that supports the statement in the alternative hypothesis and conclude $\mu \neq \mu_0$ at the level of significance, α.

EXAMPLE 6 Testing Hypotheses about a Population Mean Using a Confidence Interval

Problem: Test the hypotheses presented in Examples 2 and 4 at the $\alpha = 0.05$ level of significance by constructing a 95% confidence interval about μ, the population mean monthly cell phone bill.

Approach: We construct the 95% confidence interval using the data in Table 1 or Table 2. If the interval contains the status quo mean of \$50.64, we do not reject the null hypothesis.

Solution: We use the formula on page 410 to find the lower and upper bounds with $\bar{x} = \$65.014$, $\sigma = \$18.49$, and $n = 12$.

$$\text{Lower bound:} \quad \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \$65.014 - 1.96 \cdot \frac{18.49}{\sqrt{12}} = \$65.014 - \$10.462 = \$54.552$$

$$\text{Upper bound:} \quad \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = \$65.014 + 1.96 \cdot \frac{18.49}{\sqrt{12}} = \$65.014 + \$10.462 = \$75.476$$

We are 95% confident the mean monthly cell phone bill is between \$54.552 and \$75.476. Because the mean stated in the null hypothesis, H_0: $\mu = 50.64$, is not included in this interval, we reject the null hypothesis. There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the mean monthly cell phone bill is different from \$50.64.

Now Work Problem 35.

5 Understand the Difference between Statistical Significance and Practical Significance

In Other Words
Results are statistically significant if the difference between the observed result and the statement made in the null hypothesis is unlikely to occur due to chance alone.

When a large sample size is used in a hypothesis test, the results could be statistically significant even though the difference between the sample statistic and mean stated in the null hypothesis may have no *practical significance*.

Definition

Practical significance refers to the idea that small differences between the statistic and parameter stated in the null hypothesis are statistically significant, while the difference is not large enough to cause concern or be considered important.

EXAMPLE 7 Statistical versus Practical Significance

Problem: According to the American Community Survey, the mean travel time to work in Dallas, Texas, in 2003 was 23.6 minutes. Suppose the Department of Transportation in Dallas just reprogrammed all the traffic lights in an attempt to reduce travel time. To determine if there is evidence that indicates that travel times in Dallas have decreased as a result of the reprogramming, the Department of Transportation obtains a sample of 2500 commuters, records their travel time to work, and obtains a sample mean of 23.3 minutes. Assuming that the population standard deviation travel time to work is known to be 8.4 minutes, determine whether travel times in Dallas have decreased as a result of the reprogramming at the $\alpha = 0.05$ level of significance.

Approach: We will use both the classical approach and P-value approach to test the hypotheses.

Solution

Step 1: The Department of Transportation wants to know if the mean travel time to work has decreased from 23.6 minutes, which can be written $\mu < 23.6$. This is a left-tailed test and we have

$$H_0: \mu = 23.6 \quad \text{versus} \quad H_1: \mu < 23.6$$

Step 2: The level of significance is $\alpha = 0.05$.

Step 3: The test statistic is

$$z_0 = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{23.3 - 23.6}{\frac{8.4}{\sqrt{2500}}} = -1.79$$

Classical Approach

Step 4: This is a left-tailed test. With $\alpha = 0.05$, the critical value is $-z_{0.05} = -1.645$.

Step 5: Because the test statistic is less than the critical value (the critical value falls in the critical region), we reject the null hypothesis.

P-Value Approach

Step 4: Because this is a left-tailed test, the P-value is $P\text{-value} = P(Z < z_0) = P(Z < -1.79) = 0.0367$.

Step 5: Because the P-value is less than the level of significance $(0.0367 < 0.05)$, we reject the null hypothesis.

Step 6: There is sufficient evidence at the $\alpha = 0.05$ level of significance to conclude that the mean travel time to work has decreased.

While the difference between 23.3 and 23.6 is statistically significant, it really has no practical meaning. After all, is 0.3 minutes (18 seconds) really going to make anyone feel better about his or her commute to work?

CAUTION

Beware of studies with large sample sizes that claim statistical significance because the differences may not have any practical meaning.

The reason that the results from Example 7 were statistically significant had to do with the large sample size. The moral of the story is this:

> Large sample sizes can lead to results that are statistically significant, while the difference between the statistic and parameter in the null hypothesis is not enough to be considered practically significant.

10.2 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. State the requirements that must be satisfied to test a hypothesis regarding a population mean with σ known.
2. Determine the critical value for a right-tailed test regarding a population mean with σ known at the $\alpha = 0.01$ level of significance.
3. Determine the critical value for a two-tailed test regarding a population mean with σ known at the $\alpha = 0.05$ level of significance.
4. The procedures for testing a hypothesis regarding a population mean with σ known are robust. What does this mean?
5. Explain what a P-value is. What is the criterion for rejecting the null hypothesis using the P-value approach?
6. Suppose that we are testing the hypotheses $H_0: \mu = \mu_0$ versus $H_1: \mu < \mu_0$ and we find the P-value to be 0.23. Explain what this means. Would you reject H_0? Why?
7. Suppose that we are testing the hypotheses $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ and we find the P-value to be 0.02. Explain what this means. Would you reject H_0? Why?
8. Discuss the advantages and disadvantages of using the classical approach to hypothesis testing. Discuss the advantages and disadvantages of using the P-value approach to hypothesis testing.
9. In your own words, explain the difference between *statistical significance* and *practical significance*.
10. *True or False*: To test $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$ using a 5% level of significance, we could construct a 95% confidence interval.

Skill Building

11. To test H_0: $\mu = 50$ versus H_1: $\mu < 50$, a random sample of size $n = 24$ is obtained from a population that is known to be normally distributed with $\sigma = 12$.

(a) If the sample mean is determined to be $\bar{x} = 47.1$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, determine the critical value.
(c) Draw a normal curve that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

12. To test H_0: $\mu = 40$ versus H_1: $\mu > 40$, a random sample of size $n = 25$ is obtained from a population that is known to be normally distributed with $\sigma = 6$.

(a) If the sample mean is determined to be $\bar{x} = 42.3$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.1$ level of significance, determine the critical value.
(c) Draw a normal curve that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

13. To test H_0: $\mu = 100$ versus H_1: $\mu \neq 100$, a random sample of size $n = 23$ is obtained from a population that is known to be normally distributed with $\sigma = 7$.

(a) If the sample mean is determined to be $\bar{x} = 104.8$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.01$ level of significance, determine the critical values.
(c) Draw a normal curve that depicts the critical regions.
(d) Will the researcher reject the null hypothesis? Why?

14. To test H_0: $\mu = 80$ versus H_1: $\mu < 80$, a random sample of size $n = 22$ is obtained from a population that is known to be normally distributed with $\sigma = 11$.

(a) If the sample mean is determined to be $\bar{x} = 76.9$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.02$ level of significance, determine the critical value.
(c) Draw a normal curve that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

15. To test H_0: $\mu = 20$ versus H_1: $\mu < 20$, a random sample of size $n = 18$ is obtained from a population that is known to be normally distributed with $\sigma = 3$.

(a) If the sample mean is determined to be $\bar{x} = 18.3$, compute and interpret the P-value.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, will the researcher reject the null hypothesis? Why?

16. To test H_0: $\mu = 4.5$ versus H_1: $\mu > 4.5$, a random sample of size $n = 13$ is obtained from a population that is known to be normally distributed with $\sigma = 1.2$.

(a) If the sample mean is determined to be $\bar{x} = 4.9$, compute and interpret the P-value.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.1$ level of significance, will the researcher reject the null hypothesis? Why?

17. To test H_0: $\mu = 105$ versus H_1: $\mu \neq 105$, a random sample of size $n = 35$ is obtained from a population whose standard deviation is known to be $\sigma = 12$.

(a) Does the population need to be normally distributed to compute the P-value?
(b) If the sample mean is determined to be $\bar{x} = 101.2$, compute and interpret the P-value.
(c) If the researcher decides to test this hypothesis at the $\alpha = 0.02$ level of significance, will the researcher reject the null hypothesis? Why?

18. To test H_0: $\mu = 45$ versus H_1: $\mu \neq 45$, a random sample of size $n = 40$ is obtained from a population whose standard deviation is known to be $\sigma = 8$.

(a) Does the population need to be normally distributed to compute the P-value?
(b) If the sample mean is determined to be $\bar{x} = 48.3$, compute and interpret the P-value.
(c) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, will the researcher reject the null hypothesis? Why?

Applying the Concepts

19. Are Women Getting Taller? A researcher believes that the average height of a woman aged 20 years or older is greater than the 1994 mean height of 63.7 inches, on the basis of data obtained from the Centers for Disease Control and Prevention's *Advance Data Report*, No. 347. She obtains a simple random sample of 45 women and finds the sample mean height to be 63.9 inches. Assume that the population standard deviation is 3.5 inches. Test the researcher's belief using either the classical approach or the P-value approach at the $\alpha = 0.05$ level of significance.

NW

20. ATM Withdrawals The manufacturer of a certain type of ATM machine reports that the mean ATM withdrawal is \$60. The manager of a convenience store with an ATM machine claims that mean withdrawal from his machine is less than this amount. He obtains a simple random sample of 35 withdrawals over the past year and finds the sample mean to be \$52. Assume that the population standard deviation is \$13. Test the manager's claim using either the classical approach or the P-value approach at the $\alpha = 0.05$ level of significance.

21. SAT Exam Scores A school administrator believes that students whose first language learned is not English score worse on the verbal portion of the SAT exam than students whose first language is English. The mean SAT verbal score of students whose first language is English is 515, on the basis of data obtained from the College Board. Suppose a simple random sample of 20 students whose first language learned was not English results in a sample mean SAT verbal score of 458. SAT verbal scores are normally distributed with a population standard deviation of 112.

(a) Why is it necessary for SAT verbal scores to be normally distributed to test the hypotheses using the methods of this section?
(b) Use the classical approach or the P-value approach at the $\alpha = 0.10$ level of significance to determine if there is evidence to support the administrator's belief.

22. SAT Exam Scores A school administrator wonders if students whose first language learned is not English score differently on the math portion of the SAT exam than students whose first language is English. The mean SAT math score of students whose first language is English is 516, on the basis of data obtained from the College Board. Suppose a simple random sample of 20 students whose first language learned was not English results in a sample mean SAT math score of 522. SAT math scores are normally distributed with a population standard deviation of 114.

(a) Why is it necessary for SAT math scores to be normally distributed to test the hypotheses using the methods of this section?

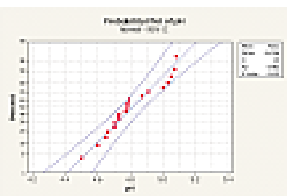
(b) Determine whether students whose first language learned is not English score differently on the math portion of the SAT exam using the classical approach or the *P*-value approach at the $\alpha = 0.10$ level of significance.

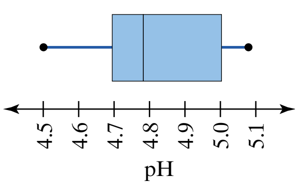
23. Acid Rain In 1990, the mean pH level of the rain in Pierce County, Washington, was 5.03. A biologist wonders if the acidity of rain has increased. (This would mean that the pH level of the rain has decreased.) From a random sample of 19 rain dates in 2004, she obtains the following data:

5.08	4.66	4.70	4.87
4.78	5.00	4.50	4.73
4.79	4.65	4.91	5.07
5.03	4.78	4.77	4.60
4.73	5.05	4.70	

Source: National Atmospheric Deposition Program

(a) Because the sample size is small, she must verify that pH level is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown next. Are the conditions for testing the hypothesis satisfied?



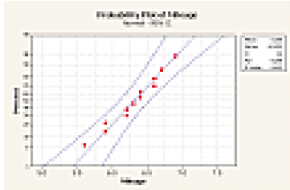


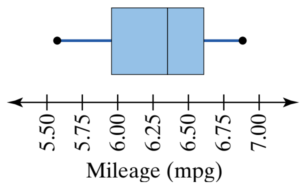
(b) Test whether the acidity of rain has increased, assuming that $\sigma = 0.2$ at the $\alpha = 0.01$ level of significance.

24. Fuel Catalyst To improve fuel efficiency and reduce pollution, the owner of a trucking fleet decides to install a new fuel catalyst in all his semitrucks. He feels that the catalyst will help to increase the number of miles per gallon. Before the installation, his trucks had a mean gas mileage of 5.6 miles per gallon. A random sample of 12 trucks after the installation gave the following gas mileages:

5.9	6.7	6.9	6.4	6.6	6.3
6.2	5.9	6.2	6.4	6.6	5.6

(a) Because the sample size is small, he must verify that mileage is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?



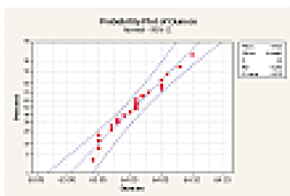


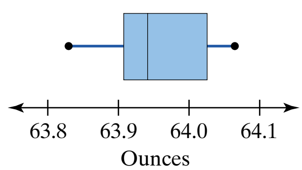
(b) Test if the fuel catalyst is effective, assuming that $\sigma = 0.5$ mile per gallon at the 0.05 level of significance.

25. Filling Bottles NW A certain brand of apple juice is supposed to have 64 ounces of juice. Because the punishment for underfilling bottles is severe, the target mean amount of juice is 64.05 ounces. However, the filling machine is not precise, and the exact amount of juice varies from bottle to bottle. The quality-control manager wishes to verify that the mean amount of juice in each bottle is 64.05 ounces so that she can be sure that the machine is not over- or underfilling. She randomly samples 22 bottles of juice and measures the content and obtains the following data:

64.05	64.05	64.03	63.97	63.95	64.02
64.01	63.99	64.00	64.01	64.06	63.94
63.98	64.05	63.95	64.01	64.08	64.01
63.95	63.97	64.10	63.98		

(a) Because the sample size is small, she must verify that the amount of juice is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





(b) Should the assembly line be shut down so that the machine can be recalibrated? Assume $\sigma = 0.06$ ounces and use a 0.01 level of significance.

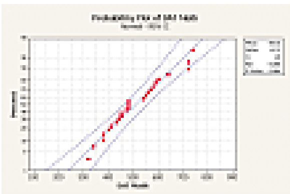
(c) Explain why a level of significance of $\alpha = 0.01$ might be more reasonable than $\alpha = 0.1$. [**Hint:** Consider the consequences of incorrectly rejecting the null hypothesis.]

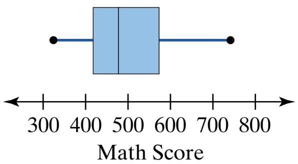
26. SAT Reasoning Test In 2005, in response to various criticisms, the College Board implemented changes to the SAT test. In particular, the math portion eliminated quantitative comparisons and expanded the topics covered. A school administrator believes that the new math portion is more difficult and will result in lower scores on the math portion compared to the 2004 average score of

516. A random sample of 25 students taking the new SAT test resulted in the following scores on the math portion.

410	720	480	560	590	340	430	400	440
540	480	450	720	570	380	740	640	600
320	580	450	480	550	470	380		

(a) Because the sample size is small, he must verify that the new math scores are normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?



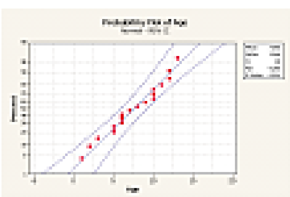


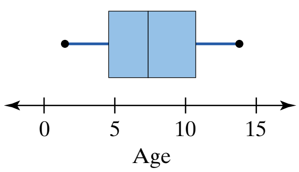
(b) Do you believe students scored lower? Assume $\sigma = 114$.

27. Are Cars Younger? Suppose you have just been hired by Ford Motor Company. Management wants to know if cars are younger today versus 1995. According to the Nationwide Personal Transportation Survey conducted by the U.S. Department of Transportation, the mean age of a car in 1995 was 8.33 years. Based on a random sample of 18 automobile owners, you obtain the ages as shown in the following table:

8	12	1	2	13	3
5	9	12	6	5	6
10	7	10	11	6	10

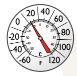
(a) Because the sample size is small, you must verify that the data are normally distributed with no outliers by drawing a normal probability plot and boxplot. Based on the following graphs, can you perform a hypothesis test?





(b) Are cars younger today? Assume that $\sigma = 3.8$ years. Use the $\alpha = 0.1$ level of significance.

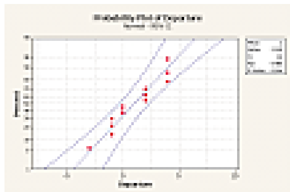
28. It's a Hot One! Recently, a friend of mine claimed that the summer of 2000 in Houston, Texas, was hotter than usual. To test his claim, I went to AccuWeather.com and randomly selected 12 days in the summer of 2000. I then recorded the departure from normal, with positive values indicating above-normal temperatures and negative values indicating below-normal temperatures, as shown in the following:

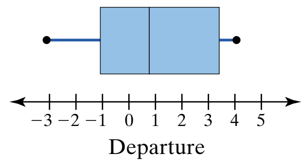


+4	−1	0	+2
+2	+4	−1	−3
−1	0	+2	+4

Source: *AccuWeather.com*

(a) Because the sample size is small, I must verify that the temperature departure is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





(b) Was the summer of 2000 hotter than normal in Houston? Assume that $\sigma = 1.8$. Use the $\alpha = 0.05$ level of significance.

29. Farm Size NW In 1990, the average farm size in Kansas was 694 acres, according to data obtained from the U.S. Department of Agriculture. A researcher claims that farm sizes are larger now due to consolidation of farms. She obtains a random sample of 40 farms and determines the mean size to be 731 acres. Assume that $\sigma = 212$ acres. Test the researcher's claim at the $\alpha = 0.05$ level of significance.

30. Oil Output An energy official claims that the oil output per well in the United States has declined from the 1998 level of 11.1 barrels per day. He randomly samples 50 wells throughout the United States and determines the mean output to be 10.7 barrels per day. Assume that $\sigma = 1.3$ barrels. Test the researcher's claim at the $\alpha = 0.05$ level of significance.

31. Volume of Dell Computer Stock The average daily volume of Dell Computer stock in 2000 was $\mu = 31.8$ million shares, with a standard deviation of $\sigma = 14.8$ million shares, according to Yahoo!Finance. Based on a random sample of 35 trading days in 2004, the sample mean number of shares traded is found to be 23.5 million. Is the volume of Dell stock different in 2004? Use the $\alpha = 0.05$ level of significance.

32. Volume of Motorola Stock The average daily volume of Motorola stock in 2000 was $\mu = 11.4$ million shares, with a standard deviation of $\sigma = 8.3$ million shares, according to Yahoo!Finance. Based on a random sample of 35 trading days in 2004, the sample mean number of shares traded is found to be 13.3 million shares. Is the volume of Motorola stock different in 2004? Use the $\alpha = 0.05$ level of significance.

33. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 25 by constructing a 99% confidence interval.

34. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 26 by constructing a 95% confidence interval.

35. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 31 by constructing a 95% confidence interval.
NW

36. Using Confidence Intervals to Test Hypotheses Test the hypotheses in Problem 32 by constructing a 95% confidence interval.

37. Statistical Significance versus Practical Significance A math teacher claims that she has developed a review course that increases the scores of students on the math portion of the SAT exam. Based on data from the College Board, SAT scores are normally distributed with $\mu = 514$ and $\sigma = 113$. The teacher obtains a random sample of 1800 students, puts them through the review class, and finds that the mean SAT math score of the 1800 students is 518.

(a) State the null and alternative hypotheses.
(b) Test the hypothesis at the $\alpha = 0.10$ level of significance. Is a mean SAT math score of 518 significantly higher than 514?
(c) Do you think that a mean SAT math score of 518 versus 514 will affect the decision of a school admissions administrator? In other words, does the increase in the score have any practical significance?
(d) Test the hypothesis at the $\alpha = 0.10$ level of significance with $n = 400$ students. Assume that the sample mean is still 518. Is a sample mean of 518 significantly more than 514? Conclude that large sample sizes cause P-values to shrink substantially, all other things being the same.

38. Statistical Significance versus Practical Significance The manufacturer of a daily dietary supplement claims that its product will help people lose weight. The company obtains a random sample of 950 adult males aged 20 to 74 who take the supplement and finds their mean weight loss after eight weeks to be 0.9 pounds. Assume the population standard deviation weight loss is $\sigma = 7.2$ pounds.

(a) State the null and alternative hypotheses.
(b) Test the hypothesis at the $\alpha = 0.1$ level of significance. Is a mean weight loss of 0.9 pound significant?
(c) Do you think that a mean weight loss of 0.9 pounds is worth the expense and commitment of a daily dietary supplement? In other words, does the weight loss have any practical significance?
(d) Test the hypothesis at the $\alpha = 0.1$ level of significance with $n = 40$ subjects. Assume that the sample mean weight loss is still 0.9 pounds. Is a sample mean weight loss of 0.9 pounds significantly more than 0 pounds? Conclude that large sample sizes cause P-values to shrink substantially, all other things being the same.

39. Simulation Simulate drawing 50 simple random samples of size $n = 20$ from a population that is normally distributed with mean 80 and standard deviation 7.

(a) Test the null hypothesis H_0: $\mu = 80$ versus the alternative hypothesis H_1: $\mu \neq 80$.
(b) Suppose we were testing this hypothesis at the $\alpha = 0.1$ level of significance. How many of the 50 samples would you expect to result in a Type I error?
(c) Count the number of samples that lead to a rejection of the null hypothesis. Is it close to the expected value determined in part (b)?
(d) Describe how we know that a rejection of the null hypothesis results in making a Type I error in this situation.

40. Simulation Simulate drawing 40 simple random samples of size $n = 35$ from a population that is exponentially distributed with mean 8 and standard deviation $\sqrt{8}$.

(a) Test the null hypothesis H_0: $\mu = 8$ versus the alternative hypothesis H_1: $\mu \neq 8$.
(b) Suppose we were testing this hypothesis at the $\alpha = 0.05$ level of significance. How many of the 40 samples would you expect to result in a Type I error?
(c) Count the number of samples that lead to a rejection of the null hypothesis. Is it close to the expected value determined in part (b)?
(d) Describe how we know that a rejection of the null hypothesis results in making a Type I error in this situation.

41. Suppose a chemical company has developed a catalyst that is meant to reduce reaction time in a chemical process. For a certain chemical process, reaction time is known to be 150 seconds. The researchers conducted an experiment with the catalyst 40 times and measured reaction time. The researchers reported that the catalysts reduced reaction time with a P-value of 0.02.

(a) Identify the null and alternative hypotheses.
(b) Explain what this result means. Do you believe that the catalyst is effective?

Technology Step by Step — Hypothesis Tests Regarding μ, σ Known

TI-83/84 Plus

***Step 1*:** If necessary, enter raw data in L1.

***Step 2*:** Press STAT, highlight TESTS, and select `1:Z-Test`.

***Step 3*:** If the data are raw, highlight DATA; make sure that List1 is set to L1 and Freq is set to 1. If summary statistics are known, highlight STATS and enter the summary statistics. Following σ, enter the population standard deviation. For the value of μ_0, enter the value of the mean stated in the null hypothesis.

***Step 4*:** Select the direction of the alternative hypothesis.

***Step 5*:** Highlight Calculate and press ENTER. The TI-83/84 Plus gives the P-value.

MINITAB

***Step 1*:** Enter raw data in column C1.

***Step 2*:** Select the **Stat** menu, highlight **Basic Statistics**, and then highlight **1-Sample Z** ....

***Step 3*:** Click Options. In the cell marked "Alternative," select the appropriate direction for the alternative hypothesis. Click OK.

***Step 4*:** Enter C1 in the cell marked "Variables." In the cell labeled "Test Mean", enter the value of the mean stated in the null hypothesis. In the cell labeled "standard deviation," enter the value of σ. Click OK.

Excel

***Step 1*:** If necessary, enter raw data in column A.

***Step 2*:** Load the PHStat Add-in.

***Step 3*:** Select the **PHStat** menu, highlight **One Sample Tests ...**, and then highlight **Z Test for the mean, sigma known ....**

***Step 4*:** Enter the value of the null hypothesis, the level of significance, α, and the value of σ. If the summary statistics are known, click "Sample statistics known" and enter the sample size and sample mean. If summary statistics are unknown, click "Sample statistics unknown." With the cursor in the "Sample cell range" cell, highlight the data in column A. Click the option corresponding to the desired test [two-tail, upper (right) tail, or lower (left) tail]. Click OK.

10.3 Hypothesis Tests for a Population Mean in Practice

Preparing for This Section Before getting started, review the following:

- Sampling distribution of $\overline{x}$ (Section 8.1, pp. 374–388)
- The t-distribution (Section 9.2, pp. 423–426)
- Using probabilities to identify unusual events (Section 5.1, p. 225)

Objective

1 Test hypotheses about a population mean with σ unknown

In Section 10.2, we assumed that the population standard deviation, σ, was known when testing hypotheses regarding the population mean. We now introduce procedures for testing hypotheses regarding a population mean when σ is not known. The only difference from the situation where σ is known is that we must use the t-distribution rather than the z-distribution.

In Other Words
When σ is known, use z; when σ is unknown, use t.

We do not replace σ with s and say that $z = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$ is normally distributed with mean 0 and standard deviation 1. Instead, $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$ follows **Student's *t*-distribution** with $n - 1$ degrees of freedom. Let's review the properties of the t-distribution.

Properties of the *t*-Distribution

1. The t-distribution is different for different degrees of freedom.
2. The t-distribution is centered at 0 and is symmetric about 0.

3. The area under the curve is 1. Because of the symmetry, the area under the curve to the right of 0 equals the area under the curve to the left of 0, which equals $\frac{1}{2}$.
4. As t increases without bound, the graph approaches, but never equals, zero. As t decreases without bound, the graph approaches, but never equals, zero.
5. The area in the tails of the t-distribution is a little greater than the area in the tails of the standard normal distribution, because using s as an estimate of σ introduces more variability to the t-statistic.
6. As the sample size n increases, the density curve of t gets closer to the standard normal density curve. This result occurs because, as the sample size n increases, the values of s get closer to the values of σ, by the Law of Large Numbers.

1 Test Hypotheses about a Population Mean with σ Unknown

From Section 10.2, we know that there are two approaches (besides using confidence intervals) we can use to test claims regarding a population mean, the classical approach and the P-value approach. We will present both methods here. Your instructor may choose one or both approaches.

Both the classical approach and the P-value approach to testing hypotheses about μ with σ unknown follow the exact same logic as testing hypotheses about μ with σ known. The only difference is that we use Student's t-distribution, rather than the normal distribution.

Testing Hypotheses Regarding a Population Mean with σ Unknown

To test hypotheses regarding the population mean with σ unknown, we use the following steps, provided that

1. The sample is obtained using simple random sampling.
2. The sample has no outliers and the population from which the sample is drawn is normally distributed or the sample size, n, is large $(n \geq 30)$.

***Step 1*:** Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two-Tailed	**Left-Tailed**	**Right-Tailed**
H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$	H_0: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$	H_1: $\mu < \mu_0$	H_1: $\mu > \mu_0$

Note: μ_0 is the assumed value of the population mean.

***Step 2*:** Select a level of significance α, depending on the seriousness of making a Type I error.

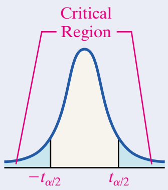
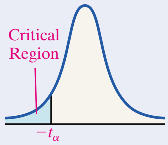
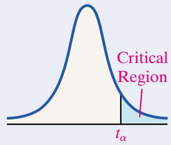
***Step 3*:** Compute the test statistic

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

which follows Student's t-distribution with $n - 1$ degrees of freedom.

Classical Approach

Step 4**:** Use Table V to determine the critical value using $n - 1$ degrees of freedom.

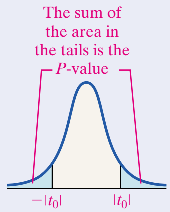
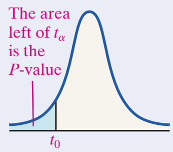
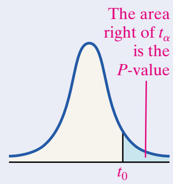
	Two-Tailed	Left-Tailed	Right-Tailed
Critical value(s)	$-t_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}}$	$-t_{\alpha}$	t_{α}
Critical region			

Step 5**:** Compare the critical value with the test statistic.

Two-Tailed	Left-Tailed	Right-Tailed
If $t_0 < -t_{\frac{\alpha}{2}}$ or $t_0 > t_{\frac{\alpha}{2}}$, reject the null hypothesis	If $t_0 < -t_{\alpha}$, reject the null hypothesis	If $t_0 > t_{\alpha}$, reject the null hypothesis

P-Value Approach

Step 4**:** Use Table V to estimate the *P*-value using $n - 1$ degrees of freedom.

Two-Tailed	Left-Tailed	Right-Tailed
		

Step 5**:** If the *P*-value $< \alpha$, reject the null hypothesis.

Step 6**:** State the conclusion.

Notice that the procedure just presented requires either that the population from which the sample was drawn be normal or that the sample size be large ($n \geq 30$). The procedure is robust, so minor departures from normality will not adversely affect the results of the test. However, if the data include outliers, the procedure should not be used.

Just as we did for hypothesis tests with σ known, we will verify these assumptions by constructing normal probability plots (to assess normality) and boxplots (to discover whether there are outliers). If the normal probability plot indicates that the data do not come from a normal population or if the boxplot reveals outliers, nonparametric tests should be performed, which are not discussed in this text.

Before we look at a couple of examples, it is important to understand that we cannot find exact *P*-values using the *t*-distribution table (Table V) because the table provides *t*-values only for certain areas. However, we can use the table to calculate lower and upper bounds on the *P*-value. To find exact *P*-values, we use statistical software or a graphing calculator with advanced statistical features.

In Other Words
When σ is unknown, exact *P*-values can be found using technology.

EXAMPLE 1 Testing a Hypothesis about a Population Mean, Large Sample

Problem: According to the Centers for Disease Control, the mean number of cigarettes smoked per day by individuals who are daily smokers is 18.1. A researcher wonders if retired adults smoke less than the general population of daily smokers, so she obtains a random sample of 40 retired adults who are current smokers and records the number of cigarettes smoked on a randomly selected day. The data result in a sample mean of 16.8 cigarettes and a standard deviation of 4.7 cigarettes. Is there sufficient evidence at the $\alpha = 0.1$ level of significance to conclude that retired adults who are daily smokers smoke less than the general population of daily smokers?

Approach: Because the sample size is large, we can follow the steps to testing a claim about a population mean given on pages 481–482.

Solution

Step 1: The researcher wants to know if retired adults smoke less than the general population. The mean number of cigarettes smoked per day by individuals who are daily smokers is 18.1, so we have

$$H_0: \mu = 18.1 \quad \text{versus} \quad H_1: \mu < 18.1$$

This is a left-tailed test.

Step 2: The level of significance is $\alpha = 0.1$.

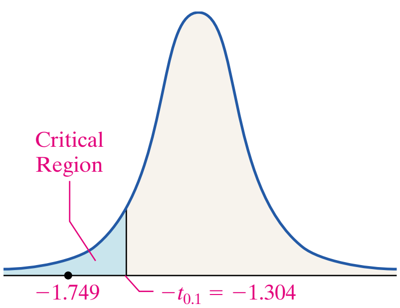
Step 3: The sample mean is $\bar{x} = 16.8$, and the sample standard deviation is $s = 4.7$. The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{16.8 - 18.1}{4.7/\sqrt{40}} = -1.749$$

Classical Approach

Step 4: Because this is a left-tailed test, we determine the critical t-value at the $\alpha = 0.1$ level of significance with $n - 1 = 40 - 1 = 39$ degrees of freedom to be $-t_{0.1} = -1.304$. The critical region is displayed in Figure 13.

Figure 13



Step 5: Because the test statistic $t_0 = -1.749$ is less than the critical value $-t_{0.1} = -1.304$, the researcher rejects the null hypothesis. We label this point in Figure 13.

P-Value Approach

Step 4: Because this is a left-tailed test, the P-value is the area under the t-distribution with $40 - 1 = 39$ degrees of freedom to the left of the test statistic, $t_0 = -1.749$, as shown in Figure 14(a). That is, P-value $= P(t < t_0) = P(t < -1.749)$, with 39 degrees of freedom.

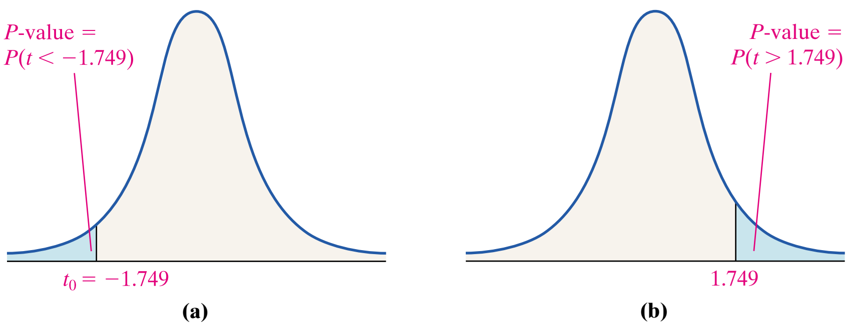
Because of the symmetry of the t-distribution, the area under the distribution to the left of -1.749 equals the area under the distribution to the right of 1.749. So P-value $= P(t < -1.749) = P(t > 1.749)$. See Figure 14(b).

Using Table V, we find the row that corresponds to 39 degrees of freedom. The value 1.749 lies between 1.685 and 2.023. The value of 1.685 has an area under the t-distribution of 0.05 to the right, with 39 degrees of freedom. The area under the t-distribution with 39 degrees of freedom to the right of 2.023 is 0.025. See Figure 15 on page 484.

Because 1.749 is between 1.685 and 2.023, the P-value is between 0.025 and 0.05. So

$$0.025 < P\text{-value} < 0.05$$

Figure 14



Step 5: Because the P-value is less than the level of significance $\alpha = 0.1$, we reject the null hypothesis.

Step 6: There is sufficient evidence to conclude that retired adults smoke less than the general population of daily smokers at the $\alpha = 0.1$ level of significance.

Figure 15

	Area in Right Tail											
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.0025**	**0.001**	**0.0005**
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657	127.321	318.289	636.558
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.328	31.600
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.214	12.924
37	0.681	0.851	1.051	1.305	1.687	2.026	2.129	2.431	2.715	2.985	3.326	3.574
38	0.681	0.851	1.051	1.304	1.686	2.024	2.127	2.429	2.712	2.980	3.339	3.566
39	0.641	0.851	1.050	1.304	1.685	2.023	2.125	2.426	2.704	2.976	3.313	3.558
40	0.681	0.851	1.050	1.303	1.654	2.021	2.123	2.423	2.706	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460

Obtaining the approximate *P*-value in Example 1 was somewhat challenging. With the aid of technology, we can find the exact *P*-value quite painlessly.

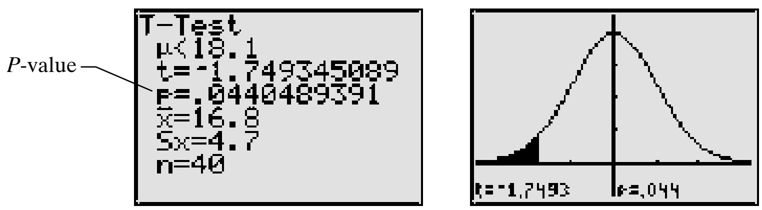
EXAMPLE 2 Testing a Hypothesis about a Population Mean Using Technology

Problem: Obtain an exact *P*-value for the problem in Example 1 using statistical software or a graphing calculator with advanced statistical features.

Approach: We will use a TI-84 Plus graphing calculator to obtain the *P*-value. The steps for testing hypotheses about a population mean with σ unknown using the TI-83/84 Plus graphing calculators, MINITAB, and Excel are given in the Technology Step by Step on page 493.

Result: Figure 16(a) shows the results from the TI-84 Plus using the Calculate option. Figure 16(b) shows the results using the Draw option. The *P*-value is 0.044. Notice the *P*-value is between 0.025 and 0.05. This agrees with the results in Example 1.

Figure 16



(a) Calculate Option **(b)** Draw Option

Now Work Problem 13.

EXAMPLE 3 Testing a Hypothesis about a Population Mean, Small Sample

Table 3

19.68	20.66	19.56
19.98	20.65	19.61
20.55	20.36	21.02
21.5	19.74	

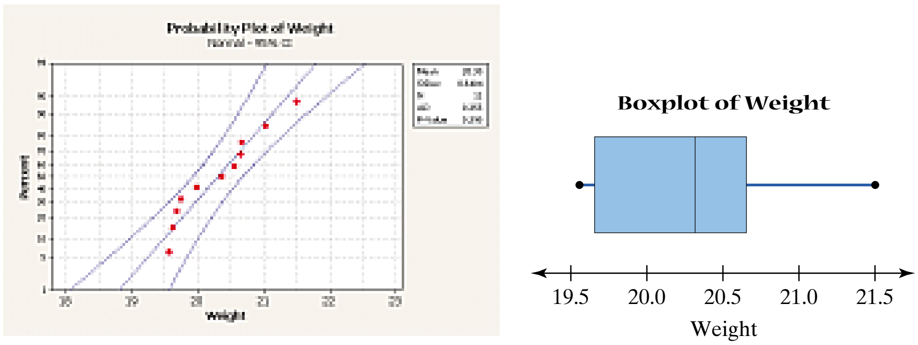
Source: Michael Carlisle, student at Joliet Junior College

Problem: The "fun size" of a Snickers bar is supposed to weigh 20 grams. Because the punishment for selling candy bars that weigh less than 20 grams is so severe, the manufacturer calibrates the machine so that the mean weight is 20.1 grams. The quality-control engineer at M&M–Mars, the company that manufactures Snickers bars, is concerned that the machine that manufacturers the candy is miscalibrated. She obtains a random sample of 11 candy bars, weighs them, and obtains the data in Table 3. Should the machine be shut down and calibrated? Because shutting down the plant is very expensive, she decides to conduct the test at the $\alpha = 0.01$ level of significance.

Approach: Before we can perform the hypothesis test, we must verify that the data come from a population that is normally distributed with no outliers. We will construct a normal probability plot and boxplot to verify these requirements. We then proceed to follow Steps 1 through 6.

Solution: Figure 17 displays the normal probability plot and boxplot.

Figure 17



The normal probability plot indicates that the data come from a population that is approximately normal. The boxplot does not show any outliers. We can proceed with the hypothesis test.

Step 1: The quality-control engineer wishes to determine whether the Snickers have a mean weight of 20.1 grams or not. The hypotheses can be written

$$H_0: \mu = 20.1 \quad \text{versus} \quad H_1: \mu \neq 20.1$$

This is a two-tailed test.

Step 2: The level of significance is $\alpha = 0.01$.

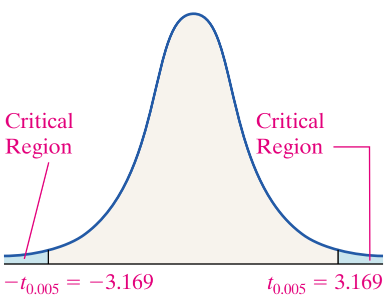
Step 3: From the data in Table 3, the sample mean is $\bar{x} = 20.3$, and the sample standard deviation is $s = 0.64$. The test statistic is

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{20.3 - 20.1}{0.64/\sqrt{11}} = 1.036$$

Classical Approach

Step 4: Because this is a two-tailed test, we determine the critical t-values at the $\alpha = 0.01$ level of significance with $n - 1 = 11 - 1 = 10$ degrees of freedom to be $-t_{0.01/2} = -t_{0.005} = -3.169$ and $t_{0.01/2} = t_{0.005} = 3.169$. The critical regions are displayed in Figure 18.

Figure 18



P-Value Approach

Step 4: Because this is a two-tailed test, the P-value is the area under the t-distribution with $11 - 1 = 10$ degrees of freedom to the left of the test statistic $-t_0 = -1.036$ and to the right of $t_0 = 1.036$, as shown in Figure 19. That is, P-value $= P(t < -1.036) + P(t > 1.036) = 2P(t > 1.036)$, with 10 degrees of freedom.

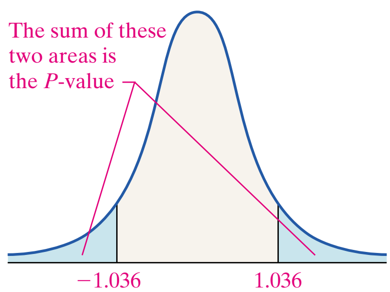
Using Table V, we find the row that corresponds to 10 degrees of freedom. The value 1.036 lies between 0.879 and 1.093. The value of 0.879 has an area under the t-distribution of 0.20 to the right. The area under the t-distribution with 10 degrees of freedom to the right of 1.093 is 0.015.

Because 1.036 is between 0.879 and 1.093, the P-value is between 2(0.20) and 2(0.15). So

$$0.30 < P\text{-value} < 0.40$$

Using MINITAB, we find that the exact P-value is 0.323.

Figure 19



***Step 5*:** The test statistic is $t_0 = -1.036$. Because the test statistic is between the critical values $-t_{0.005} = -3.169$ and $t_{0.005} = 3.169$, the quality-control engineer does not reject the null hypothesis.

***Step 5*:** Because the *P*-value is greater than the level of significance $(0.323 > 0.01)$, the quality-control engineer does not reject the null hypothesis.

***Step 6*:** There is not sufficient evidence to conclude that the Snickers do not have a mean weight of 20.1 grams at the $\alpha = 0.01$ level of significance. The machine should not be shut down.

Now Work Problem 21.

Sections 10.2 and 10.3 discussed performing hypothesis tests about a population mean. The main criterion for choosing which test to use is whether the population standard deviation, σ, is known. Provided that the population from which the sample is drawn is normal or that the sample size is large,

- if σ is known, use the z-test procedures from Section 10.2;
- if σ is unknown, use the t-test procedures from Section 10.3.

In Section 10.4, we will discuss testing hypotheses about a population proportion.

In-Class Activity: Stringing Them Along (Part I)

How skilled are people at estimating the length of a piece of rope? Do you think that they will tend to overestimate its length? Underestimate? Or are you not sure?

1. Look at the piece of rope that your instructor is holding and estimate the length of the rope in inches.
2. Using the null hypothesis H_0: $\mu = \mu_0$, where μ_0 represents the actual length of the rope, select an appropriate alternative hypothesis and a level of significance based on your responses to the questions posed at the beginning of the activity.
3. Obtain the actual length of the rope from your instructor. Combine the data for the entire class and test the hypothesis formed in part (b). What did you conclude?

[**Note:** Save the class data for use in another activity.]

MAKING AN INFORMED DECISION

What Does It Really Weigh?

Many consumer products that we purchase have labels that describe the net weight of the contents. For example, the net weight of a candy bar might be listed as 4 ounces. Choose any consumer product that reports the net weight of the contents on the packaging.

(a) Obtain a random sample of size 8 or more of the consumer product. We will treat the random purchases as a simple random sample. Weigh the contents without the packaging.

(b) If your sample size is less than 30, verify that the population from which the sample was drawn is normal and that the sample does not contain any outliers.

(c) As the consumer, you are concerned only with situations in which you are getting ripped off. Determine the null and alternative hypotheses from the point of view of the consumer.

(d) Test whether the consumer is getting ripped off at the $\alpha = 0.05$ level of significance. Are you getting ripped off? What makes you say so?

(e) Suppose you are the quality-control manager. How would you structure the alternative hypothesis? Test this hypothesis at the $\alpha = 0.05$ level of significance. Is there anything wrong with the manufacturing process? What makes you say so?

10.3 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. State the requirements that must be satisfied to test hypotheses about a population mean with σ unknown.

2. Determine the critical value for a right-tailed test of a population mean with σ unknown at the $\alpha = 0.01$ level of significance with 15 degrees of freedom.

3. Determine the critical value for a two-tailed test of a population mean with σ unknown at the $\alpha = 0.05$ level of significance with 12 degrees of freedom.

4. Determine the critical value for a left-tailed test of a population mean with σ unknown at the $\alpha = 0.05$ level of significance with 19 degrees of freedom.

Skill Building

5. To test H_0: $\mu = 50$ versus H_1: $\mu < 50$, a simple random sample of size $n = 24$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 47.1$ and $s = 10.3$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, determine the critical value.
(c) Draw a t-distribution that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

6. To test H_0: $\mu = 40$ versus H_1: $\mu > 40$, a simple random sample of size $n = 25$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 42.3$ and $s = 4.3$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.1$ level of significance, determine the critical value.
(c) Draw a t-distribution that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

7. NW To test H_0: $\mu = 100$ versus H_1: $\mu \neq 100$, a simple random sample of size $n = 23$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 104.8$ and $s = 9.2$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.01$ level of significance, determine the critical values.
(c) Draw a t-distribution that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

8. To test H_0: $\mu = 80$ versus H_1: $\mu < 80$, a simple random sample of size $n = 22$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 76.9$ and $s = 8.5$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.02$ level of significance, determine the critical value.
(c) Draw a t-distribution that depicts the critical region.
(d) Will the researcher reject the null hypothesis? Why?

9. NW To test H_0: $\mu = 20$ versus H_1: $\mu < 20$, a simple random sample of size $n = 18$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 18.3$ and $s = 4.3$, compute the test statistic.
(b) Draw a t-distribution with the area that represents the P-value shaded.
(c) Approximate and interpret the P-value.
(d) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, will the researcher reject the null hypothesis? Why?

10. To test H_0: $\mu = 4.5$ versus H_1: $\mu > 4.5$, a simple random sample of size $n = 13$ is obtained from a population that is known to be normally distributed.

(a) If $\bar{x} = 4.9$ and $s = 1.3$, compute the test statistic.
(b) Draw a t-distribution with the area that represents the P-value shaded.
(c) Approximate and interpret the P-value.
(d) If the researcher decides to test this hypothesis at the $\alpha = 0.1$ level of significance, will the researcher reject the null hypothesis? Why?

11. To test H_0: $\mu = 105$ versus H_1: $\mu \neq 105$, a simple random sample of size $n = 35$ is obtained.

(a) Does the population have to be normally distributed to test this hypothesis by using the methods presented in this section?
(b) If $\bar{x} = 101.9$ and $s = 5.9$, compute the test statistic.
(c) Draw a t-distribution with the area that represents the P-value shaded.
(d) Determine and interpret the P-value.
(e) If the researcher decides to test this hypothesis at the $\alpha = 0.01$ level of significance, will the researcher reject the null hypothesis? Why?

12. To test H_0: $\mu = 45$ versus H_1: $\mu \neq 45$, a simple random sample of size $n = 40$ is obtained.

(a) Does the population have to be normally distributed to test this hypothesis by using the methods presented in this section?
(b) If $\bar{x} = 48.3$ and $s = 8.5$, compute the test statistic.
(c) Draw a t-distribution with the area that represents the P-value shaded.
(d) Determine and interpret the P-value.
(e) If the researcher decides to test this hypothesis at the $\alpha = 0.01$ level of significance, will the researcher reject the null hypothesis? Why?

Applying the Concepts

13. Effects of Alcohol on the Brain NW In a study published in the *American Journal of Psychiatry* (157:737–744, May 2000), researchers wanted to measure the effect of alcohol on the development of the hippocampal region in adolescents. The hippocampus is the portion of the brain responsible for long-term memory storage. The researchers randomly selected 12 adolescents with alcohol use disorders. They wanted to determine whether the hippocampal volumes in the alcoholic adolescents were less than the normal volume of 9.02 cm^3. An analysis of the sample data revealed that the hippocampal volume is approximately normal with $\bar{x} = 8.10$ and $s = 0.7$. Conduct the appropriate test at the $\alpha = 0.01$ level of significance.

14. Effects of Plastic Resin Para-nonylphenol is found in polyvinyl chloride (PVC) used in the food processing and packaging industries. Researchers wanted to determine the effect this substance had on the organ weight of first-generation mice when both parents were exposed to 50 μg/L of para-nonylphenol in drinking water for 4 weeks. After 4 weeks, the mice were bred. After 100 days, the offspring of the exposed parents were sacrificed and the kidney weights were determined. The mean weight of the 12 offspring was found to be 396.9 mg with a standard deviation of 45.4 mg. Is there significant evidence to conclude that the kidney weight of the offspring whose parents were exposed to 50 μg/L of para-nonylphenol in drinking water for 4 weeks is greater than 355.7 mg, the mean weight of kidneys in normal 100-day old mice at the $\alpha = 0.05$ level of significance?
(*Source*: Vendula Kyselova et al., Effects of *p*-nonylphenol and resveratrol on body and organ weight and in vivo fertility of outbred CD-1 mice, *Reproductive Biology and Endocrinology,* 2003)

15. Got Milk? The U.S. Food and Drug Administration recommends that individuals consume 1000 mg of calcium daily. The International Dairy Foods Association (IDFA) sponsors an advertising campaign aimed at male teenagers. After the campaign, the IDFA obtained a random sample of 50 male teenagers and found that the mean amount of calcium consumed was 1081 mg, with a standard deviation of 426 mg. Conduct a test to determine if the campaign was effective. Use the $\alpha = 0.05$ level of significance.

16. Too Much Salt? A nutritionist believes that children under the age of 10 years are consuming more than the U.S. Food and Drug Administration's recommended daily allowance of sodium, which is 2400 mg. She obtains a random sample of 75 children under the age of 10 and measures their daily consumption of sodium. The mean amount of sodium consumed was determined to be 2993 mg, with a standard deviation of 1489 mg. Is there significant evidence to conclude that children under the age of 10 years are consuming too much sodium? Use the $\alpha = 0.05$ level of significance.

17. Normal Temperature Carl Reinhold August Wunderlich said that the mean temperature of humans is 98.6°F. Researchers Philip Mackowiak, Steven Wasserman, and Myron Levine [*JAMA*, Sept. 23–30 1992; 268(12):1578–80] thought that the mean temperature of humans is less than 98.6°F. They measured the temperature of 148 healthy adults 1 to 4 times daily for 3 days, obtaining 700 measurements. The sample data resulted in a sample mean of 98.2°F and a sample standard deviation of 0.7°F.

(a) Test whether the mean temperature of humans is less than 98.6°F at the $\alpha = 0.01$ level of significance using the classical approach.
(b) Determine and interpret the *P*-value.

18. Normal Temperature Carl Reinhold August Wunderlich said that the mean temperature of humans is 98.6°F. Researchers Philip Mackowiak, Steven Wasserman, and Myron Levine [*JAMA*, Sept. 23–30 1992; 268(12):1578–80] measured the temperatures of 26 females 1 to 4 times daily for 3 days to get a total of 123 measurements. The sample data yielded a sample mean of 98.4°F and a sample standard deviation of 0.7°F.

(a) Using the classical approach, judge whether the normal temperature of women is less than 98.6°F at the $\alpha = 0.01$ level of significance.
(b) Determine and interpret the *P*-value.

19. Age of Death-Row Inmates In 2002, the mean age of an inmate on death row was 40.7 years, according to data obtained from the U.S. Department of Justice. A sociologist wondered whether the mean age of a death-row inmate has changed since then. She randomly selects 32 death-row inmates and finds that their mean age is 38.9, with a standard deviation of 9.6.

(a) Do you believe the mean age has changed? Use the $\alpha = 0.05$ level of significance.
(b) Construct a 95% confidence interval about the mean age. What does the interval imply?

20. Energy Consumption In 2001, the mean household expenditure for energy was \$1493, according to data obtained from the U.S. Energy Information Administration. An economist wanted to know whether this amount has changed significantly from its 2001 level. In a random sample of 35 households, he found the mean expenditure (in 2001 dollars) for energy during the most recent year to be \$1618, with standard deviation \$321.

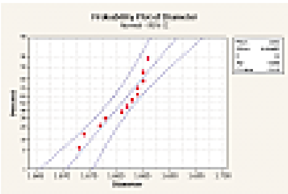
(a) Do you believe that the mean expenditure has changed significantly from the 2001 level at the $\alpha = 0.05$ level of significance?
(b) Construct a 95% confidence interval about the mean energy expenditure. What does the interval imply?

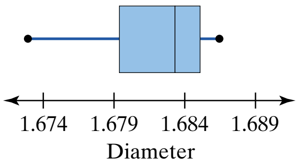
21. Conforming Golf Balls NW The United States Golf Association requires that golf balls have a diameter that is 1.68 inches. To determine if Maxfli XS golf balls conform to USGA standards, a random sample of Maxfli XS golf balls was selected. Their diameters are shown in the table.

1.683	1.677	1.681
1.685	1.678	1.686
1.684	1.684	1.673
1.685	1.682	1.674

Source: Michael McCraith, Joliet Junior College

(a) Because the sample size is small, the engineer must verify that the diameter is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





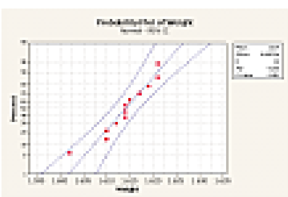
(b) Do the golf balls conform to USGA standards? Use the $\alpha = 0.05$ level of significance.
(c) Determine and interpret the *P*-value.

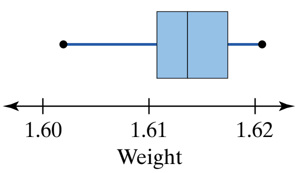
22. Conforming Golf Balls The USGA requires that golf balls have a weight that is less than 1.62 ounces. An engineer for the USGA wants to verify that Maxfli XS balls conform to USGA standards. He obtains a random sample of 12 Maxfli XS golf balls and obtains the weights in the table.

1.614	1.619	1.614
1.614	1.610	1.610
1.621	1.612	1.615
1.621	1.602	1.617

Source: Michael McCraith, Joliet Junior College

(a) Because the sample size is small, the engineer must verify that weight is normally distributed and that the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?



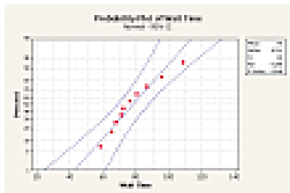


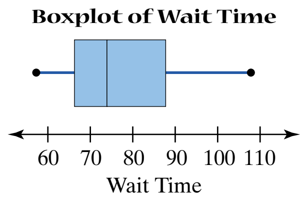
(b) Decide whether the golf balls meet Maxfli's standard at the $\alpha = 0.1$ level of significance.

23. Waiting in Line The mean waiting time at the drive-through of a fast-food restaurant from the time an order is placed to the time the order is received is 84.3 seconds. A manager devises a new drive-through system that he believes will decrease wait time. He initiates the new system at his restaurant and measures the wait time for 10 randomly selected orders. The wait times are provided in the table.

108.5	67.4	58.0	75.9	65.1
80.4	95.5	86.3	70.9	72.0

(a) Because the sample size is small, the manager must verify that wait time is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?



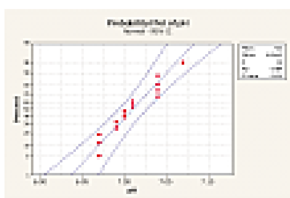


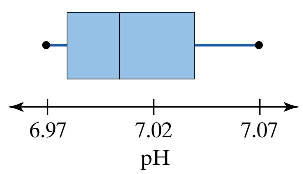
(b) Is the new system effective? Use the $\alpha = 0.1$ level of significance.

24. Calibrating a pH Meter An engineer wants to measure the bias in a pH meter. She uses the meter to measure the pH in 14 neutral substances (pH = 7.0) and obtains the data shown in the table.

7.01	7.04	6.97	7.00	6.99	6.97	7.04
7.04	7.01	7.00	6.99	7.04	7.07	6.97

(a) Because the sample size is small, the engineer must verify that pH is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





(b) Is the pH meter correctly calibrated? Use the $\alpha = 0.05$ level of significance.

25. P/E Ratio A stock analyst believes that the price-to-earnings (P/E) ratio of companies listed on the Standard and Poor's 500 (S&P 500) Index is less than its December 1, 2000, level of 22.0, in response to economic uncertainty. The P/E ratio is the price an investor is willing to pay for $1 of earnings. For example, a P/E of 23 means the investor pays $23 for each $1 of earnings. A higher P/E is an indication of investor optimism. Lower P/Es are generally assigned to companies with lower earnings growth. To test his claim, he randomly samples 14 companies listed on the S&P 500 and calculates their P/E ratios. He obtains the following data:

Company	P/E Ratio	Company	P/E Ratio
Boeing	25.1	Dow Chemical	13.7
General Motors	7.8	Citigroup	18.5
Halliburton	35.8	Merck and Co.	26.8
Norfolk Southern	25.6	Sara Lee	11.9
Agilent Technologies	22.5	Harley-Davidson	37.4
Old Kent Financial	20.0	Circuit City	14.5
Cendent	15.0	Minnesota Mining and Manufacturing	23.8

Source: Checkfree Corporation

(a) Verify that P/E ratios are normally distributed and check for outliers by drawing a normal probability plot and boxplot.
(b) Determine whether the P/E ratio is below the December 1, 2000 level. Use the $\alpha = 0.05$ level of significance.
(c) Determine and interpret the *P*-value.

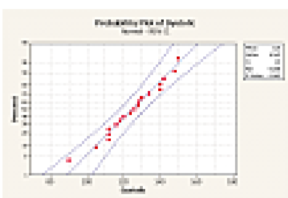
26. Enlarged Prostate Benign prostatic hyperplasia is a common cause of urinary outflow obstruction in aging males. The efficacy of Cardura (doxazosin mesylate) was measured in clinical trials of 173 patients with benign prostatic hyperplasia. Researchers wanted to discover whether Cardura significantly increased the urinary flow rate. It was found that an average increase of 0.8 mL/sec was obtained. This was said to be significant, with a *P*-value less than 0.01. State the null and alternative hypotheses of the researchers and interpret the *P*-value.

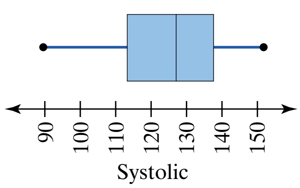
27. Systolic Blood Pressure of Surgical Patients A nursing student maintained that the mean systolic blood pressure of her male patients on the surgical floor was less than 130 mm Hg. She randomly selected 19 male surgical patients and collected the systolic blood pressures shown in the table.

116	150	140	148	105
118	128	112	124	128
140	112	126	130	120
90	134	112	142	

Source: Lora McGuire, Nursing Instructor, Joliet Junior College

(a) Because the sample size is small, the student must verify that the systolic blood pressure is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





(b) The student enters the data into MINITAB and obtains the following results:

T-Test of the Mean

```
Test of mu = 130.00 vs mu < 130.00

Variable          N     Mean    StDev   SE Mean       T        P
Systolic         19      125    15.47      3.55   −1.41    0.088
```

What are the null and alternative hypotheses? Identify the *P*-value. Will the nursing student reject the null hypothesis at the $\alpha = 0.05$ level of significance? State her conclusion.

28. Temperature of Surgical Patients A nursing student suspects that the mean temperature of surgical patients is above the normal temperature, 98.2°F. (See Problem 17.) She randomly selects 32 surgical patients and obtains the temperatures shown in the table.

97.4	98.6	98.2	98.2	98.4	98.6
99.8	97.7	97.8	98.9	97.8	97.8
96.7	97.8	98.3	98.5	98.0	98.7
96.8	98.6	98.4	97.4	99.1	98.7
97.8	99.2	99.2	98.1	98.6	98.4
99.2	98.6				

Source: Lora McGuire, Nursing Instructor, Joliet Junior College

(a) What are the null and alternative hypotheses of the student?
(b) The student enters the data into MINITAB and obtains the following results:

T-Test of the Mean

```
Test of mu = 98.200 vs mu > 98.200

Variable           N     Mean    StDev   SE Mean      T       P
Temperat          32   98.291    0.689     0.122   0.74    0.23
```

What is the P-value of the test? State the nursing student's conclusion.

---

29. Soybean Yield The mean yield per acre of soybeans on farms in the United States in 2003 was 33.5 bushels, according to data obtained from the U.S. Department of Agriculture. A farmer in Iowa claimed the yield was higher than the reported mean. He randomly sampled 35 acres on his farm and determined the mean yield to be 37.1 bushels, with a standard deviation of 2.5 bushels. He computed the P-value to be less than 0.0001 and concluded that the U.S. Department of Agriculture was wrong. Why should his conclusions be looked on with skepticism?

30. Significance Test Applet Load the hypothesis tests for a mean applet.

APPLET

(a) Set the shape to normal, the mean to 100, and the standard deviation to 15. These parameters describe the distribution of IQ scores. Obtain 1000 simple random samples of size $n = 10$ from this population, and test whether the mean is different from 100. How many samples led to a rejection of the null hypothesis if $\alpha = 0.05$? How many would we expect to lead to a rejection of the null hypothesis? For this level of significance, what is the probability of a Type I error?
(b) Set the shape to normal, the mean to 100, and the standard deviation to 15. These parameters describe the distribution of IQ scores. Obtain 1000 simple random samples of size $n = 30$ from this population, and test whether the mean is different from 100. How many samples led to a rejection of the null hypothesis if $\alpha = 0.05$? How many would we expect to lead to a rejection of the null hypothesis? For this level of significance, what is the probability of a Type I error?
(c) Compare the results of parts (a) and (b). Did the sample size have any impact on the number of samples that incorrectly rejected the null hypothesis?

31. Significance Test Applet: Violating Assumptions Load the hypothesis tests for a mean applet.

APPLET

(a) Set the shape to right skewed, the mean to 50, and the standard deviation to 10. Obtain 1000 simple random samples of size $n = 8$ from this population, and test whether the mean is different from 50. How many of the samples led to a rejection of the null hypothesis if $\alpha = 0.05$? How many would we expect to lead to a rejection of the null hypothesis if $\alpha = 0.05$? What might account for any discrepancies?
(b) Set the shape to right skewed, the mean to 50, and the standard deviation to 10. Obtain 1000 simple random samples of size $n = 40$ from this population, and test whether the mean is different from 50. How many of the samples led to a rejection of the null hypothesis if $\alpha = 0.05$? How many would we expect to lead to a rejection of the null hypothesis if $\alpha = 0.05$?

32. Simulation Simulate drawing 40 simple random samples of size $n = 20$ from a population that is normally distribution with mean 50 and standard deviation 10.

(a) Test the null hypothesis H_0: $\mu = 50$ versus the alternative hypothesis H_1: $\mu \neq 50$ for each of the 40 samples using a t-test.
(b) Suppose we were testing this hypothesis at the $\alpha = 0.05$ level of significance. How many of the 40 samples would you expect to result in a Type I error?
(c) Count the number of samples that lead to a rejection of the null hypothesis. Is it close to the expected value determined in part (b)?
(d) Describe why we know a rejection of the null hypothesis results in making a Type I error in this situation.

Consumer Reports® Eyeglass Lenses

Eyeglasses are part medical device and part fashion statement, a marriage that has always made them a tough buy. Aside from the thousands of different frames the consumer has to choose from, various lens materials and coatings can add to the durability, and the cost, of a pair of eyeglasses. One manufacturer even goes so far as to claim that its lenses are "the most scratch-resistant plastic lenses ever made." With a claim like that, we had to test the lenses.

One test involved tumbling the lenses in a drum containing scrub pads of grit of varying size and hardness. Afterward, readings of the lenses' haze were taken on a spectrometer to determine how scratched they had become. To evaluate their scratch resistance, we measured the difference between the haze reading before and after tumbling.

The photo illustrates the difference between an uncoated lens (on the left) and the manufacturer's "scratch-resistant" lens (on the right).

The following table contains the haze measurements both before and after the scratch resistance test for this manufacturer. Haze difference is measured by subtracting the before score from the after score. In other words, haze difference is computed as After–Before.

(a) Suppose it is known that the closest competitor to the manufacturer's lens has a mean haze difference of 1.0. Do the data support the manufacturer's scratch resistance claim?

(b) Write the null and alternative hypotheses, letting μ_{hdiff} represent the mean haze difference for the manufacturer's lens.

(c) We used MINITAB to perform a one-sample *t*-test. The results are shown below.

Using the MINITAB output, answer the following questions:

1. What is the value of the test statistic?
2. What is the *P*-value of the test?
3. What is the conclusion of this test? Write a paragraph for the readers of *Consumer Reports* magazine that explains your findings.

Note to Readers: In many cases, our test protocol and analytical methods are more complicated than described in these examples. The data and discussions have been modified to make the material more appropriate for the audience.

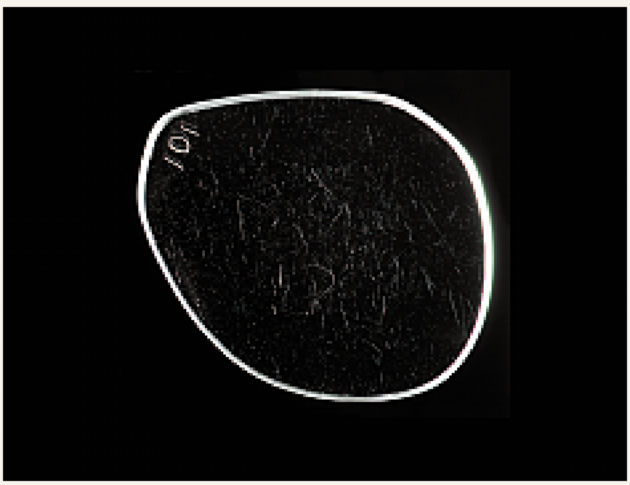
Before	After	Difference
0.18	0.72	0.54
0.16	0.85	0.69
0.20	0.71	0.51
0.17	0.42	0.25
0.21	0.76	0.55
0.21	0.51	0.30

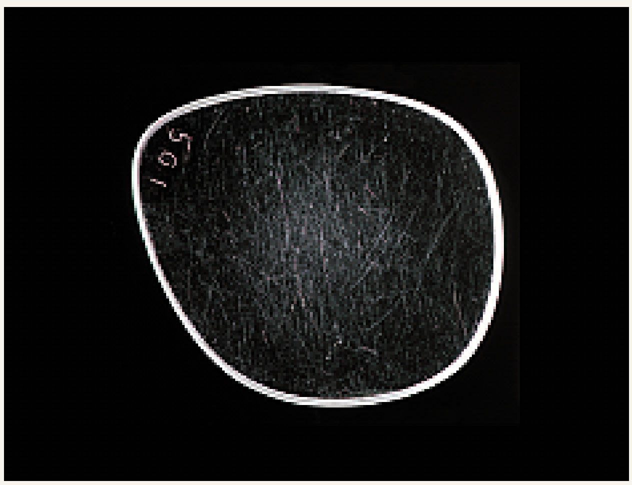
One-Sample T: Difference

```
Test of mu = 1 vs mu < 1

Variable        N          Mean     StDev   SE Mean
Difference      6        0.4733    0.1665    0.0680

Variable      95.0%   Upper Bound       T         P
Difference                0.6103    -7.75     0.000
```





Technology Step by Step **Hypothesis Tests Regarding μ, σ Unknown**

TI-83/84 Plus

***Step 1*:** If necessary, enter raw data in L1.

***Step 2*:** Press STAT, highlight TESTS, and select `2:T-Test`.

***Step 3*:** If the data are raw, highlight DATA; make sure that List1 is set to L1 and Freq is set to 1. If summary statistics are known, highlight STATS and enter the summary statistics. For the value of μ_0, enter the value of the mean stated in the null hypothesis.

***Step 4*:** Select the direction of the alternative hypothesis.

***Step 5*:** Highlight **Calculate** and press ENTER. The TI-83/84 gives the *P*-value.

MINITAB

***Step 1*:** Enter raw data in column C1.

***Step 2*:** Select the **Stat** menu, highlight **Basic Statistics**, then highlight **1-Sample t . . . .**

***Step 3*:** Enter C1 in the cell marked "samples in columns." Enter the value of the mean stated in the null hypothesis in the cell marked "Test mean:" Click Options. In the cell marked "Alternative," select the direction of the alternative hypothesis. Click OK twice.

Excel

***Step 1*:** If necessary, enter raw data in column A.

***Step 2*:** Load the PHStat Add-in.

***Step 3*:** Select the **PHStat menu**, highlight **One Sample Tests . . . ,** and then highlight **t Test for the mean, sigma known . . . .**

***Step 4*:** Enter the value of the null hypothesis and the level of significance, α. If the summary statistics are known, click "Sample statistics known" and enter the sample size, sample mean, and sample standard deviation. If summary statistics are unknown, click "Sample statistics unknown." With the cursor in the "Sample cell range" cell, highlight the data in column A. Click the option corresponding to the desired test [two-tail, upper (right) tail, or lower (left) tail]. Click OK.

10.4 Hypothesis Tests for a Population Proportion

Preparing for This Section Before getting started, review the following:

- Binomial probability distribution (Section 6.2, pp. 298–309)
- Distribution of the sample proportion (Section 8.2, pp. 392–397)

Objective **1** **Test hypotheses about a population proportion**

1 Test Hypotheses about a Population Proportion

Recall that the best point estimate of p, the proportion of the population with a certain characteristic, is given by

$$\hat{p} = \frac{x}{n}$$

where x is the number of individuals in the sample with the specified characteristic and n is the sample size. Recall from Section 8.2 that the sampling distribution of $\hat{p}$ is approximately normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$, provided that the following requirements are satisfied.

- The sample is a simple random sample.
- $np(1-p) \geq 10$.
- $n \leq 0.05N$ (that is, the sample size is no more than 5% of the population size).

Testing hypotheses about the population proportion, p, follows the same logic as the testing of hypotheses about a population mean with σ known. The only difference is that the **test statistic** is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

where p_0 is the value of the population proportion stated in the null hypothesis.

Notice that we are using p_0 in computing the standard error rather than $\hat{p}$ (as we did in computing confidence intervals about p). This is because, when we test a hypothesis, the null hypothesis is always assumed true. Therefore, we are assuming that the population proportion is p_0.

CAUTION
When determining the standard error for the sampling distribution of $\hat{p}$ for hypothesis testing, use the assumed value of the population proportion, p_0.

Testing Hypotheses Regarding a Population Proportion, *p*

To test hypotheses regarding the population proportion, we can use the following steps, provided that

1. The sample is obtained by simple random sampling.
2. $np_0(1 - p_0) \geq 10$ with $n \leq 0.05N$ (the sample size, n, is no more than 5% of the population size, N).

***Step 1*:** Determine the null and alternative hypotheses. The hypotheses can be structured in one of three ways:

Two-Tailed	**Left-Tailed**	**Right-Tailed**
H_0: $p = p_0$	H_0: $p = p_0$	H_0: $p = p_0$
H_1: $p \neq p_0$	H_1: $p < p_0$	H_1: $p > p_0$

Note: p_0 is the assumed value of the population proportion.

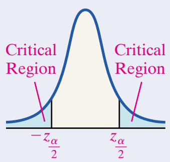
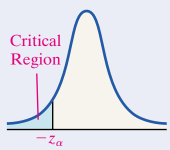
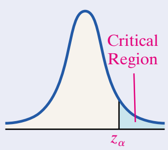
***Step 2*:** Select a level of significance α, depending on the seriousness of making a Type I error.

***Step 3*:** Compute the test statistic

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$$

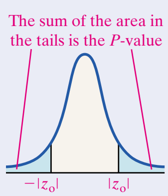
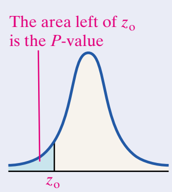
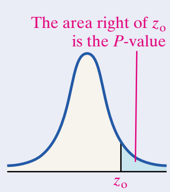
Classical Approach

***Step 4*:** Use Table IV to determine the critical value.

	Two-Tailed	**Left-Tailed**	**Right-Tailed**
Critical value	$-z_{\frac{\alpha}{2}}$ and $z_{\frac{\alpha}{2}}$	$-z_\alpha$	z_α
Critical region			

P-Value Approach

***Step 4*:** Use Table IV to determine the P-value.

Two-Tailed	**Left-Tailed**	**Right-Tailed**
		

Step 5: Compare the critical value with the test statistic.

Two-Tailed	Left-Tailed	Right-Tailed
If $z_0 < -z_{\frac{\alpha}{2}}$ or $z_0 > z_{\frac{\alpha}{2}}$, reject the null hypothesis.	If $z_0 < -z_\alpha$, reject the null hypothesis.	If $z_0 > z_\alpha$, reject the null hypothesis.

Step 5: If P-value $< \alpha$, reject the null hypothesis.

Step 6: State the conclusion.

EXAMPLE 1 Testing Hypotheses about a Population Proportion: Right-Tailed Test

Problem: In 2004, 65% of adult Americans thought that the death penalty was morally acceptable. In a poll conducted by the Gallup Organization May 2–5, 2005, a simple random sample of 1005 adult Americans resulted in 704 respondents stating that they believe the death penalty was morally acceptable when asked, "Do you believe the death penalty is morally acceptable or morally wrong?" The choices "morally acceptable" and "morally wrong" were randomly interchanged for each interview. Is there significant evidence to indicate that the proportion of adult Americans who believe that the death penalty is morally acceptable has increased from the level reported in 2004 at the $\alpha = 0.05$ level of significance?

CAUTION
Always verify the requirements before conducting a hypothesis test.

Approach: We must verify the requirements to perform the hypothesis test; that is, the sample must be a simple random sample, $np_0(1 - p_0) \geq 10$, and the sample size cannot be more than 5% of the population size. Then we follow Steps 1 through 6.

Solution: We want to know if the proportion of adult Americans who believe that the death penalty is morally acceptable has increased since 2004. This can be written $p > 0.65$. The sample is a simple random sample. Also, $np_0(1 - p_0) = (1005)(0.65)(1 - 0.65) = 228.6 > 10$. There are over 200 million adult Americans, so the sample size is less than 5% of the population size. The requirements are satisfied, so we now proceed to Steps 1 through 6.

Step 1: We want to show that $p > 0.65$. We assume "no difference" between opinions in 2004 and 2005, so we have

$$H_0: p = 0.65 \quad \text{versus} \quad H_1: p > 0.65$$

Step 2: The level of significance is $\alpha = 0.05$.

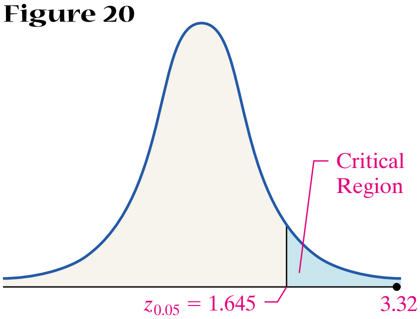
Step 3: The assumed value of the population proportion is $p_0 = 0.65$. The point estimate of the population proportion is $\hat{p} = \frac{x}{n} = \frac{704}{1005} = 0.70$. The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.70 - 0.65}{\sqrt{\frac{0.65(1 - 0.65)}{1005}}} = 3.32$$

Classical Approach

***Step 4*:** Because this is a right-tailed test, we determine the critical value at the $\alpha = 0.05$ level of significance to be $z_{0.05} = 1.645$. The critical region is displayed in Figure 20.

Figure 20

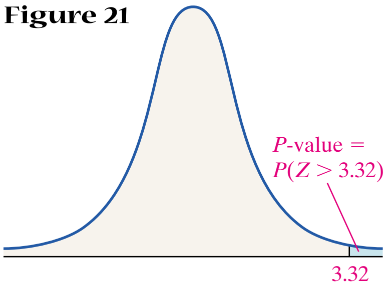


***Step 5*:** The test statistic is $z_0 = 3.32$. We label this point in Figure 20. Because the test statistic is greater than the critical value $(3.32 > 1.645)$, we reject the null hypothesis.

P-Value Approach

***Step 4*:** Because this is a right-tailed test, the P-value is the area under the standard normal distribution to the right of the test statistic, $z_0 = 3.32$, as shown in Figure 21. That is, $P\text{-value} = P(Z > z_0) = P(Z > 3.32) = 1 - P(Z \leq 3.32) = 1 - 0.9995 = 0.0005$.

Figure 21



***Step 5*:** Because the P-value is less than the level of significance $\alpha = 0.05$ $(0.0005 < 0.05)$, we reject the null hypothesis.

***Step 6*:** There is sufficient evidence to conclude that the proportion of adult Americans who believe that the death penalty is morally acceptable has increased since 2004 at the $\alpha = 0.05$ level of significance.

Now Work Problem 9.

EXAMPLE 2 Testing Hypotheses about a Population Proportion, Two-Tailed Test

***Problem*:** The drug Prevnar is a vaccine meant to prevent meningitis. (It also helps control ear infections.) It is typically administered to infants. In clinical trials, the vaccine was administered to 710 randomly sampled infants between 12 and 15 months of age. Of the 710 infants, 121 experienced a loss of appetite. Is there significant evidence to conclude that the proportion of infants who receive Prevnar and experience loss of appetite is different from 0.135, the proportion of children who experience a loss of appetite with competing medications at the $\alpha = 0.01$ level of significance?

***Approach*:** We must verify the requirements to perform the hypothesis test: the sample must be a simple random sample with $np_0(1 - p_0) \geq 10$, and the sample size cannot be more than 5% of the population. Then we follow Steps 1 through 6, as laid out previously.

***Solution*:** We want to know if the proportion of infants who experience a loss of appetite is different from 0.135; that is, $p \neq 0.135$. The sample is a simple random sample. Also, $np_0(1 - p_0) = (710)(0.135)(1 - 0.135) = 82.9 > 10$. Because there are about 1 million babies between 12 and 15 months of age, the sample size is less than 5% of the population size. The requirements are satisfied, so we now proceed to follow Steps 1 through 6.

Step 1: We want to know whether $p \neq 0.135$. This is a two-tailed test.

$$H_0: p = 0.135 \quad \text{versus} \quad H_1: p \neq 0.135$$

Step 2: The level of significance is $\alpha = 0.01$.

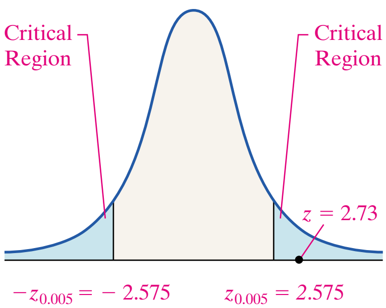
Step 3: The assumed value of the population proportion is $p_0 = 0.135$. The point estimate of the population proportion is $\hat{p} = \frac{x}{n} = \frac{121}{710} = 0.170$. The test statistic is

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.170 - 0.135}{\sqrt{\frac{0.135(1-0.135)}{710}}} = 2.73$$

Classical Approach

Step 4: Because this is a two-tailed test, we determine the critical values at the $\alpha = 0.01$ level of significance to be $-z_{0.01/2} = -z_{0.005} = -2.575$ and $z_{0.01/2} = z_{0.005} = 2.575$. The critical regions are displayed in Figure 22.

Figure 22

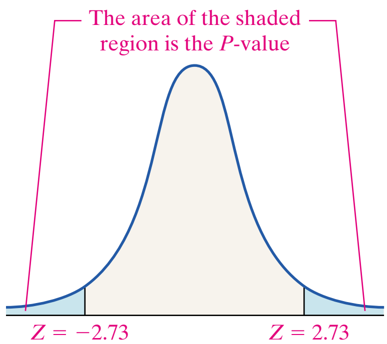


Step 5: The test statistic is $z_0 = 2.73$. We label this point in Figure 22. Because the test statistic is greater than the critical value $(2.73 > 2.575)$, we reject the null hypothesis.

P-Value Approach

Step 4: Because this is a two-tailed test, the P-value is the area under the standard normal distribution to the left of $-z_0 = -2.73$ and to the right of $z_0 = 2.73$ as shown in Figure 23. That is, P-value $= P(Z < -|z_0|) + P(Z > |z_0|) = 2P(Z < -2.73) = 2(0.0032) = 0.0064$.

Figure 23



Step 5: Because the P-value is less than the level of significance $\alpha = 0.01$ $(0.0064 < 0.01)$, we reject the null hypothesis.

Step 6: There is sufficient evidence to conclude that the proportion of infants who experienced a loss of appetite when receiving Prevnar is different from 0.135 at the $\alpha = 0.01$ level of significance.

Now Work Problem 15.

EXAMPLE 3 Testing Hypotheses Regarding a Population Proportion Using Technology

Problem: Test the hypotheses presented in Example 2 by obtaining the P-value using statistical software or a graphing calculator with advanced statistical features.

Approach: We will use MINITAB to obtain the P-value. The steps for testing hypotheses regarding a population proportion using the TI-83/84 Plus graphing calculator, MINITAB, and Excel are given in the Technology Step by Step on page 500.

Result: Figure 24 shows the results using MINITAB. The P-value is 0.006. Because the P-value is less than the level of significance $(0.006 < 0.01)$, we reject the null hypothesis.

Figure 24 **Test and CI for One Proportion**

```
Test of p = 0.135 vs p not = 0.135

Sample      X      N      Sample          99% CI            Z-Value    P-Value
1         121    710    0.170423    (0.134075, 0.206770)      2.76      0.006
```

In-Class Activity (Hypothesis Testing): Taste the Rainbow

The advertising campaign for Skittles® Brand candy in 2005 said to "Taste the Rainbow!". While the original candies do not have all the colors of the rainbow, they do come in red, orange, yellow, green, and purple (violet). But are the proportions of each color the same? If so, the proportion of each color would be $p = 0.2$.

(a) Obtain a 1-pound bag of Skittles (original flavor).
(b) Select one of the five original colors. Count the total number of candies in your bag, as well as the number for the color you selected.
(c) Is $np_0(1 - p_0) \geq 10$? If not, what could you do?
(d) Determine whether the proportion of the selected color is different from $p = 0.2$.
(e) Compare your results to others in the class. Did everyone arrive at the same conclusion for the same color? What about different colors?

10.4 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. State the assumptions required to test a hypothesis about a population proportion.

2. A poll conducted by CNN, *USA Today*, and Gallup reported the following results: "According to the most recent CNN/*USA Today*/Gallup poll, conducted June 28–July 1, a majority of Americans (52%) approve of the job Bush is doing as president." The poll results were obtained by conducting simple random sample of 1014 adults aged 18 years old or older, with a margin of error of ±3 percentage points. State what is wrong with the conclusions presented by the pollsters.

Skill Building

In Problems 3–8, test the hypothesis, using (a) the classical approach and then (b) the P-value approach. Be sure to verify the requirements of the test.

3. H_0: $p = 0.3$ versus H_1: $p > 0.3$
$n = 200$; $x = 75$; $\alpha = 0.05$

4. H_0: $p = 0.6$ versus H_1: $p < 0.6$
$n = 250$; $x = 124$; $\alpha = 0.01$

5. H_0: $p = 0.55$ versus H_1: $p < 0.55$
$n = 150$; $x = 78$; $\alpha = 0.1$

6. H_0: $p = 0.25$ versus H_1: $p < 0.25$
$n = 400$; $x = 96$; $\alpha = 0.1$

7. H_0: $p = 0.9$ versus H_1: $p \neq 0.9$
$n = 500$; $x = 440$; $\alpha = 0.05$

8. H_0: $p = 0.4$ versus H_1: $p \neq 0.4$
$n = 1000$; $x = 420$; $\alpha = 0.01$

Applying the Concepts

9. Lipitor The drug Lipitor is meant to reduce total cholesterol and LDL cholesterol. In clinical trials, 19 out of 863 patients taking 10 mg of Lipitor daily complained of flulike symptoms. Suppose that it is known that 1.9% of patients taking competing drugs complain of flulike symptoms. Is there sufficient evidence to conclude that more than 1.9% of Lipitor users experience flulike symptoms as a side effect at the $\alpha = 0.01$ level of signifi-cance?

NW

10. Nexium Nexium is a drug that can be used to reduce the acid produced by the body and heal damage to the esophagus due to acid reflux. Suppose the manufacturer of Nexium claims that more than 94% of patients taking Nexium are healed within 8 weeks. In clinical trials, 213 of 224 patients suffering from acid reflux disease were healed after 8 weeks. Test the manufacturer's claim at the $\alpha = 0.01$ level of significance.

11. Americans Reading In a Gallup Poll conducted May 20–22, 2005, 835 of 1006 adults aged 18 or older said they had read at least one book during the previous year. In December 1990, 81% of adults aged 18 or older had read at least one book during the previous year. Is there sufficient evidence to conclude that the percent of adults who have read at least one book in the last year is different from 1990 at the $\alpha = 0.05$ significance level?

12. Haunted Houses In September 1996, 33% of adult Americans believed in haunted houses. In a Gallup Poll conducted June 6–8, 2005, 370 of 1002 adult Americans aged 18 or older believed in haunted houses. Is there sufficient evidence to conclude that the proportion of adult Americans who believe in haunted houses has increased at the $\alpha = 0.05$ level of significance?

13. Eating Together In December 2001, 38% of adults with children under the age of 18 reported that their family ate dinner together 7 nights a week. Suppose in a recent poll, 337 of 1122 adults with children under the age of 18 reported that their family ate dinner together 7 nights a week. Has the proportion of families with children under the age of 18 who eat dinner together 7 nights a week decreased? Use the $\alpha = 0.05$ significance level.

14. Critical Job Skills In August 2003, 56% of employed adults in the United States reported that basic mathematical skills were critical or very important to their job. The supervisor of the job placement office at a 4-year college thinks this percentage has increased due to increased use of technology in the workplace. He takes a random sample of 480 employed adults and finds that 297 of them feel that basic mathematical skills are critical or very important to their job. Has the percentage of employed adults who feel basic mathematical skills are critical or very important to their job increased? Use the $\alpha = 0.05$ level of significance.

15. Chance for Promotion In October 1998, 30% of employed adults were satisfied with their chances for promotion. A human resource manager wants to determine if this percentage has changed significantly since then. She randomly selects 280 employed adults and finds that 112 of them are completely satisfied with their chances for promotion. Is the percentage of employed adults satisfied with their chances for promotion changed significantly from the percentage in 1998, at the $\alpha = 0.1$ level of significance?

NW

16. Living Alone? In 2000, 58% of females aged 15 years of age and older lived alone, according to the U.S. Census Bureau. A sociologist wants to know if this percentage is different today, so she obtains a random sample of 500 females aged 15 years of age and older and finds that 285 are living alone. Is there evidence at the $\alpha = 0.1$ level of significance to support belief in a change?

17. Confidence in Schools In 1995, 40% of adults aged 18 years or older reported that they had "a great deal" of confidence in the public schools. On June 1, 2005, the Gallup Organization released results of a poll in which 372 of 1004 adults aged 18 years or older stated that they had "a great deal" of confidence in the public schools. Does the evidence suggest at the $\alpha = 0.05$ level of significance that the proportion of adults aged 18 years or older having "a great deal" of confidence in the public schools is significantly lower in 2005 than the 1995 proportion?

18. Pathological Gamblers Pathological gambling is an impulse-control disorder. The American Psychiatric Association lists 10 characteristics that indicate the disorder in its DSM-IV manual. The National Gambling Impact Study Commission randomly selected 2417 adults and found that 35 were pathological gamblers. Is there evidence to conclude that more than 1% of the adult population are pathological gamblers at the $\alpha = 0.05$ level of significance?

19. Talk to the Animals In a survey conducted by the American Animal Hospital Association, 37% of respondents stated that they talk to their pets on the answering machine or telephone. A veterinarian found this result hard to believe, so he randomly selected 150 pet owners and discovered that 54 of them spoke to their pet on the answering machine or telephone. Does the veterinarian have a right to be skeptical? Use the $\alpha = 0.05$ level of significance.

20. Eating Salad According to a survey conducted by the Association for Dressings and Sauces (this is an actual association!), 85% of American adults eat salad at least once a week. A nutritionist suspects that the percentage is higher than this. She conducts a survey of 200 American adults and finds that 171 of them eat salad at least once a week. Conduct the appropriate test that addresses the nutritionist's suspicions. Use the $\alpha = 0.10$ level of significance.

21. Statistics in the Media One of the more popular statistics reported in the media is the president's job approval rating. The approval rating is reported as the proportion of Americans who approve of the job that the sitting president is doing and is typically based on a random sample of about 1100 Americans.

(a) This proportion tends to fluctuate from week to week. Name some reasons for the fluctuation in the statistic.

(b) A recent article had the headline "Bush Ratings Show Decline." This headline was written because an April poll showed President Bush's approval rating to be 0.48 (48%). A poll in June based on 1100 randomly selected Americans showed that 506 approved of the job Bush was doing. Do the results of the June poll indicate that the proportion of Americans who approve of the job Bush is doing is significantly less than April's level? Explain.

22. Statistics in the Media In May 2002, 71% (0.71) of Americans favored the death penalty for a person convicted of murder. In May 2005, 1005 adult Americans were asked by the Gallup Organization, "Are you in favor of the death penalty for a person convicted of murder?" Of the 1005 adults surveyed, 744 responded yes. The headline in the article reporting the survey results stated, "Americans' Views of Death Penalty More Positive This Year." Use a test of significance to support or refute this headline.

Technology Step by Step — Hypothesis Tests Regarding a Population Proportion

TI-83/84 Plus

***Step 1*:** Press STAT, highlight TESTS, and select `5:1-PropZTest`.

***Step 2*:** For the value of p_0, enter the "status quo" value of the population proportion.

***Step 3*:** Enter the number of successes, x, and the sample size, n.

***Step 4*:** Select the direction of the alternative hypothesis.

***Step 5*:** Highlight Calculate or Draw, and press ENTER. The TI-83 or TI-84 gives the P-value.

MINITAB

***Step 1*:** Select the **Stat** menu, highlight **Basic Statistics**, then highlight **1-Proportion**.

***Step 2*:** Select "Summarized data."

***Step 3*:** Enter the number of trials, n, and the number of successes, x.

***Step 4*:** Click Options. Enter the "status quo" value of the population proportion in the cell "Test proportion." Enter the direction of the alternative hypothesis. If $np_0(1 - p_0) \geq 10$, check the box marked "Use test and interval based on normal distribution." Click OK twice.

Excel

***Step 1*:** Load the PHStat Add-in.

***Step 2*:** Select the **PHStat menu**, highlight **One Sample Tests . . .**, then highlight **Z Test for proportion**.

***Step 3*:** Enter the value of the null hypothesis, the level of significance, α, the number of successes, x, and the number of trials, n. Click the option corresponding to the desired test [two-tail, upper (right) tail, or lower (left) tail]. Click OK.

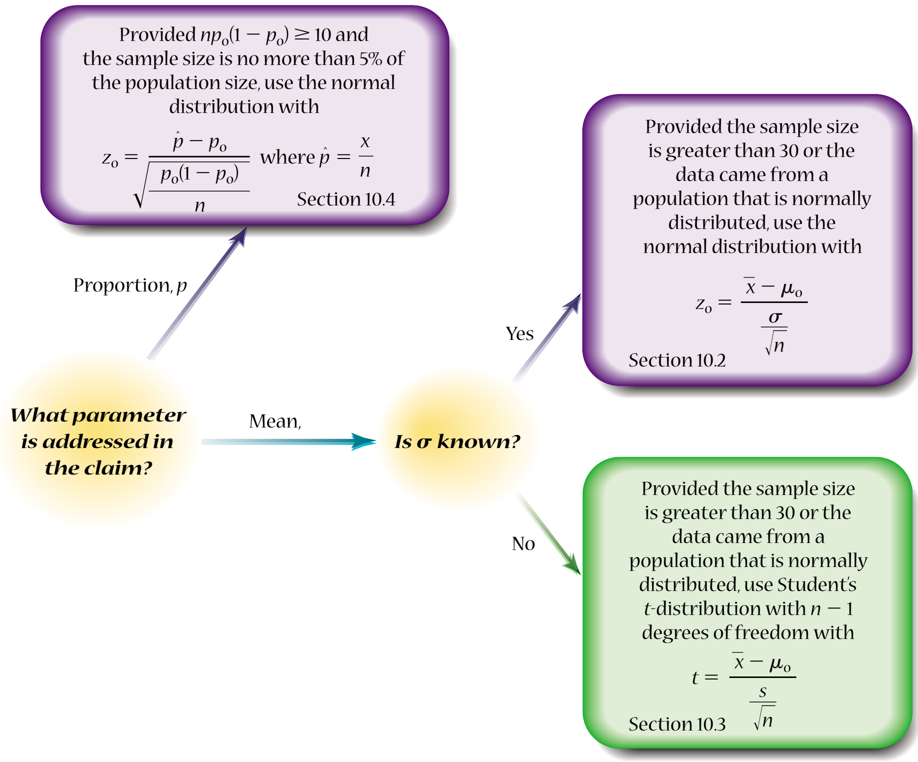
10.5 Putting It All Together: Which Method Do I Use?

Objective 1 Determine the appropriate hypothesis test to perform

1 Determine the Appropriate Hypothesis Test to Perform

Perhaps the most difficult aspect of testing hypotheses is determining which hypothesis test to conduct. To assist in the decision making, we present Figure 25 which shows which approach to take in testing hypotheses for the three parameters discussed in this chapter.

Figure 25



10.5 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. What are the requirements that must be satisfied to test a hypothesis about a population mean? When do we use the normal model to test a hypothesis about a population mean? When do we use Student's t-distribution to test a hypothesis about a population mean?

2. What are the requirements that must be satisfied before we can test a hypothesis about a population proportion?

Skill Building

In Problems 3–12, conduct the appropriate test.

3. A simple random sample of size $n = 14$ is drawn from a population that is normally distributed with $\sigma = 20$. The sample mean is found to be $\bar{x} = 60$. Test whether the population mean is less than 70 at the $\alpha = 0.1$ level of significance.

4. A simple random sample of size $n = 19$ is drawn from a population that is normally distributed. The sample mean is found to be 0.8, and the sample standard deviation is found to be 0.4. Test whether the population mean is less than 1.0 at the $\alpha = 0.01$ level of significance.

5. A simple random sample of size $n = 200$ individuals with a valid driver's license is asked if they drive an American-made automobile. Of the 200 individuals surveyed, 115 responded that they drive an American-made automobile. Test whether more than half of those with a valid driver's license drive an American-made automobile at the $\alpha = 0.05$ level of significance.

6. A simple random sample of size $n = 65$ is drawn from a population. The sample mean is found to be 583.1, and the sample standard deviation is found to be 114.9. Test the claim that the population mean is different from 600 at the $\alpha = 0.1$ level of significance.

7. A simple random sample of size $n = 15$ is drawn from a population that is normally distributed. The sample mean is found to be 23.8, and the sample standard deviation is found to be 6.3. Test whether the population mean is different from 25 at the $\alpha = 0.01$ level of significance.

8. A simple random sample of size $n = 25$ is drawn from a population that is normally distributed with $\sigma = 7$. The sample mean is found to be 53.2. Test whether the population mean is different from 55 at the $\alpha = 0.05$ level of significance.

9. A simple random sample of size $n = 40$ is drawn from a population. The sample mean is found to be 108.5, and the sample standard deviation is found to be 17.9. Test the claim that the population mean is greater than 100 at the $\alpha = 0.05$ level of significance.

10. A simple random sample of size $n = 320$ adults was asked their favorite ice cream flavor. Of the 320 individuals surveyed, 58 responded that they preferred mint chocolate chip. Test the claim that less than 25% of adults prefer mint chocolate chip ice cream at the $\alpha = 0.01$ level of significance.

Applying the Concepts

11. Family Size In 1985, the mean for the ideal number of children for a family was considered to be 2.5. A Gallup Poll of 1006 adults aged 18 or older conducted February 16–17, 2004, reported the mean for the ideal number of children to be 2.6.

(a) Assuming $\sigma = 1.2$, is there sufficient evidence to conclude that the mean ideal number of children has changed since 1985 at the $\alpha = 0.1$ level of significance?

(b) Do the results have any practical significance? Explain.

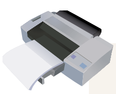
12. Americans Online In spring 2002, 66% of adults in the United States aged 18 years or older had Internet access. A Harris Interactive poll in February and May of 2005 found that 1496 of 2022 adults surveyed had Internet access. Is this enough evidence to conclude that more adults had Internet access in 2005 than in 2002 at the $\alpha = 0.05$ level of significance?

13. Tattoos In 2001, 23% of American university undergraduate students had at least one tattoo. A health practitioner obtains a random sample of 1026 university undergraduates and finds that 254 have at least one tattoo. Has the proportion of American university undergraduate students with at least one tattoo changed since 2001? Use the $\alpha = 0.1$ level of significance.

14. Mortgage Rates In 2001, the mean contract interest rate for a conventional 30-year first loan for the purchase of a single-family home was 6.3 percent, according to the U.S. Federal Housing Board. A real estate agent believes that interest rates are lower today and obtains a random sample of 41 recent 30-year conventional loans. The mean interest rate was found to be 6.05 percent, with a standard deviation of 1.75 percent. Is this enough evidence to conclude that interest rates are lower at the $\alpha = 0.05$ level of significance?

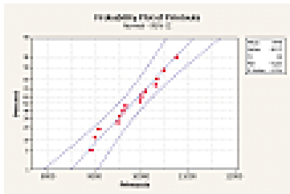
15. Auto Insurance According to the Insurance Information Institute, the mean expenditure for auto insurance in the United States was \$774 for 2002. An insurance salesman obtains a random sample of 35 auto insurance policies and determines the mean expenditure to be \$735 with a standard deviation of \$48.31. Is there enough evidence to conclude that the mean expenditure for auto insurance is different from the 2002 amount at the $\alpha = 0.01$ level of significance?

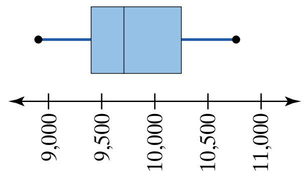
16. Toner Cartridge The manufacturer of a toner cartridge claims the mean number of printouts is 10,000 for each cartridge. A consumer advocate believes the actual mean number of printouts is lower. He selects a random sample of 14 such cartridges and obtains the following number of printouts:



9600	10,300	9000	10,750	9490	9080	9655
9520	10,070	9999	10,470	8920	9964	10,330

(a) Because the sample size is small, he must verify that the number of printouts is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





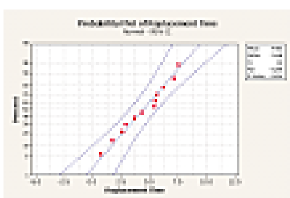
(b) Are the consumer advocate's beliefs justified? Use the $\alpha = 0.05$ level of significance.

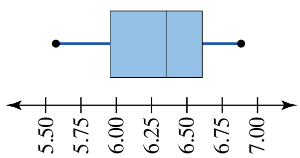
17. Vehicle Emission Inspection A certain vehicle emission inspection station states that the mean wait time for customers is less than 8 minutes. A local resident is skeptical and collects a random sample of 49 wait times for customers at the testing station. He finds that the sample mean is 7.34 minutes, with a standard deviation of 3.2 minutes. Is the resident's skepticism justified? Use the $\alpha = 0.01$ level of significance.

18. Lights Out With a previous contractor, the mean time to replace a streetlight was 3.2 days. A city councilwoman thinks that the new contractor is not getting the streetlights replaced as quickly. She selects a random sample of 12 streetlight service calls and obtains the following times to replacement (in days).

6.2	7.1	5.4	5.5	7.5	2.6
4.3	2.9	3.7	0.7	5.6	1.7

(a) Because the sample size is small, she must verify that replacement time is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown below. Are the conditions for testing the hypothesis satisfied?





(b) Is there evidence that the contractor is not performing? Use the $\alpha = 0.05$ level of significance.

CHAPTER 10 Review

Summary

In this chapter, we discussed hypothesis testing. A statement is made regarding a population parameter, which leads to a null and an alternative hypothesis. The null hypothesis is assumed true. Given sample data, we either reject or do not reject the null hypothesis. In performing a hypothesis test, there is always the possibility of making a Type I error (rejecting the null hypothesis when it is true) or of making a Type II error (not rejecting the null hypothesis when it is false). The probability of making a Type I error is equal to the level of significance, α, of the test.

We discussed three types of hypothesis tests in this chapter. First, we performed tests about a population mean with σ known. Second, we performed tests about a population mean with σ unknown.

In both of these cases, we required that either the sample size be large ($n \geq 30$) or the population be approximately normal with no outliers. For small sample sizes, we verified the normality of the data by using normal probability plots. Boxplots were used to check for outliers.

The third test we performed regarded hypothesis tests about a population proportion. To perform these tests, we required a large sample size so that $np(1 - p) \geq 10$, yet the sample size could be no more than 5% of the population size.

All three hypothesis tests were performed by using classical methods and the P-value approach. The P-value approach to testing hypotheses has appeal, because the rejection rule is always to reject the null hypothesis if the P-value is less than the level of significance, α.

Formulas

Test Statistics

- $z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ follows the standard normal distribution if the population from which the sample was drawn is normal or if the sample size is large ($n \geq 30$).
- $t_0 = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ follows Student's t-distribution with $n - 1$ degrees of freedom if the population from which the sample was drawn is normal or if the sample size is large ($n \geq 30$).
- $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1 - p_0)}{n}}}$ follows the standard normal distribution if $np_0(1 - p_0) \geq 10$ and $n \leq 0.05N$.

Type I and Type II Errors

- $\alpha = P(\text{Type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$
- $\beta = P(\text{Type II error}) = P(\text{not rejecting } H_0 \text{ when } H_1 \text{ is true})$

Vocabulary

Hypothesis (p. 454)
Hypothesis testing (p. 455)
Null hypothesis (p. 455)
Alternative hypothesis (p. 455)
Two-tailed test (p. 455)
Left-tailed test (p. 455)
Right-tailed test (p. 455)
Type I error (p. 457)
Type II error (p. 457)
Level of significance (p. 459)
Statistically significant (p. 463)
Test statistic (p. 466)
Critical value (p. 466)
Critical region (p. 466)
Decision rule (p. 466, 470)
Robust (p. 467)
P-value (p. 469)
Practical significance (p. 474)

Objectives

Section	You should be able to . . .	Examples	Review Exercises
10.1	1 Determine the null and alternative hypothesis (p. 454)	2	1, 2
	2 Understand Type I and Type II errors (p. 457)	3	1–4, 15(b), (c); 16(b), (c)
	3 State conclusions to hypothesis tests (p. 459)	4	1–2
10.2	1 Understand the logic of hypothesis testing (p. 462)	pp. 462–465	22
	2 Test hypotheses about a population mean with σ known using the classical approach (p. 465)	1 and 2	5, 6, 13, 14, 17, 18
	3 Test hypotheses about a population mean with σ known using P-values (p. 469)	3–5	5, 6, 13, 14, 17, 18

	4 Test hypotheses about a population mean with σ known using confidence intervals (p. 473)	6	6(f)
	5 Understand the difference between statistical significance and practical significance (p. 474)	7	19
10.3	**1** Test hypotheses about a population mean with σ unknown (p. 480)	1–3	7, 8, 11, 12, 15(a), 16(a), 21
10.4	**1** Test hypotheses about a population proportion (p. 493)	1–3	9, 10, 19, 20
10.5	**1** Determine the appropriate hypothesis test to perform (p. 500)	pp. 500–501	5–21

Review Exercises

In Problems 1 and 2, (a) determine the null and alternative hypotheses, (b) explain what it would mean to make a Type I error, (c) explain what it would mean to make a Type II error, (d) state the conclusion that would be drawn if the null hypothesis is not rejected, and (e) state the conclusion that would be reached if the null hypothesis is rejected.

1. Credit Card Debt According to the *Statistical Abstract of the United States*, the mean outstanding credit card debt per cardholder was $4277 in 2000. A consumer credit counselor believes that the mean outstanding credit card debt per cardholder is now more than the 2000 amount.

2. Downloading Music According to a study by Ipsos-Reid, 61% of Internet users aged 18 to 24 years old had downloaded music from the Internet by the end of 2000. A researcher believes that the percentage is now higher than 61%.

3. Suppose a test is conducted at the $\alpha = 0.05$ level of significance. What is the probability of a Type I error?

4. Suppose β is computed to be 0.113. What is the probability of a Type II error?

5. To test H_0: $\mu = 30$ versus H_1: $\mu < 30$, a simple random sample of size $n = 12$ is obtained from a population that is known to be normally distributed with $\sigma = 4.5$.
(a) If the sample mean is determined to be $\bar{x} = 28.6$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, determine the critical value.
(c) Draw a normal curve that depicts the rejection region.
(d) Will the researcher reject the null hypothesis? Why?
(e) What is the P-value?

6. To test H_0: $\mu = 65$ versus H_1: $\mu \neq 65$, a simple random sample of size $n = 23$ is obtained from a population that is known to be normally distributed with $\sigma = 12.3$.
(a) If the sample mean is determined to be $\bar{x} = 70.6$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.1$ level of significance, determine the critical values.
(c) Draw a normal curve that depicts the rejection region.
(d) Will the researcher reject the null hypothesis? Why?
(e) What is the P-value?
(f) Conduct the test by constructing a 90% confidence interval.

7. To test H_0: $\mu = 8$ versus H_1: $\mu \neq 8$, a simple random sample of size $n = 15$ is obtained from a population that is known to be normally distributed.
(a) If $\bar{x} = 7.3$ and $s = 1.8$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.02$ level of significance, determine the critical values.
(c) Draw a t-distribution that depicts the rejection region.
(d) Will the researcher reject the null hypothesis? Why?
(e) Determine the P-value.

8. To test H_0: $\mu = 3.9$ versus H_1: $\mu < 3.9$, a simple random sample of size $n = 25$ is obtained from a population that is known to be normally distributed.
(a) If $\bar{x} = 3.5$ and $s = 0.9$, compute the test statistic.
(b) If the researcher decides to test this hypothesis at the $\alpha = 0.05$ level of significance, determine the critical value.
(c) Draw a t-distribution that depicts the rejection region.
(d) Will the researcher reject the null hypothesis? Why?
(e) Determine the P-value.

In Problems 9 and 10, test the hypothesis at the $\alpha = 0.05$ level of significance, using (a) the classical approach and (b) the P-value approach. Be sure to verify the requirements of the test.

9. H_0: $p = 0.6$ versus H_1: $p > 0.6$
$n = 250$; $x = 165$; $\alpha = 0.05$

10. H_0: $p = 0.35$ versus H_1: $p \neq 0.35$
$n = 420$; $x = 138$; $\alpha = 0.01$

11. Linear Rotary Bearing A linear rotary bearing is designed so that the distance between the retaining rings is 0.875 inch. The quality-control manager suspects that the manufacturing process needs to be recalibrated and that the mean distance between the retaining rings is longer than 0.875 inch. In a random sample of 36 bearings, he finds the sample mean distance between the retaining rings to be 0.876 inch with standard deviation 0.005 inch. Should the machine be recalibrated? Use the $\alpha = 0.05$ level of significance.

12. Education Pays The U.S. Census Bureau reported that the mean annual salary in 2002 was $51,194 for an individual whose highest degree was a bachelor's. A government economist believes that the mean annual salary for individuals whose highest degree is a bachelor's is different today. She obtains a random sample of 300 employed adults whose highest degree is a bachelor's and determines the mean annual salary to be $55,988 with a standard deviation of $26,855 (both in 2002 dollars). Does the evidence indicate salaries are different today? Use the $\alpha = 0.05$ level of significance.

13. SAT Math Scores A mathematics instructor wanted to know whether or not use of a calculator improves SAT math scores. In 2000, the SAT math scores of students who used a calculator once or twice weekly were normally distributed, with a mean of 474 and a standard deviation 103. In a random sample of 50 students who use a calculator every day, the mean score was 539. Does the evidence support the instructor's assertion that students who use a calculator "frequently" score better on the SAT math portion than those who use a calculator "infrequently"? Use the $\alpha = 0.01$ level of significance.

14. Birth Weight An obstetrician wants to determine whether a new diet significantly increases the birth weight of babies. In 2002, birth weights of full-term babies (gestation period of 37 to 41 weeks) were normally distributed, with mean 7.53 pounds and standard deviation 1.15 pounds, according to the *National Vital Statistics Report*, Vol. 48, No. 3. The obstetrician randomly selects 50 recently pregnant mothers and persuades them to partake of this new diet. The obstetrician then records the birth weights of the babies and obtains a mean of 7.79 pounds. Does the new diet increase the birth weights of newborns? Use the $\alpha = 0.01$ level of significance.

15. High Cholesterol A nutritionist maintains that 20- to 39-year-old males consume too much cholesterol. The USDA-recommended daily allowance of cholesterol is 300 mg. In a survey conducted by the U.S. Department of Agriculture of 404 20- to 39-year-old males, it was determined the mean daily cholesterol intake was 326 milligrams, with standard deviation 342 milligrams.

(a) Is there evidence to support the nutritionist's belief at the $\alpha = 0.05$ level of significance?
(b) What would it mean for the nutritionist to make a Type I error? A Type II error?
(c) What is the probability the nutritionist will make a Type I error?

16. Sodium A nutritionist thinks that 20- to 39-year-old females consume too much sodium. The USDA-recommended daily allowance of sodium is 2400 mg. In a survey conducted by the U.S. Department of Agriculture of 257 20- to 39-year-old females, it was determined the mean daily sodium intake was 2919 milligrams and the standard deviation was 906 milligrams.

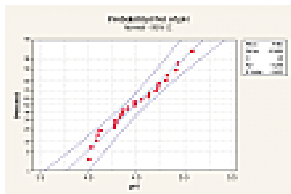
(a) Is there evidence to support the nutritionist's belief at the $\alpha = 0.10$ level of significance?
(b) What would it mean for the nutritionist to make a Type I error? A Type II error?
(c) What is the probability the nutritionist will make a Type I error?

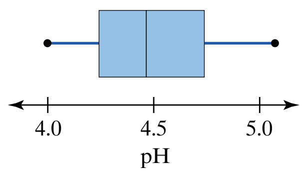
17. Acid Rain In 1990, the mean pH level of the rain in Barnstable County, Massachussetts, was 4.61. A biologist fears that the acidity of rain has increased (in other words that the pH level of the rain has decreased). She draws a random sample of 25 rain dates in 2004 and obtains the following data:

4.80	4.27	4.09	4.55	5.08	4.34	4.08
4.36	4.82	4.70	4.40	4.73	4.62	4.48
4.28	4.28	4.48	4.72	4.12	4.00	4.93
4.91	4.32	4.63	4.03			

Source: National Atmospheric Deposition Program

(a) Because the sample size is small, she must verify that the pH level is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





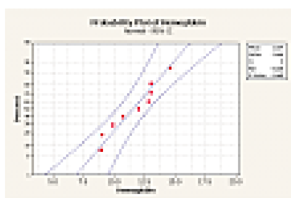
(b) Assuming $\sigma = 0.26$, judge whether the biologist's fears are justified. Use the $\alpha = 0.01$ level of significance.

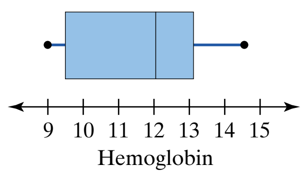
18. Hemoglobin A medical researcher maintains that the mean hemoglobin reading of surgical patients is different from 14.0 grams per deciliter. She randomly selects nine surgical patients and obtains the following data:

14.6	12.8	8.9	9.0	9.9
10.7	13.0	12.0	13.0	

Source: Lora McGuire, Nursing Instructor, Joliet Junior College

(a) Because the sample size is small, the student must verify that hemoglobin is normally distributed and the sample does not contain any outliers. The normal probability plot and boxplot are shown. Are the conditions for testing the hypothesis satisfied?





(b) Assuming that $\sigma = 2.001$, test the researcher's beliefs. Use the $\alpha = 0.01$ level of significance.

19. Tuberculosis According to the Centers for Disease Control, 56% of all tuberculosis cases in 1999 were of foreign-born residents of the United States. A researcher believes that this proportion has increased from its 1999 level. She obtains a simple random sample of 300 tuberculosis cases in the United States and determines that 170 of them are foreign-born. Is there sufficient evidence to support the belief that the percentage of cases of tuberculosis of foreign-born residents has increased at the $\alpha = 0.01$ level of significance?

20. Phone Purchases In 1997, 39.4% of females had ordered merchandise or services by phone in the last 12 months. A market research analyst feels that the percentage of females ordering merchandise or services by phone has declined from the 1997 level because of Internet purchases. She obtains a random sample of 500 females and determines that 191 of them have ordered merchandise or services by phone in the last 12 months. Does the evidence suggest that the percentage of females ordering merchandise or services by phone has decreased from its 1997 proportion at the $\alpha = 0.10$ level of significance?

21. A New Teaching Method A large university has a college algebra enrollment of 5000 students each semester. Because of space limitations, the university decides to offer its college algebra courses in a self-study format in which students learn independently but have access to tutors and other help in a lab setting. Historically, students in traditional college algebra scored 73.2 points on the final exam and the coordinator of this course is concerned that test scores are going to decrease in the new format. At the end of the first semester using the new delivery system, 3851 students took the final exams and had a mean score of 72.8 and a standard deviation of 12.3. Treating these students as a simple random sample of all students, determine whether or not the scores decreased significantly at the $\alpha = 0.05$ level of significance. Do you think that the decrease in scores has any practical significance?

22. In your own words, explain the procedure for conducting a test regarding a population mean when assuming the population standard deviation is known.

THE CHAPTER 10 CASE STUDY IS LOCATED ON THE CD THAT ACCOMPANIES THIS TEXT.